Jakob Schwichtenberg

Physics from Symmetry

Second Edition

Springer

Undergraduate Lecture Notes in Physics

Undergraduate Lecture Notes in Physics (ULNP) publishes authoritative texts covering topics throughout pure and applied physics. Each title in the series is suitable as a basis for undergraduate instruction, typically containing practice problems, worked examples, chapter summaries, and suggestions for further reading.

ULNP titles must provide at least one of the following:

- An exceptionally clear and concise treatment of a standard undergraduate subject.
- A solid undergraduate-level introduction to a graduate, advanced, or non-standard subject.
- A novel perspective or an unusual approach to teaching a subject.

ULNP especially encourages new, original, and idiosyncratic approaches to physics teaching at the undergraduate level.

The purpose of ULNP is to provide intriguing, absorbing books that will continue to be the reader's preferred reference throughout their academic career.

Series editors

Neil Ashby
University of Colorado, Boulder, CO, USA

William Brantley
Department of Physics, Furman University, Greenville, SC, USA

Matthew Deady
Physics Program, Bard College, Annandale-on-Hudson, NY, USA

Michael Fowler
Department of Physics, University of Virginia, Charlottesville, VA, USA

Morten Hjorth-Jensen
Department of Physics, University of Oslo, Oslo, Norway

Michael Inglis
SUNY Suffolk County Community College, Long Island, NY, USA

More information about this series at http://www.springer.com/series/8917

Jakob Schwichtenberg

Physics from Symmetry

Second Edition

Springer

Jakob Schwichtenberg
Karlsruhe
Germany

NATURE ALWAYS CREATES THE BEST OF ALL OPTIONS

ARISTOTLE

AS FAR AS I SEE, ALL A PRIORI STATEMENTS IN PHYSICS HAVE THEIR
ORIGIN IN SYMMETRY.

HERMANN WEYL

THE IMPORTANT THING IN SCIENCE IS NOT SO MUCH TO OBTAIN NEW FACTS
AS TO DISCOVER NEW WAYS OF THINKING ABOUT THEM.

WILLIAM LAWRENCE BRAGG

Dedicated to my parents

Preface to the Second Edition

In the two years since the first edition of this book was published I've received numerous messages from readers all around the world. I was surprised by this large number of responses and how positive most of them were. Of course, I was cautiously confident that readers would like the book. Otherwise I wouldn't have spent so many months writing it. However, there is certainly no shortage of books on group theory or on the role of symmetries in physics. To quote Predrag Cvitanovic[1]

> Almost anybody whose research requires sustained use of group theory (and it is hard to think of a physical or mathematical problem that is wholly devoid of symmetry) writes a book about it.

Moreover, I'm not a world renowned expert. Therefore, I knew no one would buy the book because my name is written on the cover. So the chances were high that "Physics from Symmetry" would simply drown in the flood of new textbooks that are published every year.

Therefore, it's reasonable to wonder: Why and how did "Physics from Symmetry" avoid this fate?

I think the main reason for the success of the first edition is what I framed as something negative above: it wasn't written by a world renowned expert. I wrote the book while I was still a student and, as I remarked in the preface to the first edition, "I wrote the book I wished had existed when I started my journey in physics". So my motivation for writing the book wasn't to create an authoritative reference or a concise text that experts would love. Instead, my only focus was to write a book that helps students understand. As a student myself I always had still fresh in memory what I found confusing and what finally helped me understand.

This point of view is nicely summarized in the following quote by C.S. Lewis[2]

> It often happens that two schoolboys can solve difficulties in their work for one another better than the master can. When you took the problem

to a master, as we all remember, he was very likely to explain what you understood already, to add a great deal of information which you didn't want, and say nothing at all about the thing that was puzzling you. [...] The fellow-pupil can help more than the master because he knows less. The difficulty we want him to explain is one he has recently met. The expert met it so long ago he has forgotten. He sees the whole subject, by now, in a different light that he cannot conceive what is really troubling the pupil; he sees a dozen other difficulties which ought to be troubling him but aren't.

While most readers liked the student-friendly spirit of the first edition, it was, of course, not perfect. Several readers pointed out typos and paragraphs with confusing notation or explanations. This feedback guided me during the preparation of this second edition. I focused on correcting typos, improving the notation and I rewrote entire sections that were causing confusion.

I hope these changes make "Physics from Symmetry" even more student-friendly and useful[3].

Karlsruhe, September 2017 *Jakob Schwichtenberg*

[3] No book is ever perfect and I'm always happy to receive feedback. So if you find an error, have an idea for improvement or simply want to comment on something, always feel free to write me at `mail@jakobschwichtenberg.com`

Preface to the First Edition

The most incomprehensible thing about the world is that it is at all comprehensible.

- Albert Einstein[4]

In the course of studying physics I became, like any student of physics, familiar with many fundamental equations and their solutions, but I wasn't really able to see their connection.

I was thrilled when I understood that most of them have a common origin: **symmetry**. To me, the most beautiful thing in physics is when something incomprehensible, suddenly becomes comprehensible, because of a deep explanation. That's why I fell in love with symmetries.

For example, for quite some time I couldn't really understand spin, which is some kind of curious internal angular momentum that almost all fundamental particles carry. Then I learned that spin is a direct consequence of a symmetry, called Lorentz symmetry, and everything started to make sense.

Experiences like this were the motivation for this book and in some sense, I wrote the book I wished had existed when I started my journey in physics. Symmetries are beautiful explanations for many otherwise incomprehensible physical phenomena and this book is based on the idea that we can derive the fundamental theories of physics from symmetry.

One could say that this book's approach to physics starts at the end: Before we even talk about classical mechanics or non-relativistic quantum mechanics, we will use the (as far as we know) exact symmetries of nature to derive the fundamental equations of quantum field theory. Despite its unconventional approach, this book is about standard physics. We will not talk about speculative, experimentally unverified theories. We are going to use standard assumptions and develop standard theories.

Depending on the reader's experience in physics, the book can be used in two different ways:

- It can be used as a quick primer for those who are relatively new to physics. The starting points for classical mechanics, electrodynamics, quantum mechanics, special relativity and quantum field theory are explained and after reading, the reader can decide which topics are worth studying in more detail. There are many good books that cover every topic mentioned here in greater depth and at the end of each chapter some further reading recommendations are listed. If you feel you fit into this category, you are encouraged to start with the mathematical appendices at the end of the book[5] before going any further.

- Alternatively, this book can be used to connect loose ends for more experienced students. Many things that may seem arbitrary or a little wild when learnt for the first time using the usual historical approach, can be seen as being inevitable and straightforward when studied from the symmetry point of view.

In any case, you are encouraged to read this book from cover to cover, because the chapters build on one another.

We start with a short chapter about special relativity, which is the foundation for everything that follows. We will see that one of the most powerful constraints is that our theories must respect special relativity. The second part develops the mathematics required to utilize symmetry ideas in a physical context. Most of these mathematical tools come from a branch of mathematics called group theory. Afterwards, the Lagrangian formalism is introduced, which makes working with symmetries in a physical context straightforward. In the fifth and sixth chapters the basic equations of modern physics are derived using the two tools introduced earlier: The Lagrangian formalism and group theory. In the final part of this book these equations are put into action. Considering a particle theory we end up with quantum mechanics, considering a field theory we end up with quantum field theory. Then we look at the non-relativistic and classical limits of these theories, which leads us to classical mechanics and electrodynamics.

Every chapter begins with a brief summary of the chapter. If you catch yourself thinking: "Why exactly are we doing this?", return to the summary at the beginning of the chapter and take a look at how this specific step fits into the bigger picture of the chapter. Every page has a big margin, so you can scribble down your own notes and ideas while reading[6].

[5] Starting with Chapter A. In addition, the corresponding appendix chapters are mentioned when a new mathematical concept is used in the text.

[6] On many pages I included in the margin some further information or pictures.

I hope you enjoy reading this book as much as I have enjoyed writing it.

Karlsruhe, January 2015   *Jakob Schwichtenberg*

Acknowledgments

Contents

Part III The Equations of Nature

Part IV Applications

Part I
Foundations

1

Introduction

1.1   What we Cannot Derive

Before we talk about what we can derive from symmetry, let's clarify what we need to put into the theories by hand. First of all, there is presently no theory that is able to derive the constants of nature. These constants need to be extracted from experiments. Examples are the coupling constants of the various interactions and the masses of the elementary particles.

Besides that, there is something else we cannot explain: **The number three.** This should not be some kind of number mysticism, but we cannot explain all sorts of restrictions that are directly connected with the number three. For instance,

- there are three gauge theories[1], corresponding to the three fundamental forces described by the standard model: The electromagnetic, the weak and the strong force. These forces are described by gauge theories that correspond to the symmetry groups $U(1), SU(2)$ and $SU(3)$. Why is there no fundamental force following from $SU(4)$? Nobody knows!

- There are three lepton generations and three quark generations. Why isn't there a fourth? We only know from experiments with high accuracy that there is no fourth generation[2].

- We only include the three lowest orders in Φ in the Lagrangian (Φ^0, Φ^1, Φ^2), where Φ denotes here something generic that describes our physical system and the Lagrangian is the object we use to derive our theory from, in order to get a sensible theory describing free (=non-interacting) fields/particles.

- We only use the three lowest-dimensional representations of the double cover of the Poincaré group, which correspond to spin $0, \frac{1}{2}$

[1] Don't worry if you don't understand some terms, like gauge theory or double cover, in this introduction. All these terms will be explained in great detail later in this book and they are included here only for completeness.

[2] For example, the element abundance in the present universe depends on the number of generations. In addition, there are strong evidence from collider experiments. (See e.g. Phys. Rev. Lett. 109, 241802) .

and 1, respectively, to describe fundamental particles. There is no fundamental particle with spin $\frac{3}{2}$.

In the present theory, these things are assumptions we have to put in by hand. We know that they are correct from experiments, but there is presently no deeper principle why we have to stop after three.

In addition, there is one thing that can't be derived from symmetry, but which must be taken into account in order to get a sensible theory:

We are only allowed to include the lowest-possible, non-trivial order in the differential operator ∂_μ in the Lagrangian. For some theories these are first order derivatives ∂_μ, for other theories Lorentz invariance forbids first order derivatives and therefore second order derivatives $\partial_\mu \partial^\mu$ are the lowest-possible, non-trivial order. Otherwise, we don't get a sensible theory. Theories with higher order derivatives are unbounded from below, which means that the energy in such theories can be arbitrarily negative. Therefore states in such theories can always transition into lower energy states and are never stable.

Finally, there is another thing we cannot derive in the way we derive the other theories in this book: **gravity**. Of course there is a beautiful and correct theory of gravity, called general relativity. But this theory works quite differently than the other theories and a complete derivation lies beyond the scope of this book. Quantum gravity, as an attempt to fit gravity into the same scheme as the other theories, is still a theory under construction that no one has successfully derived. Nevertheless, some comments regarding gravity will be made in the last chapter.

1.2 Book Overview

Double Cover of the Poincaré Group

Irreducible Representations

$(0,0):$ Spin 0 Rep $\qquad$ $(\frac{1}{2},0) \oplus (0,\frac{1}{2}):$ Spin $\frac{1}{2}$ Rep $\qquad$ $(\frac{1}{2},\frac{1}{2}):$ Spin 1 Rep

acts on $\qquad$ acts on $\qquad$ acts on

Scalars $\qquad$ Spinors $\qquad$ Vectors

Constraint that Lagrangian is invariant $\qquad$ Constraint that Lagrangian is invariant $\qquad$ Constraint that Lagrangian is invariant

Free Spin 0 Lagrangian $\qquad$ Free Spin $\frac{1}{2}$ Lagrangian $\qquad$ Free Spin 1 Lagrangian

Euler-Lagrange equations $\qquad$ Euler-Lagrange equations $\qquad$ Euler-Lagrange equations

Klein-Gordon equation $\qquad$ Dirac equation $\qquad$ Proka equation

This book uses **natural units**, which means we set the Planck constant to $\hbar = 1$ and the speed of light to $c = 1$. This is helpful for theoretical considerations, because it avoids a lot of unnecessary writing. For applications the constants need to be added again to return to standard SI units.

Our starting point will be the basic assumptions of **special relativity**. These are: The velocity of light has the same value c in all inertial frames of reference, which are frames moving with constant velocity relative to each other and physics is the same in all inertial frames of reference.

The set of all transformations permitted by these symmetry constraints is called the **Poincaré group**. To be able to utilize them, we discuss the mathematical theory that enables us to work with symmetries. This branch of mathematics is called group theory. We will derive the irreducible representations of the Poincaré group[3], which you can think of as basic building blocks of all other representations. These representations are what we use later in this text to describe particles and fields of different spin. On the one hand, spin is an abstract label for different kinds of particles/fields and on the other hand can be seen as something like internal angular momentum. We will discuss in detail how this comes about.

Afterwards, the **Lagrangian formalism** is introduced, which makes working with symmetries in a physical context very conve-

[3] To be technically correct: We will derive the representations of the **double-cover** of the Poincaré group instead of the Poincaré group itself. The term "double-cover" comes from the observation that the map between the double-cover of a group and the group itself maps **two** elements of the double cover to **one** element of the group. This is explained in Section 3.3.1 in detail.

nient. The central object is the **Lagrangian**. Different Lagrangians describe different physical systems and we will derive several Lagrangians using symmetry considerations. In addition, the **Euler-Lagrange equations** are derived. These enable us to derive the equations of motion from a given Lagrangian. Using the irreducible representations of the Poincaré group, the fundamental equations of motion for fields and particles with different spin can be derived.

The central idea here is that the Lagrangian must be invariant (=does not change) under any transformation of the Poincaré group. This makes sure the equations of motion take the same form in all frames of reference, which we stated above as "physics is the same in all inertial frames".

Then, we will discover another symmetry of the Lagrangian for free spin $\frac{1}{2}$ fields: Invariance under $U(1)$ transformations. Similarly an **internal symmetry** for spin 1 fields can be found. Demanding **local** $U(1)$ symmetry will lead us to **coupling terms** between spin $\frac{1}{2}$ and spin 1 fields. The Lagrangian with this coupling term is the correct **Lagrangian for quantum electrodynamics**. A similar procedure for local $SU(2)$ and $SU(3)$ transformations will lead us to the correct **Lagrangian for weak and strong interactions**.

In addition, we discuss **symmetry breaking** and a special way to break symmetries called the **Higgs mechanism**. The Higgs mechanism enables us to describe particles with mass[4].

[4] Without the Higgs mechanism, terms describing mass in the Lagrangian spoil the symmetry and are therefore forbidden.

Afterwards, **Noether's theorem** is derived, which reveals a deep connection between symmetries and conserved quantities. We will utilize this connection by identifying each physical quantity with the corresponding symmetry generator. This leads us to the most important equation of quantum mechanics

$$[\hat{x}_i, \hat{p}_j] = i\delta_{ij} \tag{1.1}$$

and quantum field theory

$$[\hat{\Phi}(x), \hat{\pi}(y)] = i\delta(x - y). \tag{1.2}$$

[5] Non-relativistic means that everything moves slowly compared to the speed of light and therefore especially curious features of special relativity are too small to be measurable.

We continue by taking the non-relativistic[5] limit of the equation of motion for spin 0 particles, called Klein-Gordon equation, which result in the famous **Schrödinger equation**. This, together with the identifications we made between physical quantities and the generators of the corresponding symmetries, is the foundation of **quantum mechanics**.

Then we take a look at **free quantum field theory**, by starting

with the solutions of the different equations of motion[6] and Eq. 1.2. Afterwards, we take interactions into account, by taking a closer look at the Lagrangians with coupling terms between fields of different spin. This enables us to discuss how the **probability amplitude for scattering processes** can be derived.

By deriving the **Ehrenfest theorem** the connection between quantum and **classical mechanics** is revealed. Furthermore, the fundamental equations of **classical electrodynamics**, including the **Maxwell equations** and the **Lorentz force law**, are derived.

Finally, the basic structure of the modern theory of **gravity**, called **general relativity**, is briefly introduced and some remarks regarding the difficulties in the derivation of a **quantum theory of gravity** are made.

The major part of this book is about the tools we need to work with symmetries mathematically and about the derivation of what is commonly known as the **standard model**. The standard model uses quantum field theory to describe the behavior of all known elementary particles. Until the present day, all experimental predictions of the standard model have been correct. Every other theory introduced here can then be seen to follow from the standard model as a special case. For example in the limit of macroscopic objects we get classical mechanics or in the limit of elementary particles with low energy, we get quantum mechanics. For those readers who have never heard about the presently-known elementary particles and their interactions, a really quick overview is included in the next section.

1.3 Elementary Particles and Fundamental Forces

There are two major categories for elementary particles: **bosons** and **fermions**. There can be never two fermions in exactly the same state, which is known as **Pauli's exclusion principle**, but infinitely many bosons. This curious fact of nature leads to the completely different behavior of these particles:

- fermions are responsible for matter

- bosons for the forces of nature.

This means, for example, that atoms consist of fermions[7], but the electromagnetic-force is mediated by bosons. The bosons that are responsible for electromagnetic interactions are called photons. One of the most dramatic consequences of thie exclusion principle is that there is stable matter at all. If there could be infinitely many fermions in the same state, there would be no stable matter[8].

There are four presently known fundamental forces

- The electromagnetic force, which is mediated by massless **photons**.

- The weak force, which is mediated by massive $\mathbf{W}^+, \mathbf{W}^-$ and **Z-bosons**.

- The strong force, which is mediated by massless **gluons**.

- Gravity, which is (maybe) mediated by **gravitons**.

Some of the these bosons are massless and some are not and this tells us something deep about nature. We will fully understand this after setting up the appropriate framework. For the moment, just take note that each force is closely related to a symmetry. The fact that the bosons mediating the weak force are massive means the related symmetry is broken. This process of **spontaneous symmetry breaking** is responsible for the masses of **all** elementary particles. We will see later that this is possible through the coupling to another fundamental boson, the **Higgs boson**.

Fundamental particles interact via some force if they carry the corresponding **charge**[9].

- For the electromagnetic force this is the **electric charge** and consequently only electrically charged particles take part in electromagnetic interactions.

- For the weak force, the charge is called **isospin**[10]. All known fermions carry isospin and therefore interact via the weak force.

- The charge of the strong force is called **color**, because of some curious features it shares with the humanly visible colors. Don't let this name confuse you, because this charge has nothing to do with the colors you see in everyday life.

The fundamental fermions are divided into two subcategories: **quarks**, which are the building blocks of protons and neutrons, and **leptons**. Famous leptons are, for example, electrons and neutrinos. The difference is that quarks carry color charge and therefore take part in strong interactions, whereas leptons do not. There are three quark and lepton **generations**, which consist each of two particles:

[9] All charges have a beautiful common origin that will be discussed in Chapter 7.

[10] Often the charge of the weak force carries the extra prefix "weak", i.e. is called weak isospin, because there is another concept called isospin for composite objects that interact via the strong force. Nevertheless, this is not a fundamental charge and in this book the prefix "weak" is omitted.

	Generation 1	Generation 2	Generation 3	Electric charge	Isospin	Color
Quarks:	Up	Charm	Top	$\frac{+2}{3}e$	$\frac{1}{2}$	✓
	Down	Strange	Bottom	$\frac{-1}{3}e$	$\frac{-1}{2}$	✓
Leptons:	Electron-Neutrino	Muon-Neutrino	Tauon-Neutrino	0	$\frac{+1}{2}$	-
	Electron	Muon	Tauon	$-e$	$\frac{-1}{2}$	-

INTRODUCTION 9

Different particles can be identified through **labels**. In addition to the charges and the mass there is another incredibly important label called **spin**, which can be seen as some kind of internal angular momentum. Bosons carry integer spin, whereas fermions carry half-integer spin. The fundamental fermions we listed above have spin $\frac{1}{2}$ and almost all fundamental bosons have spin 1. There is only one known fundamental particle with spin 0: the Higgs boson.

There is an anti-particle for each particle, which carries exactly the same labels with opposite sign[11]. For the electron the anti-particle is called positron, but in general there is no extra name and only a prefix "anti". For example, the antiparticle corresponding to an up-quark is called anti-up-quark. Some particles, like the photon[12] are their own anti-particle.

All these notions will be explained in more detail later in this text. Now it's time to start with the derivation of the theory that describes correctly the interplay of the different characters in this particle zoo. The first cornerstone towards this goal is Einstein's famous theory of special relativity, which is the topic of the next chapter.

[11] Maybe except for the mass label. This is currently under experimental investigation, for example at the AEGIS, the ATRAP and the ALPHA experiment, located at CERN in Geneva, Switzerland.

[12] And maybe the neutrinos, which is currently under experimental investigation in many experiments that search for a neutrinoless double-beta decay.

2

Special Relativity

The famous Michelson-Morley experiment discovered that the speed of light has the same value in all reference frames[1]. Albert Einstein recognized the far reaching consequences of this observation and around this curious fact of nature he built the theory of special relativity. Starting from the constant speed of light, Einstein was able to predict many interesting and strange consequences that all proved to be true. We will see how powerful this idea is, but first let's clarify what special relativity is all about. The two basic postulates are

- **The principle of relativity:** Physics is the same in all inertial frames of reference, i.e. frames moving with constant velocity relative to each other.

- **The invariance of the speed of light:** The velocity of light has the same value c in all inertial frames of reference.

In addition, we will assume that the stage our physical laws act on is homogeneous and isotropic. This means it does not matter where (=homogeneity) we perform an experiment and how it is oriented (=isotropy), the laws of physics stay the same. For example, if two physicists, one in New-York and the other one in Tokyo, perform exactly the same experiment, they would find the same[2] physical laws. Equally a physicist on planet Mars would find the same physical laws.

The laws of physics, formulated correctly, shouldn't change if you look at the experiment from a different perspective or repeat it tomorrow. In addition, the first postulate tells us that a physical experiment should come up with the same result regardless of if you perform it on a wagon moving with constant speed or at rest in a laboratory. These things coincide with everyday experience. For example, if you close your eyes in a car moving with constant speed, there is no way to tell if you are really moving or if you're at rest.

[1] The speed of objects we observe in everyday life depend on the frame of reference. For example, when an observer standing at a train station measures that a train moves with $50 \frac{km}{h}$, another observer running with $15 \frac{km}{h}$ next to the same train, measures that the train moves with $35 \frac{km}{h}$. In contrast, light always moves with $1,08 \cdot 10^9 \frac{km}{h}$, no matter how you move relative to it.

[2] Besides from changing constants, as, for example, the gravitational acceleration

Without homogeneity and isotropy physics would be in deep trouble: If the laws of nature we deduce from experiment would hold only at one point in space, for a specific orientation of the experiment such laws would be rather useless.

The only unintuitive thing is the second postulate, which is contrary to all everyday experience. Nevertheless, all experiments until the present day show that it is correct.

2.1   The Invariant of Special Relativity

Before we dive into the details, here's a short summary of what we want to do in the following sections. We use the postulates of special relativity to derive the Minkowski metric, which tells us how to compute the "distance" between two physical events. Another name for physical events in this context is points in Minkowski space, which is how the stage the laws of special relativity act on is called. It then follows that all transformations connecting different inertial frames of reference must leave the Minkowski metric unchanged. This is how we are able to find all transformations that connect allowed frames of reference, i.e. frames with a constant speed of light. In the rest of the book we will use the knowledge of these transformations, to find equations that are unchanged by these transformations. Let's start with a thought experiment that enables us to derive one of the most fundamental consequences of the postulates of special relativity.

Imagine, we have a spectator, standing at the origin of his co-ordinate system and sending a light pulse straight up, where it is reflected by a mirror and finally reaches again the point from where it was sent. An illustration of this can be seen in Fig. 2.1

We have three important events:

- **A** : the light leaves the starting point

- **B** : the light is reflected at a mirror

- **C** : the light returns to the starting point.

The time-interval between **A** and **C** is[3]

$$\Delta t = t_C - t_A = \frac{2L}{c},\tag{2.1}$$

where L denotes the distance between the starting point and the mirror.

Next imagine a second spectator, standing at t_A at the origin of his coordinate system and moving with constant velocity u to the left,

Fig. 2.1: Illustration of the thought experiment

[3] For constant speed v we have $v = \frac{\Delta s}{\Delta t}$, with the distance covered Δs and the time needed Δt, and therefore $\Delta t = \frac{\Delta s}{v}$

relative to the first spectator[4]. For simplicity let's assume that the origin of this second spectators' coordinate system coincides at t_A with the coordinate origin of the first spectator. The second spectator sees things a little differently. In his frame of reference the point where the light ends up will not have the same coordinates as the starting point (see Fig. 2.2).

We can express this mathematically

$$x'_A = 0 \neq x'_C = u\Delta t' \qquad \rightarrow \qquad \Delta x' = u\Delta t', \qquad (2.2)$$

where the primed coordinates denote the moving spectator. For the first spectator in the rest-frame we have of course

$$x_A = x_C \qquad \rightarrow \qquad \Delta x = 0. \qquad (2.3)$$

We assume movement along the x-axis, therefore

$$y'_A = y'_C \quad \text{and} \quad z'_A = z'_C \qquad \rightarrow \qquad \Delta y' = 0 \quad \text{and} \quad \Delta z' = 0 \qquad (2.4)$$

and equally of course

$$y_A = y_C \quad \text{and} \quad z_A = z_C \qquad \rightarrow \qquad \Delta y = 0 \quad \text{and} \quad \Delta z = 0. \qquad (2.5)$$

   The next question is: **What about the time interval the second spectator measures?** Because we postulate a constant velocity of light, the second spectator measures a different time interval between **A** and **C**! The time interval $\Delta t' = t'_C - t'_A$ is equal to the distance l the light travels, as the second spectator observes it, divided by the speed of light c.

$$\Delta t' = \frac{l}{c} \qquad (2.6)$$

We can compute the distance traveled l using the good old Pythagorean theorem (see Fig. 2.2)

$$l = 2\sqrt{\left(\frac{1}{2}u\Delta t'\right)^2 + L^2}. \qquad (2.7)$$

Using Eq. 2.6, we therefore conclude

$$c\Delta t' = 2\sqrt{\left(\frac{1}{2}u\Delta t'\right)^2 + L^2}. \qquad (2.8)$$

If we now use $\Delta x' = u\Delta t'$ from Eq. 2.2, we can write

$$c\Delta t' = 2\sqrt{\left(\frac{1}{2}\Delta x'\right)^2 + L^2}$$

[4] Transformations that allow us to transform the description of one observer into the description of a second observer, moving with constant speed relative to first observer, are called **boosts**. We derive later a formal description of such transformations.
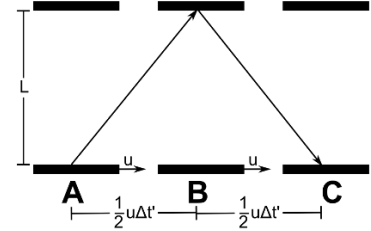
Fig. 2.2: Illustration of the thought experiment for a moving spectator. The second spectator moves to the left and therefore the first spectator (and the experiment) moves relative to him to the right.

$$\to (c\Delta t')^2 = 4\left(\left(\frac{1}{2}\Delta x'\right)^2 + L^2\right)$$

$$\to (c\Delta t')^2 - (\Delta x')^2 = 4\left(\left(\frac{1}{2}\Delta x'\right)^2 + L^2\right) - (\Delta x')^2 = 4L^2 \qquad (2.9)$$

and now recalling from Eq. 2.1 that $\Delta t = \frac{2L}{c}$, we can write

$$(c\Delta t')^2 - (\Delta x')^2 = 4L^2 = (c\Delta t)^2 = (\Delta tc)^2 - \underbrace{(\Delta x)^2}_{=0 \text{ see Eq. 2.3}}. \qquad (2.10)$$

So finally, we arrive[5] at

$$(c\Delta t')^2 - (\Delta x')^2 - \underbrace{(\Delta y')^2}_{=0} - \underbrace{(\Delta z')^2}_{=0} = (c\Delta t)^2 - \underbrace{(\Delta x)^2}_{=0} - \underbrace{(\Delta y)^2}_{=0} - \underbrace{(\Delta z)^2}_{=0}.$$
$$(2.11)$$

Considering a third observer, moving with a different velocity relative to the first observer, we can use the same reasoning to arrive at

$$(c\Delta t'')^2 - (\Delta x'')^2 - (\Delta y'')^2 - (\Delta z'')^2 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$$
$$(2.12)$$

Therefore, we have found something which is the same for all observers: the quadratic form

$$(\Delta s)^2 \equiv (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2. \qquad (2.13)$$

In addition, we learned in this section that $(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$ or $(c\Delta t)^2$ aren't the same for different observers. We will talk about the implications of this curious property in the next section.

2.2 Proper Time

We derived in the last section the invariant of special relativity Δs^2, i.e. a quantity that has the same value for all observers. Now, we want to think about the physical meaning of this quantity.

For brevity, let's restrict ourselves to one spatial dimension. An object at rest, relative to some observer, has a **spacetime diagram** as drawn in Fig. 2.3. In contrast, an object moving with constant velocity, relative to the same observer, has a spacetime diagram as drawn in Fig. 2.4.

The lines we draw to specify the position of objects in spacetime are called **world lines**. World lines are always observer dependent. Two different observers may draw completely different world lines for the same object. The moving object with world line drawn in

[5] Take note that what we are doing here is just the shortest path to the result, because we chose the origins of the two coordinate systems to coincide at t_A. Nevertheless, the same can be done, with more effort, for arbitrary choices, because physics is the same in all inertial frames. We used this freedom to choose two inertial frames where the computation is easy. In an arbitrarily moving second inertial system we do not have $\Delta y' = 0$ and $\Delta z' = 0$. Nevertheless, the equation holds, because physics is the same in all inertial frames.
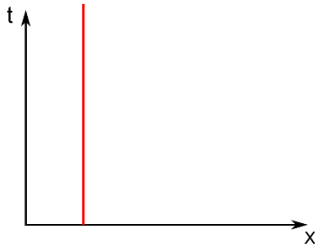
Fig. 2.3: World line of an object at rest. The position of the object stays the same as time goes on.
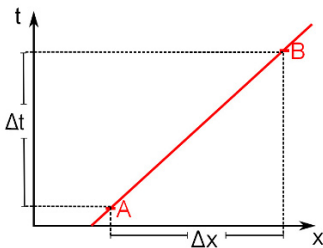
Fig. 2.4: World line of a moving object with two events A and B. The distance travelled between A and B is Δx and the time that passed the events is Δt.

Fig. 2.4, looks for a second observer who moves with the same constant speed as the object, as drawn in Fig. 2.5. For this second observer the object is at rest. Take note, to account for the two different descriptions we introduce primed coordinates for the second observer: x' and t'.

We can see that both observers do not agree on the distance the object travels between some events A and B in spacetime. For the first observer we have $\Delta x \neq 0$, but for the second observer $\Delta x' = 0$. For both observers the time interval between A and B is non-zero: $\Delta t \neq 0$ and $\Delta t' \neq 0$. Both observers agree on the value of the quantity $(\Delta s)^2$, because as we derived in the last section, this invariant of special relativity has the same value for all observers. A surprising consequence is that both observers do not agree on the time elapsed between the events A and B

$$(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 \tag{2.14}$$

$$(\Delta s')^2 = (c\Delta t')^2 - \underbrace{(\Delta x')^2}_{=0} = (c\Delta t')^2 \tag{2.15}$$

$$(\Delta s)^2 = (\Delta s')^2 \rightarrow (\Delta t')^2 \neq (\Delta t)^2 \quad \text{because } (\Delta x)^2 \neq 0. \tag{2.16}$$

This is one of the most famous phenomena of special relativity and commonly called **time-dilation**. Time-intervals and spatial distances are observer dependent. The clocks tick differently for different observers and therefore they observe a different number of ticks between two events.

Now that the concept of time has become relative, a new notion of time that all observers agree on may be useful. In the example above we can see that for the second observer, moving with the same speed as the object, we have

$$(\Delta s)^2 = (c\Delta t')^2. \tag{2.17}$$

This means the invariant of special relativity is equivalent, up to a constant c, to the time interval measured by this observer. With this in mind, we can interpret $(\Delta s)^2$ and define a notion of time that all observers agree on. We define

$$(\Delta s)^2 = (c\Delta \tau)^2, \tag{2.18}$$

where τ is called the **proper time**. The proper time is the time measured by an observer in the special frame of reference where the object in question is at rest.
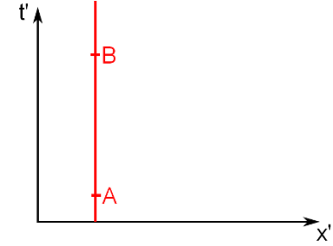
Fig. 2.5: World line of the same moving object, as observed from someone moving with the same constant speed as the object. The distance travelled between A and B is for this observer $\Delta x' = 0$.

Of course objects in the real world aren't restricted to motion with constant speed, but if the time interval is short enough, in the extremal case infinitesimal, any motion is linear and the notion of proper time is sensible. In mathematical terms this requires we make the transition to infinitesimal intervals $\Delta \to d$:

$$(ds)^2 = (cd\tau)^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2. \qquad (2.19)$$

Therefore, even if an object moves wildly, we can still imagine some observer with a clock travelling with the object and therefore observing the object at rest. The time interval this special observer measures is the proper time and all observers agree on its value, because $(ds)^2 = (cd\tau)^2$ has the same value for all observers. Again, this does not mean that all observers measure the same time interval! They just agree on the value of the time interval measured by someone who travels with the object in question.

2.3   Upper Speed Limit

Now that we have an interpretation for the invariant of special relativity, we can go a step further and explore one of the most stunning consequences of the postulates of special relativity.

It follows from the minus sign in the definition of Δs^2 that it can be zero for two events that are separated in space and time. It even can be negative, but then we would get a complex value for the proper time[6], which is commonly discarded as unphysical. We conclude, we have a minimal proper time $\tau = 0$ for two events if $\Delta s^2 = 0$. Then we can write

$$\Delta s_{\text{min}}^2 = 0 = (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2$$

$$\to (c\Delta t)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2$$

$$\to c^2 = \frac{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}{(\Delta t)^2}. \qquad (2.20)$$

On the right-hand side we have a squared velocity v^2, i.e. distance divided by time. We can rewrite this in the infinitesimal limit

$$\to c^2 = \frac{(dx)^2 + (dy)^2 + (dz)^2}{(dt)^2}. \qquad (2.21)$$

The functions $x(t), y(t), z(t)$ describe the path between the two events. Therefore, we have on the right-hand side the velocity between the events.

[6] Recall $(ds)^2 = (cd\tau)^2$ and therefore if $(ds)^2 < 0 \to d\tau$ is complex.

We conclude the lowest value for the proper time is measured by someone travelling with speed

$$\to c^2 = v^2. \tag{2.22}$$

This means nothing can move with a velocity larger than c! **We have an upper speed limit for everything in physics**. Two events in spacetime can't be connected by anything faster than c.

From this observation follows the **principle of locality**, which means that everything in physics can only be influenced by its immediate surroundings. Every interaction must be local and there can be no action at a distance, because everything in physics needs time to travel from some point to another.

2.4   The Minkowski Notation

> Henceforth space by itself, and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality.

> **- Hermann Minkowski**[7]

We can rewrite the invariant of special relativity

$$ds^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \tag{2.23}$$

by using a new notation, which looks quite complicated at first sight, but will prove to be invaluable:

$$ds^2 = \eta^{\mu\nu} dx_\mu dx_\nu = \eta^{00}(dx_0)^2 + \eta^{11}(dx_1)^2 + \eta^{22}(dx_2)^2 + \eta^{33}(dx_3)^2$$

$$= (dx_0)^2 - (dx_1)^2 - (dx_2)^2 - (dx_3)^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2. \tag{2.24}$$

Here we use several new notations and conventions one needs to become familiar with, because they are used everywhere in modern physics:

- Einsteins summation convention: If an index occurs twice, a sum is implicitly assumed : $\sum_{i=1}^{3} a_i b_i = a_i b_i = a_1 b_1 + a_2 b_2 + a_3 b_3$, but $\sum_{i=1}^{3} a_i b_j = a_1 b_j + a_2 b_j + a_3 b_j \neq a_i b_j$

- Greek indices[8], like μ, ν or σ, are always summed from 0 to 3: $x_\mu y_\mu = \sum_{\mu=0}^{3} x_\mu y_\mu$.

- Renaming of the variables $x_0 \equiv ct$, $x_1 \equiv x$, $x_2 \equiv y$ and $x_3 \equiv z$, to make it obvious that time and space are now treated equally and to be able to use the rules introduced above

$$^9 \eta = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

- Introduction of the Minkowski metric $\eta^{00} = 1$, $\eta^{11} = -1$, $\eta^{22} = -1$, $\eta^{33} = -1$ and $\eta^{\mu\nu} = 0$ for $\mu \neq \nu$ (an equal way of writing this is[9] $\eta = \mathrm{diag}(1, -1, -1, -1)$)

In addition, it's conventional to introduce the notion of a **four-vector**

$$dx_\mu = \begin{pmatrix} dx_0 \\ dx_1 \\ dx_2 \\ dx_3 \end{pmatrix}, \tag{2.25}$$

because the equation above can be written equally using four-vectors and the Minkowski metric in matrix form

$$(ds)^2 = dx_\mu \eta^{\mu\nu} dx_\nu = \begin{pmatrix} dx_0 & dx_1 & dx_2 & dx_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} dx_0 \\ dx_1 \\ dx_2 \\ dx_3 \end{pmatrix}$$

$$= dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 \tag{2.26}$$

This is really just a clever way of writing things. A physical interpretation of ds is that it is the "distance" between two events in spacetime. Take note that we don't mean here only the spatial distance, but also have to consider a separation in time. If we consider 3-dimensional Euclidean[10] space the squared (shortest) distance between two points is given by[11]

$$(ds)^2 = dx_i \delta_{ij} dx_j = \begin{pmatrix} dx_1 & dx_2 & dx_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix}$$

$$= (ds)^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2 \tag{2.27}$$

The mathematical tool that tells us the distance between two infinitesimal separated points is called **metric**. In boring Euclidean space the metric is just the identity matrix δ. In the curved spacetime of general relativity much more complicated metrics can occur. The geometry of the spacetime of special relativity is encoded in the relatively simple Minkowski metric η. Because the metric is the tool to compute length, we need it to define the **length of a four-vector**, which is given by the scalar product of the vector with itself[12]

$$x^2 = x \cdot x \equiv x_\mu x_\nu \eta^{\mu\nu}.$$

Analogously, the **scalar product** of two arbitrary four-vectors is

[10] 3-dimensional Euclidean space is just the space of classical physics, where time was treated differently from space and therefore it was not included into the geometric considerations. The notion of spacetime, with time as a fourth coordinate was introduced with special relativity, which enables mixing of time and space coordinates as we will see.

[11] The Kronecker delta δ_{ij}, which is the identity matrix in index notation, is defined in Appendix B.5.5.

[12] The same is true in Euclidean space: $\mathrm{length}^2(v) = \vec{v} \cdot \vec{v} = v_1^2 + v_2^2 + v_3^2$, because the metric is here simply

$$\delta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

defined by

$$x \cdot y \equiv x_\mu y_\nu \eta^{\mu\nu}. \tag{2.28}$$

There is another, notational convention to make computations more streamlined. We define a four-vector with upper index as[13]

$$x^\mu \equiv \eta^{\mu\nu} x_\nu \tag{2.29}$$

or equally

$$y^\nu \equiv \eta^{\mu\nu} y_\mu \underbrace{=}_{\text{The Minkowski metric is symmetric } \eta^{\mu\nu}=\eta^{\nu\mu}} \eta^{\nu\mu} y_\mu \,. \tag{2.30}$$

Therefore, we can write the scalar product of two four-vectors as[14]

$$x \cdot y \equiv x_\mu y_\nu \eta^{\mu\nu} = x_\mu y^\mu = x^\nu y_\nu. \tag{2.31}$$

It doesn't matter which index we transform to an upper index. This is just a way of avoiding writing the Minkowski metric all the time, just as Einstein's summation convention is introduced to avoid writing the summation sign.

2.5   Lorentz Transformations

Next, we try to figure out in what ways we can transform our description in a given frame of reference without violating the postulates of special-relativity. We learned above that it follows directly from the two postulates that $ds^2 = \eta^{\mu\nu} dx_\mu dx_\nu$ is the same in all inertial frames of reference:

$$ds'^2 = dx'_\mu dx'_\nu \eta^{\mu\nu} = ds^2 = dx_\mu dx_\nu \eta^{\mu\nu}. \tag{2.32}$$

Therefore, allowed transformations are those which leave this quadratic form or equally the scalar product of Minkowski spacetime invariant. Denoting a generic transformation that transforms the description in one frame of reference into the description in another frame with Λ, the transformed coordinates dx'_μ can be written as:

$$dx_\mu \to dx'_\mu = \Lambda_\mu{}^\sigma dx_\sigma. \tag{2.33}$$

Then we can write the invariance condition as

$$(ds)^2 = (ds')^2$$

$$\rightarrow dx \cdot dx \stackrel{!}{=} dx' \cdot dx'$$

$$\rightarrow dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} dx'_\mu dx'_\nu \eta^{\mu\nu} \underbrace{=}_{\text{Eq. 2.33}} \Lambda_\mu{}^\sigma dx_\sigma \Lambda_\nu{}^\gamma dx_\gamma \eta^{\mu\nu}$$

$$\underbrace{\rightarrow}_{\text{Renaming dummy indices}} dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} \Lambda_\sigma{}^\mu dx_\mu \Lambda_\gamma{}^\nu dx_\nu \eta^{\sigma\gamma}$$

$$\underbrace{\rightarrow}_{\text{Because the equation holds for arbitrary } dx_\mu} \eta^{\mu\nu} \stackrel{!}{=} \Lambda_\sigma{}^\mu \eta^{\sigma\gamma} \Lambda_\gamma{}^\nu . \tag{2.34}$$

Or written in matrix notation[15]

$$\eta = \Lambda^T \eta \Lambda \tag{2.35}$$

This is the condition that transformations Λ between allowed frames of reference must fulfil.

If this seems strange at this point don't worry, because we will see that such a condition is a quite natural thing. In the next chapter we will learn that, for example, rotations in ordinary Euclidean space are defined as those transformations O that leave the scalar product of Euclidean space invariant[16]

$$\vec{a} \cdot \vec{b} \stackrel{!}{=} \vec{a}' \cdot \vec{b}' \underbrace{=}_{\text{Take note that } (Oa)^T = a^T O^T} \vec{a}^T O^T O \vec{b}. \tag{2.36}$$

Therefore[17] $O^T 1 O \stackrel{!}{=} 1$ and we can see that the metric of Euclidean space, which is just the unit matrix 1, plays the same role as the Minkowski metric in Eq. 2.35. This is one part of the definition for rotations, because the defining feature of rotations is that they leave the length of a vector unchanged, which corresponds mathematically to the invariance of the scalar product[18]. Additionally we must include that rotations do not change the orientation[19] of our coordinate system, which means mathematically $\det O \stackrel{!}{=} 1$, because there are other transformations which leave the length of any vector invariant: spatial inversions[20]

We define the **Lorentz transformations** as those transformations that leave the scalar product of Minkowski spacetime invariant. In physical terms this means that Lorentz transformations describe changes between frames of references that respect the postulates of special relativity. In turn this does mean, of course, that everytime we want to get a term that does not change under Lorentz transformations, we must combine an upper with a lower index:

[15] If you wonder about the transpose here have a look at Appendix C.1.

[16] The $\cdot$ is used for the scalar product of vectors, which corresponds to $\vec{a} \cdot \vec{b} = \vec{a}^T \vec{b}$ for ordinary matrix multiplication, where a vector is an 1×3 matrix. The fact $(Oa)^T = a^T O^T$ is explained in Appendix C.1, specifically Eq. C.3.

[17] This condition is often called **orthogonality**, hence the symbol O. A matrix satisfying $O^T O = 1$ is called orthogonal, because its columns are orthogonal to each other. In other words: Each column of a matrix can be thought of as a vector and the orthogonality condition for matrices means that each such vector is orthogonal to all other column vectors.

[18] Recall that the length of a vector is given by the scalar product of a vector with itself.

[19] This is explained in Appendix A.5.

[20] A spatial inversion is simply a map $\vec{x} \rightarrow -\vec{x}$. Mathematically such transformations are characterized by the conditions $\det O \stackrel{!}{=} -1$ and $O^T O = 1$. Therefore, if we only want to talk about rotations we have the extra condition $\det O \stackrel{!}{=} 1$. Another name for spatial inversions are parity transformations.

$x_\mu y^\mu = x_\mu y_\nu \eta^{\mu\nu}$. We will construct explicit matrices for the allowed transformations in the next chapter, after we have learned some very elegant techniques for dealing with conditions like this.

2.6   Invariance, Symmetry and Covariance

Before we move on, we have to talk about some very important notions. Firstly, we call something **invariant**, if it does not change under transformations. For instance, let's consider something arbitrary like $F = F(A, B, C, ...)$ that depends on different quantities $A, B, C, ...$ If we transform $A, B, C, ... \to A', B', C', ...$ and we have

$$F(A', B', C', ...) = F(A, B, C, ...) \tag{2.37}$$

F is called invariant under this transformation. We can express this differently using the word symmetry. **Symmetry** is defined as invariance under a transformation or class of transformations. For example, some physical system is symmetric under rotations if we can rotate it arbitrarily and it always stays **exactly** the same. Another example would be a room with constant temperature. The quantity temperature does not depend on the position of measurement. In other words, the quantity temperature is invariant under translations. A translation means that we move every point a given distance in a specified direction. Therefore, we have translational symmetry within this room.

Covariance means something similar, but may not be confused with invariance. An equation is called covariant, if it takes the same **form** when the objects in it are transformed. For instance, if we have an equation

$$E_1 = aA^2 + bBA + cC^4$$

and after the transformation this equation reads

$$E_1' = aA'^2 + bB'A' + cC'^4$$

the equation is called covariant, because the form stayed the same. Another equation

$$E_2 = x^2 + 4axy + z$$

that after a transformation looks like

$$E_2' = y'^3 + 4az'y' + y'^2 + 8z'x'$$

is not covariant, because it changed its form completely.

All physical laws must be covariant under Lorentz transformations, because only such laws are valid in all reference frames. Formulating the laws of physics in a non-covariant way would be a very bad idea, because such laws would only hold in **one** frame of reference. The laws of physics would look differently in Tokyo and New York. There is no preferred frame of reference and we therefore want our laws to hold in all reference frames. We will learn later how we can formulate the laws of physics in a covariant manner.

Further Reading Tips

- **E. Taylor and J. Wheeler - Spacetime Physics: Introduction to Special Relativity**[21] is a very good book to start with.

- **D. Fleisch - A Student's Guide to Vectors and Tensors**[22] has very creative explanations for the tensor formalism used in special relativity, for example, for the differences between covariant and contravariant components.

- **N. Jeevanjee - An Introduction to Tensors and Group Theory for Physicists**[23] is another good source for the mathematics needed in special relativity.

- **A. Zee - Einstein Gravity in a nutshell**[24] is a book about general relativity, but has many great explanations regarding special relativity, too.

[21] Edwin F. Taylor and John Archibald Wheeler. *Spacetime Physics*. W. H. Freeman, 2nd edition, 3 1992. ISBN 9780716723271

[22] Daniel Fleisch. *A Student's Guide to Vectors and Tensors*. Cambridge University Press, 1st edition, 11 2011. ISBN 9780521171908

[23] Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

[24] Anthony Zee. *Einstein Gravity in a Nutshell*. Princeton University Press, 1st edition, 5 2013. ISBN 9780691145587

Part II
Symmetry Tools

3

Lie Group Theory

Chapter Overview

The final goal of this chapter is the derivation of the **fundamental representations of the double cover of the Poincaré group**, which we assume is the fundamental symmetry group of spacetime. These fundamental representations are the tools needed to describe all elementary particles, each representation for a different kind of elementary particle. The representations will tell us what types of elementary particles exist in nature.

We start with the definition of a **group**, which is motivated by two easy examples. Then, as a first step towards Lie theory we introduce two ways for describing rotations in two dimensions:

- 2×2 rotation matrix and

- unit complex numbers.

Then we will try to find a similar second description of rotations in three dimensions. This leads us to an extremely important group, called[1] **SU(2)**. After that, we will learn about **Lie algebras**, which enable us to learn a lot about something difficult (a Lie group) by using something simpler (the corresponding Lie algebra). There are in general many groups with the same Lie algebra, but only one of them is truly fundamental. We will use this knowledge to reveal the true fundamental symmetry group of nature, which doubly covers the Poincaré group. We usually start with some known transformations, derive the Lie algebra and use this Lie algebra to get different representations of the symmetry transformations. This will enable us to see that the representation we started with is just one special case out of many. This knowledge can then be used to learn something fundamental about the **Lorentz group**, which is an important part of the

2D Rotations

$U(1)$ $SO(2)$

3D Rotations

$SU(2)$ $SO(3)$

Lorentz Transformations

Lie algebra $\hat{=} \mathfrak{su}(2) \oplus \mathfrak{su}(2)$

Representations of the Double Cover

Lorentz Transformations + Translations
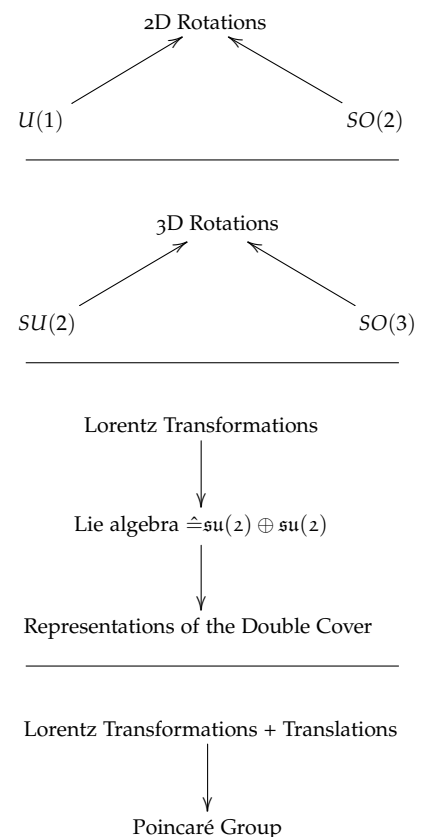
Poincaré Group

[1] The S stands for special, which means $\det(M) = 1$. U stands for unitary: $M^\dagger M = 1$ and the number 2 is used because the group is defined in the first place by 2×2 matrices.

Poincaré group. We will see that the Lie algebra of the double cover of the Lorentz group consists of two copies of the $SU(2)$ Lie algebra. Therefore, we can directly use everything we learned about $SU(2)$. Finally, we include translations into the considerations, which leads us to the Poincaré group. The Poincaré group is the Lorentz group plus translations. At this point, we will have everything at hand to classify the fundamental representations of the double cover of the Poincaré group. We use these fundamental representations in later chapters to derive the fundamental laws of physics.

3.1 Groups

If we want to utilize the power of symmetry, we need a framework to deal with symmetries mathematically. The branch of mathematics that deals with symmetries is called **group theory**. A special branch of group theory that deals with continuous symmetries is **Lie Theory**.

Symmetry is defined as invariance under a set of of transformations and therefore, one defines a group as a collection of transformations. Let us get started with two easy examples to get a feel for what we want to do:

Fig. 3.1: Illustration of a square

1. A square is mathematically a set of points (for example, the four corner points are part of this set) and a symmetry of the square is a transformation that maps this set of points into itself.

   Examples of symmetries of the square are rotations about the origin by 90°, 180°, 270° or 0°. These rotations map the square into itself. This means they map every point of the set to a point that lies again in the set and one says the set is invariant under such transformations.

   Take note that not every rotation is a symmetry of the square. This becomes obvious when we focus on the corner points of the square. Transforming the set by a clockwise rotation by, say 5°, maps these points into points outside the original set that defines the square. For example, as shown in Figure 3.2, the corner point A is mapped to the point A', which is not found inside the set that defined the square in the first place. Therefore a rotation by 5° is not a symmetry of the square. Of course the rotated object is still a square, but a different square (=different set of points). Nevertheless, a rotation by 90° is a symmetry of the square because the point A is mapped to the point B, which lies again in the original set. This is shown in Fig. 3.3.
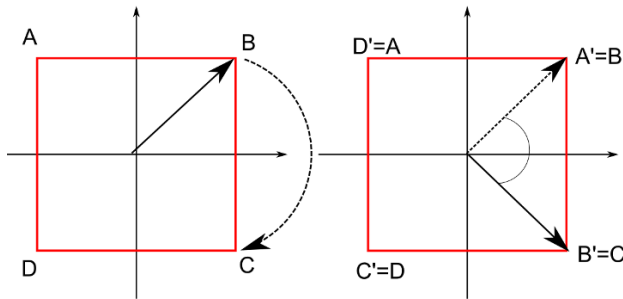
Fig. 3.2: Illustration of a square, rotated by 5°

Here's another helpful perspective: Imagine you close your eyes for a moment, and someone transforms a square in front of you. If you can't tell after opening your eyes whether the person changed anything at all, then the transformation was a symmetry transformation.

The set of transformations that leave the square invariant is called a group. The transformation parameter, here the rotation angle, can't take on arbitrary values and the group is called a discrete group.

2. Another example is the set of transformations that leave the unit circle invariant. Again, the unit circle is defined as a set of points and a symmetry transformation is a map that maps this set into itself.

The unit circle is invariant under all rotations about the origin, not just a few. In other words: the transformation parameter (the rotation angle) can take on arbitrary values, and the group is said to be a continuous group.

Fig. 3.4: Illustration of the rotation of the unit circle. For arbitrary rotations about the origin, all rotated points lie again in the initial set.

We are, of course, not only interested in symmetries of geometric shapes. For examples, considering vectors, we can look at the set of transformations that leave the length of any vector unchanged. For this reason, the definition of symmetry at the beginning of this chapter was very general: Symmetry means invariance under a transformation. Luckily, there is **one** mathematical theory, called group theory, that lets us work with **all** kinds of symmetries[2]

To make the idea of a mathematical theory that lets us deal with symmetries precise, we need to distill the defining features of symmetries into mathematical terms:

- Leaving the object in question unchanged ("doing nothing") is always a symmetry and therefore, every group needs to contain an identity element. In the examples above, the identity element is the rotation by 0°.

[2] As a side-note: Group theory was invented historically to describe symmetries of equations

- Transforming something and afterwards doing the inverse transformation must be equivalent to doing nothing. Therefore, there must be, for every element in the set, an inverse element. A transformation followed by its inverse transformation is, by definition of the inverse transformation, the same as the identity transformation. In the above examples this means that the inverse transformation of a rotation by 90° is a rotation by -90°. A rotation by 90° followed by a rotation by -90° is the same as a rotation by 0°.

- Performing a symmetry transformation followed by a second symmetry transformation is again a symmetry transformation. A rotation by 90° followed by a rotation by 180° is a rotation by 270°, which is a symmetry transformation, too. This property of the set of transformations is called closure.

³ But not commutative! For example rotations around different axes do not commute. This means in general: $R_x(\theta)R_z(\Phi) \neq R_z(\Phi)R_x(\theta)$

- The combination of transformations must be associative[3]. A rotation by 90° followed by a rotation by 40°, followed by a rotation by 110° is the same as a rotation by 130° followed by a rotation by 110°, which is the same as a rotation by 90° followed by a rotation by 150°. In a symbolic form:

$$R(110°)R(40°)R(90°) = R(110°)\big(R(40°)R(90°)\big) = R(110°)R(130°) \tag{3.1}$$

and

$$R(110°)R(40°)R(90°) = \big(R(110°)R(40°)\big)R(90°) = R(150°)R(90°) \tag{3.2}$$

This property is called **associativity**.

⁴ If you want to know more about the derivation of rotation matrices have a look at Appendix A.2.

⁵ For example, rotations in the plane can be described alternatively by multiplication with unit complex numbers. The rule for combining group elements is then complex number multiplication. This will be discussed later in this chapter.

⁶ Representation theory is the topic of Section 3.5.

- To be able to talk about the things above, one needs a rule, to be precise: a **binary operation**, for the combination of group elements. In the above examples, the standard approach would be to use rotation matrices[4] and the rule for combining the group elements (the corresponding rotation matrices) would be ordinary matrix multiplication. Nevertheless, there are often different ways to describe the same thing[5] and group theory enables us to study this systematically. The branch of group theory that deals with different descriptions of the same transformations is called **representation theory**[6].

To work with ideas like these in a rigorous, mathematical way, one distils the defining features of such transformations and promotes them to axioms. All structures satisfying these axioms are then called groups. This paves the way for a whole new branch of mathematics, called group theory. It is possible to find very abstract structures

satisfying the group axioms, but we will stick with groups that are very similar to the rotations we explored above.

After the discussion above, we can see that the abstract definition of a group simply states (obvious) properties of symmetry transformations:

A group $(G, \circ)$ is a set G, together with a binary operation $\circ$ defined on G, that satisfies the following axioms[7]

- Closure: For all $g_1, g_2 \in G$, $g_1 \circ g_2 \in G$

- Identity element: There exists an identity element $e \in G$ such that for all $g \in G$, $g \circ e = g = e \circ g$

- Inverse element: For each $g \in G$, there exists an inverse element $g^{-1} \in G$ such that $g \circ g^{-1} = e = g^{-1} \circ g$.

- Associativity: For all $g_1, g_2, g_3 \in G$, $g_1 \circ (g_2 \circ g_3) = (g_1 \circ g_2) \circ g_3$.

To summarize: The set of all transformations that leave a given object invariant is called a symmetry group. For Minkowski spacetime, the object that is left invariant is the Minkowski metric[8] and the corresponding symmetry group is called the Poincaré group.

Take note that the characteristic properties of a group are defined completely independent of the object the transformations act on. We can therefore study such symmetry transformations without making references to any object we extracted them from. This is very useful, because there can be many objects with the same symmetry or at least the same kind of symmetry. Using group theory we no longer have to inspect each object on its own, but are now able to study general properties of symmetry transformations.

3.2 Rotations in two Dimensions

As a first step, we start with an easy, but important, example. What transformations in two dimensions leave the length of any vector unchanged? After thinking about it for a while, we come up with[9] rotations and reflections. These transformations are of course the same ones that map the unit circle into the unit circle. This is an example of how one group may act on different kinds of objects: On the circle, which is a geometric shape, and on a vector. Considering vectors, one can represent these transformations by rotation matrices[10],

[7] Do not worry too much about this. In practice one checks for some kind of transformation if they obey these axioms. If they do, the transformations form a group and one can use the results of group theory to learn more about the transformations in question.

[8] Recall, this is the tool which we use to compute distances and lengths in Minkowski space.

[9] Another kind of transformation that leaves the length of a vector unchanged are translations, which means we move every point a constant distance in a specified direction. These are described mathematically a bit different and we are going to talk about them later.

[10] For an explicit derivation of these matrices have a look a look at Appendix A.2.

which are of the form

$$R_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \qquad (3.3)$$

and describe two-dimensional rotations about the origin by angle θ. Reflections at the axes can be performed using the matrices:

$$P_x = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad P_y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (3.4)$$

You can check that these matrices, together with the ordinary matrix multiplication as binary operation $\circ$, satisfy the group axioms and therefore these transformations form a group.

We can formulate the task of finding "all transformations in two dimensions that leave the length of any vector unchanged" in a more abstract way. The length of a vector is given by the scalar product of the vector with itself. If the length of the vector is the same after the transformation $a \to a'$, the equation

$$a' \cdot a' \overset{!}{=} a \cdot a \qquad (3.5)$$

must hold. We denote the transformation with O and write the transformed vector as $a \to a' = Oa$. Thus

$$a \cdot a = a^T a \to a'^T a' = (Oa)^T Oa = a^T O^T Oa \overset{!}{=} a^T a = a \cdot a, \qquad (3.6)$$

where we can see the condition a transformation must fulfil to leave the length of a vector unchanged is

$$O^T O = I, \qquad (3.7)$$

where I denotes the unit matrix[11]. You can check that the well-known rotation and reflection matrices we cited above fulfil exactly this condition[12]. This condition for two dimensional matrices defines the group $O(2)$, which is the group of all[13] orthogonal 2×2 matrices. We can find a subgroup of this group that includes only rotations, by observing that it follows from the condition in Eq. 3.7 that

$$\det(O^T O) \overset{!}{=} \det(I) = 1$$

$$\to \det(O^T O) = \det(O^T) \det(O) \overset{!}{=} \det(I) = 1$$

$$\to (\det(O))^2 \overset{!}{=} 1 \to \det(O) \overset{!}{=} \pm 1. \qquad (3.8)$$

The transformations of the group with $\det(O) = 1$ are rotations[14] and the two conditions

$$O^T O = I \qquad (3.9)$$

[11] $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

[12] With the matrix from Eq. 3.3 we have $R_\theta^T R = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 \\ 0 & \sin^2(\theta) + \cos^2(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \checkmark$

[13] Every orthogonal 2×2 matrix can be written either in the form of Eq. 3.3, as in Eq. 3.4, or as a product of these matrices.

[14] As can be easily seen by looking at the matrices in Eq. 3.3 and Eq. 3.4. The matrices with $\det O = -1$ are reflections.

$$\det O = 1 \tag{3.10}$$

define the $SO(2)$ group, where the "S" denotes special and the "O" orthogonal. The special thing about $SO(2)$ is that we now restrict it to transformations that keep the system orientation, i.e., a right-handed[15] coordinate system stays right-handed. In the language of linear algebra this means that the determinant of our matrices must be $+1$.

[15] If you don't know the difference between a right-handed and a left-handed coordinate system have a look at the Appendix A.5.

3.2.1   Rotations with Unit Complex Numbers

There is a different way to describe rotations in two dimensions that makes use of complex numbers: rotations about the origin by angle θ can be described by multiplication with a unit complex number ($z = a + ib$ which fulfils the condition[16] $|z|^2 = z^\star z = 1$).

[16] The $\star$ symbol denotes complex conjugation: $z = a + ib \to z^\star = a - ib$

The unit complex numbers together with ordinary complex number multiplication are a group, called[17] $U(1)$, as you can check by looking at the group axioms. To establish the connection with the group definitions for $O(2)$ and $SO(2)$ introduced above, we write the defining condition for unit numbers as[18]

$$U^\star U = 1. \tag{3.11}$$

Another way to write a unit complex number is[19]

$$R_\theta = e^{i\theta} = \cos(\theta) + i\sin(\theta), \tag{3.12}$$

because then

$$R_\theta^\star R_\theta = e^{-i\theta}e^{i\theta} = \big(\cos(\theta) - i\sin(\theta)\big)\big(\cos(\theta) + i\sin(\theta)\big) = 1 \tag{3.13}$$

Let's have a look at an example: we rotate the complex number $z = 3 + 5i$, by $90°$:

$$z \to z' = e^{i90°}z = \big(\underbrace{\cos(90°)}_{=0} + i\underbrace{\sin(90°)}_{=1}\big)(3 + 5i) = i(3 + 5i) = 3i - 5.$$
$$\tag{3.14}$$

The two complex numbers are plotted in Fig. 3.6 and we see the multiplication with $e^{i90°}$ does indeed rotate the complex number by $90°$. In this description, the rotation operator $e^{i90°}$ acts on complex numbers instead of on vectors. To describe a rotation in two dimensions, one parameter is necessary: the angle of rotation θ. A complex number has two degrees of freedom and with the constraint to unit complex numbers $|z| = 1$, one degree of freedom is left as needed.

[17] The U stands for unitary, which means the condition $U^\dagger U = 1$, where the symbol "†", called "dagger", denotes transposition plus complex conjugation: $U^\dagger = U^{T\star}$. For numbers, the dagger operation † is the same as taking only the complex conjugate, because every number satisfies trivially $z^T = z$. Hence the condition reduces here simply to $U^\star = U$.

[18] For more general information about the definition of groups involving a complex product, have a look at the appendix in Section 3.10.

[19] This is known as Euler's formula, which is derived in Appendix B.4.2. For a complex number $z = a + ib$, a is called the real part of z: $Re(z) = a$ and b the imaginary part: $Im(z) = b$. In Euler's formula $\cos(\theta)$ is the real part, and $\sin(\theta)$ the imaginary part of R_θ.
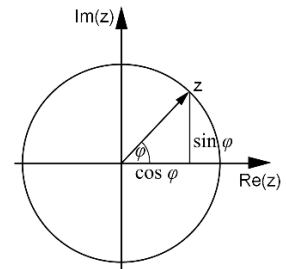
Fig. 3.5: The unit complex numbers lie on the unit circle in the complex plane.

We can connect this description to the previous one, where we used rotation matrices, by representing complex numbers by real 2×2 matrices. We define

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad i = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{3.15}$$

You can check that these matrices fulfil

$$1^2 = 1, \qquad i^2 = -1, \qquad 1i = i1 = i. \tag{3.16}$$

So now, the complex representation of rotations of the plane reads

$$R_\theta = \cos(\theta) + i\sin(\theta) = \cos(\theta) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin(\theta) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \tag{3.17}$$

By making the map

$$i \to \text{real matrix,}$$

we go back to the familiar representation of rotations of the plane. Maybe you have noticed a subtle point: The familiar rotation matrix in 2-dimensions acts on vectors, but here we identified the complex unit i with a real matrix (Eq. 3.15). Therefore, the rotation matrix will act on a 2×2 matrix, because the complex number we act on becomes a matrix, too.

A generic complex number in this description reads

$$z = a + ib = a \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}. \tag{3.18}$$

Let us take a look at how rotations act on such a matrix that represents a complex number:

$$z' = \begin{pmatrix} a' & -b' \\ b' & a' \end{pmatrix} = R_\theta z = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

$$= \begin{pmatrix} \cos(\theta)a - \sin(\theta)b & -\cos(\theta)b - \sin(\theta)a \\ \sin(\theta)a + \cos(\theta)b & -\sin(\theta)b + \cos(\theta)a \end{pmatrix}. \tag{3.19}$$

By comparing the left-hand side with the right-hand side, we get

$$\to a' = \cos(\theta)a - \sin(\theta)b$$
$$\to b' = \sin(\theta)a + \cos(\theta)b\,, \tag{3.20}$$

Fig. 3.6: Rotation of a complex number, by multiplication with a unit complex number

which is the same result that we get when we act with R_θ on a column vector

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \cos(\theta)a - \sin(\theta)b \\ \sin(\theta)a + \cos(\theta)b \end{pmatrix} = \begin{pmatrix} a' \\ b' \end{pmatrix}. \quad (3.21)$$

We see that both representations do exactly the same thing and mathematically speaking we have an isomorphism[20] between $SO(2)$ and $U(1)$. This is a very important discovery and we will elaborate on such lines of thought in the following chapters.

[20] An isomorphism is a one-to-one map Π that preserves the product structure $\Pi(g_1)\Pi(g_2) = \Pi(g_1 g_2) \; \forall g_1, g_2 \in G.$

Next we want to describe rotations in three dimensions and find similarly two descriptions for rotations in three dimensions[21].

[21] Things are about to get really interesting! Analogous to the two-dimensional case we discussed in the preceding section, we will find a second description of rotations in three dimensions and this alternative description will reveal something fundamental about nature.

3.3   Rotations in three Dimensions

The standard method to rotate vectors in three dimensions is to use 3×3 rotation matrices. The "basis rotations" around the three axes can be described by the following matrices:

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \qquad R_y = \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}$$

$$R_z = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (3.22)$$

If we want to rotate the vector

$$\vec{v} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

around the z-axis[22], we multiply it with the corresponding rotation matrix

[22] A general, rotated vector is derived explicitly in Appendix A.2.

$$R_z(\theta)\vec{v} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ 0 \end{pmatrix}. \quad (3.23)$$

To get a second description for rotations in three dimensions, the first thing we have to do is find a generalisation of complex numbers in higher dimensions. A first guess may be to go from 2-dimensional complex numbers to 3-dimensional complex numbers, but it turns out that there are no 3-dimensional complex numbers. Instead, we can find 4-dimensional complex numbers, called quaternions. The

quaternions will prove to be the correct second tool to describe rotations in 3-dimensions and the fact that this tool is 4-dimensional reveals something deep about the universe. We could have anticipated this result, because to describe an arbitrary rotation in 3-dimensions, 3 parameters are needed. Four dimensional complex numbers, with the constraint to unit quaternions[23], have exactly 3 degrees of freedom.

3.3.1   Quaternions

The 4-dimensional complex numbers can be constructed analogous to the 2-dimensional complex numbers. Instead of just one complex "unit" we introduce three, named $\mathbf{i}, \mathbf{j}, \mathbf{k}$. These fulfil

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1. \tag{3.24}$$

Then a 4-dimensional complex number, called a quaternion, can be written as

$$q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}. \tag{3.25}$$

We now need multiplication rules for $ij =?$ etc., because products like this will occur when one multiplies two quaternions. The extra condition

$$\mathbf{ijk} = -1 \tag{3.26}$$

suffices to compute all needed relations, for example $\mathbf{ij=k}$ follows from multiplying Eq. 3.26 with $\mathbf{k}$:

$$\mathbf{ij}\underbrace{\mathbf{kk}}_{=-1} = -\mathbf{k} \rightarrow \mathbf{ij=k}. \tag{3.27}$$

The set of *unit* quaternions $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ satisfy the condition[24]

[24] The symbol †, here is called "*dagger*" and denotes transposition plus complex conjugation: $a^\dagger = (a^\star)^T$. The ordinary scalar product always includes a transposition $a \cdot b = a^T b$, because matrix multiplication requires that we multiply a row with a column. In addition, for complex entities we include complex conjugation that makes sure we get something real, which is important if we want to interpret things in terms of length.

$$q^\dagger q \overset{!}{=} 1$$
$$\rightarrow (a\mathbf{1} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k})(a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = a^2 + b^2 + c^2 + d^2 \overset{!}{=} 1. \tag{3.28}$$

Exactly as the unit complex numbers formed a group under complex number multiplication, the unit quaternions form a group under quaternion multiplication.

Analogous to what we did for two-dimensional complex numbers, we now represent each of the three complex units with a matrix. There are different ways of doing this, but one choice that does the job is as complex 2×2 matrices:

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

$$\mathbf{j} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \quad , \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \tag{3.29}$$

You can check that these matrices fulfil the defining conditions in Eq. 3.24 and Eq. 3.26. Using these matrices a generic quaternion can then be written

$$q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = \begin{pmatrix} a + di & b + ci \\ -b + ci & a - di \end{pmatrix}. \tag{3.30}$$

Furthermore, we have

$$\det(q) = a^2 + b^2 + c^2 + d^2. \tag{3.31}$$

Comparing this with Eq. 3.28 tells us that the set of unit quaternions is given by matrices of the above form with unit determinant. The unit quaternions, written as complex 2×2 matrices therefore fulfil the conditions

$$U^\dagger U = 1 \quad \text{and} \quad \det(U) = 1. \tag{3.32}$$

Take note that the way we define $SU(2)$ here, is analogously to how we defined $SO(2)$. The S denotes special, which means $\det(U) = 1$ and U stands for unitary, which means the property[25] $U^\dagger U = 1$. Through the map in Eq. 3.29, every unit quaternion can be identified with an element of $SU(2)$.

Now, how is $SU(2)$ and with it the unit quaternions related to rotations? Unfortunately, the map between $SU(2)$ and[26] $SO(3)$ is not as simple as the one between $U(1)$ and $SO(2)$.

In 2-dimensions the 2 parameters of a complex number $z = a + \mathbf{i}b$ could be easily identified with the two spatial axes, i.e. $v = x + \mathbf{i}y$. The restriction to unit complex numbers automatically makes sure that the resulting matrix preserves the length of any vector[27]

$$(Uz)^\star Uz = z^\star U^\star Uz = z^\star z.$$

The quaternions have 4 parameters, so the identification with the 3 coordinates of a three-dimensional vector is not obvious. When we define

$$v \equiv x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \tag{3.33}$$

and use the matrix representation of the quaternions (Eq. 3.29), we can compute

$$\det(v) = x^2 + y^2 + z^2. \tag{3.34}$$

[25] For some more information about this, have a look at the Appendix 3.10 at the end of this chapter.

[26] Recall that $SO(3)$ is the set of the usual rotation matrices acting on 3 dimensional vectors.

[27] Recall that here R is a unit complex number, because complex numbers can be rotated by multiplication with unit complex numbers. Therefore we have $U^\star U = 1$, which is the defining condition for unit complex numbers.

[28] This follows from the general rule $\det(BA) = \det(B)\det(A)$. Therefore, if B has determinant 1, the product matrix BA has the same determinant as A.

[29] We identify our spatial components x, y, z as components in the subspace $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$. If, as a result of a transformation, we end up with a coefficient that does not belong to this subspace, i.e. does not appear together with $\mathbf{i}$, $\mathbf{j}$ or $\mathbf{k}$, we can't interpret it.

[30] For a derivation have a look at Appendix B.4.2.

Therefore, if we want to consider transformations that preserve the length of the vector (x, y, z), we must use matrix transformations that preserve determinants. Therefore the restriction to **unit** quaternions means that we must restrict to matrices with **unit** determinant[28]. Everything may now seem straight forward, but there is a subtle point. A first guess would be that a unit quaternion u induces a rotation on v simply by multiplication. This is not the case, because the product of u and v may not belong to $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$. Therefore, the transformed vector can have a component we are not able to interpret[29]. Instead the transformation that does the job is given by

$$v' = qvq^{-1}. \tag{3.35}$$

It turns out that by making this identification unit quaternions can describe rotations in 3-dimensions.

Let's take a look at an explicit example: To make the connection to our example in two dimensions, we define u as a unit quaternion that only takes values in $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ and denote a general unit quaternion with

$$t = \cos(\theta) + \sin(\theta)u. \tag{3.36}$$

Using Eq. 3.33 a generic vector can be written

$$\vec{v} = (v_x, v_y, v_z)^T = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} \underbrace{=}_{\text{Eq. 3.30}} \begin{pmatrix} iv_z & v_x + iv_y \\ -v_x + iv_y & -iv_z \end{pmatrix}. \tag{3.37}$$

With the identifications made above, we want to rotate, as an example, a vector $\vec{v} = (1, 0, 0)^T$ around the z-axis. We will make a particular choice for the vector and for the quaternion representing the rotation and show that it works. We write, using quaternions in their matrix representation (Eq. 3.29)

$$\vec{v} = (1, 0, 0)^T \rightarrow v = 1\mathbf{i} + 0\mathbf{j} + 0\mathbf{k} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{3.38}$$

In addition, we use the following quaternion

$$R_z(\theta) = \cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{k} = \begin{pmatrix} \cos(\theta) + i\sin(\theta) & 0 \\ 0 & \cos(\theta) - i\sin(\theta) \end{pmatrix}, \tag{3.39}$$

and then calculate that it is the correct quaternion that describes a rotation around the z-axis. We can rewrite R_z using Euler's formula[30] $e^{ix} = \cos(x) + i\sin(x)$:

$$\Rightarrow R_z(\theta) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}. \tag{3.40}$$

Inverting the quaternion rotation matrix yields

$$R_z(\theta)^{-1} = \begin{pmatrix} \cos(\theta) - i\sin(\theta) & 0 \\ 0 & \cos(\theta) + i\sin(\theta) \end{pmatrix} = \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}. \tag{3.41}$$

Using Eq. 3.35 the rotated vector reads

$$v' = R_z(\theta)vR_z^{-1}(\theta) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & e^{i2\theta} \\ -e^{-2i\theta} & 0 \end{pmatrix} = \begin{pmatrix} 0 & \cos(2\theta) + i\sin(2\theta) \\ -\cos(2\theta) + i\sin(2\theta) & 0 \end{pmatrix}. \tag{3.42}$$

On the other hand, an arbitrary vector can be written in our quaternion notation (Eq. 3.37)

$$v' = \begin{pmatrix} iv'_z & v'_x + iv'_y \\ -v'_x + iv'_y & -iv'_z \end{pmatrix}, \tag{3.43}$$

which we now compare with Eq. 3.42. This yields

$$v'_x = \cos(2\theta) \quad , \quad v'_y = \sin(2\theta) \quad , \quad v'_z = 0. \tag{3.44}$$

Therefore, written again in the conventional vector notation

$$\rightarrow \vec{v}' = (\cos(2\theta), \sin(2\theta), 0)^T. \tag{3.45}$$

Our identifications do indeed induce rotations[31], but something needs our attention. We haven't rotated $\vec{v}$ by θ, but by 2θ. Therefore, we define $\phi \equiv 2\theta$, because then ϕ really represents the angle we rotate. Using this definition we rewrite Eq. 3.36, which yields

$$t = \cos(\frac{\phi}{2}) + \sin(\frac{\phi}{2})u. \tag{3.46}$$

We can now see that the identifications we made are not one-to-one, but rather we have two unit-quaternions describing the same rotation. For example[32]

$$t_{\phi=\pi} = u \qquad\qquad t_{\phi=3\pi} = -u$$

Vector Rotation by π

This is the reason $SU(2)$ is called the **double-cover** of $SO(3)$. It is always possible to go unambiguously from $SU(2)$ to $SO(3)$ but not

[31] See the example in Eq. 3.23 where we rotated the vector, using the conventional rotation matrix

[32] A rotation by π is the same as a rotation by $3\pi = 2\pi + \pi$ for ordinary vectors, because $2\pi = 360°$ is a full rotation. In other words: We can see that two quaternions u and $-u$ can be used to rotate a vector by π.

vice versa. One may think this is just a mathematical side-note, but we will understand later that groups which cover other groups are indeed more fundamental[33].

To be able to discover the group that covers a given group, we need to introduce the most important tool of Lie theory: Lie algebras. This is the topic of the next section.

Take note that the fact we had one quaternion parameter too many, may be interpreted as a hint towards relativity. One may argue that a more natural identification would have been, as in the two-dimensional case, $v = t\mathbf{1} + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. We see that pure mathematics pushes us towards the idea of using a 4th component and what could it be, if not time? If we now want to describe rotations in 4-dimensions, because we know that the universe we live in is 4-dimensional, we have two choices:

- We could search for even higher dimensional complex numbers or

- we could again try to work with quaternions.

From the last paragraph it may seem that quaternions have something to say about rotations in 4 dimensions, too. An arbitrary rotation in 4 dimensions is described by[34] 6 parameters. There is no 7-dimensional generalisation of complex numbers, which together with the constraint to unit objects would have 6 free parameters. However, **two** unit quaternions have exactly 6 free parameters. Therefore, maybe it's possible to describe a rotation in 4-dimensions by two quaternions? We will learn later that there is indeed a close connection between two copies of $SU(2)$ and rotations in four dimensions.

3.4   Lie Algebras

Lie theory is all about continuous symmetries. An example is the continuous symmetry of the unit circle we discussed at the beginning of this chapter. Continuous here means that there are infinitely many symmetry transformations, that can be parametrized continuously by one or several parameters. For example, the rotation angle ϕ in the circle example can be any value: $0.1°$, $0.11°$, $0.11991°$, ... In contrast, for discrete symmetries like a reflection symmetry there is no continuous transformation parameter.

An important observation in Lie theory is that for continuous symmetries there are elements of the group which are arbitrarily close to the identity transformation[35]. In contrast, for a discrete group there is no continuous transformation parameter and therefore no element arbitrarily close to the identity. Consider again the symmetries of a

square. A rotation by 0,000001°, which is very close to the identity transformation (= a rotation by 0°), is not in the set of symmetry transformations of the square. In contrast, a rotation by 0,000001° is a symmetry of the circle. The symmetry group of a circle is continuous, because the rotation parameter (the rotation angle) can take on arbitrary (continuous) values. Mathematically, with the identity denoted I, an element g close to the identity is denoted

$$g(\epsilon) = I + \epsilon X, \tag{3.47}$$

where ϵ is, as always in mathematics, a really, really small number and X is an object, called generator, we will talk about in a moment. Such small transformations, when acting on some object change barely anything. In the smallest possible case such transformations are called infinitesimal transformation. Nevertheless, repeating such an infinitesimal transformation often, results in a finite transformation. Think about rotations: many small rotations in one direction are equivalent to one big rotation in the same direction. Mathematically, we can write the idea of repeating a small transformation many times

$$h(\theta) = (I + \epsilon X)(I + \epsilon X)(I + \epsilon X)... = (I + \epsilon X)^k, \tag{3.48}$$

where k denotes how often we repeat the small transformation. If θ denotes some finite transformation parameter, e.g. 50° or so, and N is some really big number which makes sure we are close to the identity, we can write the element close to the identity as

$$g(\theta) = I + \frac{\theta}{N} X. \tag{3.49}$$

The transformations we want to consider are the smallest possible, which means N must be the biggest possible number, i.e. $N \to \infty$. To get a finite transformation from such a infinitesimal transformation, one has to repeat the infinitesimal transformation infinitely. Mathematically

$$h(\theta) = \lim_{N \to \infty} (I + \frac{\theta}{N} X)^N, \tag{3.50}$$

which is in the limit just the exponential function[36]

$$h(\theta) = \lim_{N \to \infty} (I + \frac{\theta}{N} X)^N = e^{\theta X}. \tag{3.51}$$

In some sense the object X generates the finite transformation h, which is why it's called the **generator**.

If we want to calculate the generator X of a given transformation, we can differentiate the above formula

[36] This is often used as a definition of the exponential function. A proof, showing the equivalence of this limit and the exponential series we derive in Appendix B.4.1, can be found in most books about analysis.

$$\frac{d}{d\theta}h(\theta) = \frac{d}{d\theta}e^{\theta X} = Xe^{\theta X}.$$ (3.52)

Thus, if we evaluate this formula at $\theta = 0$, we get

$$X = \frac{dh(\theta)}{d\theta}\Big|_{\theta=0}.$$ (3.53)

The idea behind such lines of thought is that one can learn a lot about a group by looking at the important part of the infinitesimal elements (denoted X above): the **generators**. This will be made more precise in a moment, but first let's look at this from another perspective that makes it even clearer in what sense the generators generate a finite transformation:

If we consider a continuous group of transformations that are given by matrices, we can make a Taylor expansion[37] of an element of the group about the identity. The Taylor series is given by

$$h(\theta) = I + \frac{dh}{d\theta}\cdot|_{\theta=0}\theta + \frac{1}{2}\frac{d^2h}{d\theta^2}\cdot|_{\theta=0}\theta^2 + ... = \sum_n \frac{1}{n!}\frac{d^nh}{d\theta^n}|_{\theta=0}\theta^n.$$ (3.54)

This shows nicely how the generators generate transformations.

For matrix Lie groups one defines the corresponding Lie algebra as the collection of objects that give an element of the group when exponentiated. This is an easy definition one can use when restricting to matrix Lie groups. Later we will introduce a more general definition. In mathematical terms[38]

[38] The Lie algebra which belongs to a group G is conventionally denoted by the corresponding "Fraktur" letter $\mathfrak{g}$

For a Lie Group G (given by $n \times n$ matrices), the Lie algebra $\mathfrak{g}$ of G is given by those $n \times n$ matrices X such that $e^{tX} \in G$ for $t \in \mathbb{R}$, together with an operation, called the Lie bracket $[,]$ that tells us how we can combine these matrices.

The last part of this definition can be a bit confusing and thus we now spent some time discussing it.

We know from the definition of a group, that a group is more than just a collection of transformations. The definition of a group includes a binary operation $\circ$ that tells us how to combine group elements. For matrix Lie groups this is just ordinary matrix multiplication. Naively one may think that the same combination rule $\circ$ is valid for elements of the Lie algebra, but this is not the case! The elements of the Lie algebra are given by matrices[39], but the multiplication of two Lie algebra elements doesn't need to be an element of the Lie algebra. Instead there is another combination rule for the Lie

[39] A famous theorem of Lie group theory, called Ado's Theorem, states that every Lie algebra is isomorphic to a matrix Lie algebra.

algebra, already mentioned in the definition above, that is directly connected to the combination rule of the corresponding Lie group.

The connection between the combination rule of the Lie group and the combination rule of the Lie algebra is given by the famous Baker-Campbell-Hausdorff formula[40]

$$e^X \circ e^Y = e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+...}$$ (3.55)

On the left hand side, we have the multiplication of two elements of the Lie group, let's name them g and h, which we can write in terms of the corresponding generators (=elements of the Lie algebra)

$$\underbrace{g}_{\in G} \circ \underbrace{h}_{\in G} = e^X \circ e^Y = \underbrace{e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+...}}_{\in G}$$ (3.56)

with the generators[41] $X, Y \in \mathfrak{g}$. On the right-hand side we have a single object of the group and the multiplication of the group elements have been translated to a sum of Lie algebra elements. The new symbol in this sum $[\ ,\ ]$ is called **Lie bracket** and for matrix Lie groups it is given by $[X,Y] = XY - YX$, which is called the **commutator** of X and Y. The elements XY and YX need not to be part of the Lie algebra, but their difference always is[42]!

We learn from the Baker-Campbell-Hausdorff-Formula that the natural product of the Lie algebra is, not as one would naively think, ordinary matrix multiplication, but the Lie bracket $[,]$. One says, the Lie algebra is **closed** under the Lie bracket, just as the group is closed under the corresponding composition rule $\circ$, e.g. matrix multiplication. Closure means that the composition of two elements lies again in the same set[43].

After looking at an example to illustrate these new notions, we will have a look at the modern definition of a Lie algebra. The main component of this definition is how the generators of a group behave when put into the Lie bracket. By using this general definition we will see that it is possible to say that different groups have the **same** Lie algebra. With the definition above saying something like this would make little sense. Nevertheless, this new way of thinking about Lie algebras will enable us to reveal the most fundamental description corresponding to a given transformation. This is possible because there is a theorem in Lie theory that tells us that there is exactly one **distinguished** Lie group for each Lie algebra. The Lie algebra however, according to the abstract definition, corresponds to many Lie groups. We will make this more concrete after introducing the modern definition of a Lie group.

[40] We will not talk about the proof of this formula in this book. Proofs can be found in most books about Lie theory, for example in William Fulton and Joe Harris. *Representation Theory: A First Course*. Springer, 1st corrected edition, 8 1999. ISBN 9780387974958

[41] The Lie algebra which belongs to a group G is conventionally denoted by the corresponding "Fraktur" letter $\mathfrak{g}$.

[42] A very illuminating proof of this fact can be found in John Stillwell. *Naive Lie Theory*. Springer, 1st edition, August 2008. ISBN 978-0387782140

[43] For group elements $g, h \in G$ we have $g \circ h \in G$. For elements of the Lie algebra $X, Y \in \mathfrak{g}$ we have $[X,Y] \in \mathfrak{g}$ and in general $X \circ Y \notin \mathfrak{g}$

Now we want to take a look at an explicit example of how we can derive the Lie algebra of a given group.

3.4.1   The Generators and Lie Algebra of $SO(3)$

The defining conditions of the group $SO(3)$ are (Eq. 3.10)

$$O^T O \overset{!}{=} I \quad \text{and} \quad \det(O) \overset{!}{=} 1. \tag{3.57}$$

We can write every group element O in terms of a generator J:

$$O = e^{\theta J}. \tag{3.58}$$

Putting this into the first defining condition yields

$$O^T O = e^{\theta J^T} e^{\theta J} \overset{!}{=} 1 \to J^T + J \overset{!}{=} 0. \tag{3.59}$$

Using the second condition in Eq. 3.57 and the identity[44] $\det(e^A) = e^{\text{tr}(A)}$ for the matrix exponential, we see

$$(O) \overset{!}{=} 1 \quad \to \quad \det(e^{\theta J}) = e^{\theta \text{tr}(J)} \overset{!}{=} 1$$

$$\to \quad \text{tr}(J) \overset{!}{=} 0. \tag{3.60}$$

Three linearly independent[45] matrices fulfilling both conditions (Eq. 3.59, Eq. 3.60) are

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \qquad J_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \qquad J_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.61}$$

These matrices form a **basis** for the generators of the group $SO(3)$. This means any generator of the group can be written as a linear combination of these basis generators: $J = aJ_1 + bJ_2 + cJ_3$, where a, b, c denote real constants. These generators can be written more compactly by using the Levi-Civita symbol[46]

$$(J_i)_{jk} = -\epsilon_{ijk}, \tag{3.62}$$

where j, k denote the components of the generator J_i. For example,

$$(J_1)_{jk} = -\epsilon_{1jk} \leftrightarrow \begin{pmatrix} (J_1)_{11} & (J_1)_{12} & (J_1)_{13} \\ (J_1)_{21} & (J_1)_{22} & (J_1)_{23} \\ (J_1)_{31} & (J_1)_{32} & (J_1)_{33} \end{pmatrix} = - \begin{pmatrix} \epsilon_{111} & \epsilon_{112} & \epsilon_{113} \\ \epsilon_{121} & \epsilon_{122} & \epsilon_{123} \\ \epsilon_{131} & \epsilon_{132} & \epsilon_{133} \end{pmatrix}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.63}$$

[44] $\text{tr}(A)$ denotes the trace of the matrix A, which means the sum of all elements on the main diagonal. For example for $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, we have $\text{tr}(A) = A_{11} + A_{22}$.

[45] This is explained in Appendix A.1.

[46] The Levi-Civita symbol is explained in Appendix B.5.5.

Let's see what finite transformation matrix we get from the first of these basis generators. The connection between the generators and the finite transformations is given by the exponential function: $R_{1 \text{ fin}} = e^{\theta J_1}$. To calculate this, we can focus on the lower right 2×2 matrix[47] j_1 in J_1 and ignore the zeroes for a moment:

$$J_1 = \begin{pmatrix} 0 & & \\ & \underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}_{\equiv j_1} \end{pmatrix}. \tag{3.64}$$

We can immediately compute

$$(j_1)^2 = -1, \tag{3.65}$$

therefore

$$(j_1)^3 = \underbrace{(j_1)^2}_{=-1} j_1 = -j_1, \quad (j_1)^4 = +1, \quad (j_1)^5 = +j_1. \tag{3.66}$$

In general, we have

$$(j_1)^{2n} = (-1)^n I \quad \text{and} \quad (j_1)^{2n+1} = (-1)^n j_1, \tag{3.67}$$

which we can use if we evaluate the exponential function as series expansion[48]

$$\begin{aligned}
\tilde{R}_{1 \text{ fin}} = e^{\theta j_1} &= \sum_{n=0}^{\infty} \frac{\theta^n j_1^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{\theta^{2n}}{(2n)!} \underbrace{(j_1)^{2n}}_{(-1)^n I} + \sum_{n=0}^{\infty} \frac{\theta^{2n+1}}{(2n+1)!} \underbrace{(j_1)^{2n+1}}_{(-1)^n j_1} \\
&= \underbrace{\left(\sum_{n=0}^{\infty} \frac{\theta^{2n}}{(2n)!} (-1)^n \right)}_{=\cos(\theta)} I + \underbrace{\left(\sum_{n=0}^{\infty} \frac{\theta^{2n+1}}{(2n+1)!} (-1)^n \right)}_{=\sin(\theta)} j_1 \\
&= \cos(\theta) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin(\theta) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.
\end{aligned} \tag{3.68}$$

Using $e^0 = 1$ for the upper-left component, the complete, finite transformation matrix therefore reads

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}, \tag{3.69}$$

which we can recognize as one of the well-known rotation matrices in 3-dimensions that were quoted at the beginning of this chapter

[50] For example, we have
$[J_1, J_2] = J_1 J_2 - J_2 J_1$

$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} -$

$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} =$

$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} =$

$\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \underbrace{\epsilon_{12k} J_k}_{=0 \text{ except for } k=3}$

$= \epsilon_{123} J_3 = J_3$

[51] We will call the Lie bracket relation of the basis generators **the** Lie algebra, because everything important is encoded here.

[52] For example now we have $J_1^\star = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix}$ and therefore

$J_1^\dagger = (J_1^\star)^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} = J_1$

[53] This will be discussed in Section 8.3.

(Eq. 3.22). Following the same steps, we can derive the matrices for rotations around the other axes, too.

We now have the generators of the group in explicit matrix form (Eq. 3.61) and this allows us to compute the Lie bracket relations between the basis generators by brute force[49]. The result is[50]

$$[J_i, J_j] = \epsilon_{ijk} J_k, \tag{3.70}$$

where ϵ_{ijk} is again the Levi-Civita symbol. In physics it's conventional to define the generators of $SO(3)$ with an extra "i". Concretely this means that instead of $e^{\tilde\phi J}$, we write $e^{i\phi J}$ with $\phi = -\tilde\phi$. Our generators are then

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad J_2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \qquad J_3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\tag{3.71}$$

and the Lie algebra[51] reads

$$[J_i, J_j] = i\epsilon_{ijk} J_k. \tag{3.72}$$

We introduce the additional "i" in physics to get Hermitian generators, which means generators that satisfy[52] $J^\dagger = (J^\star)^T = J$. This is a nice property, because Hermitian matrices have *real* eigenvalues and this becomes important in quantum mechanics where the eigenvalues of the generators are the values we expect to measure in experiments[53]. Without the "i", the generators are anti-Hermitian $J^\dagger = (J^\star)^T = -J$ and the corresponding eigenvalues are imaginary. This would make it less intuitive to interpret the eigenvalues as something that we can observe in experiments.

We can derive the basis generators in another way, by starting with the well known rotation matrices and use Eq. 3.53: $X = \frac{dh}{d\theta}|_{\theta=0}$. For the first rotation matrix, as quoted in Eq. 3.22 and derived in Eq. 3.69, this yields

$$J_1 = \frac{dR_1}{d\theta}\Big|_{\theta=0} = \frac{d}{d\theta} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \Big|_{\theta=0}$$

$$= \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin(\theta) & -\cos(\theta) \\ 0 & \cos(\theta) & -\sin(\theta) \end{pmatrix} \Big|_{\theta=0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \tag{3.73}$$

which is exactly the first generator in Eq. 3.61. Nevertheless, the first method is more general, because we will not always start with

given finite transformation matrices. For the Lorentz group we will start with the definition of the group, derive the basis generators and only afterwards compute an explicit matrix form of the Lorentz transformations. If you already have explicit transformation matrices, you can always use Eq. 3.53 to derive the corresponding generators.

Before we discuss the Lorentz group in more detail, we take a small detour and have a look at the modern definition of a Lie algebra. This modern definition is essential to get a deep understanding of the symmetries of nature.

3.4.2 The Abstract Definition of a Lie Algebra

Up to this point we used a simplified definition: The Lie algebra consists of all elements X that result in an element of the corresponding group G, when put into the exponential function $e^X \in G$, and an operation, called Lie bracket $[,]$, that we use to combine the Lie algebra elements.

We already discussed that the last part of this definition is crucial. As for the group, we also need a rule to combine Lie algebra elements. By looking at the Baker-Campbell-Hausdorff-Formula, we learned how the rule for the combination of group elements, is connected to the rule for the combination of Lie algebra elements. An important observation was that the rule for the combination of Lie algebra elements is not simply matrix multiplication, but a more complicated rule called Lie bracket.

While this operation already appears in our simplified definition, we now introduce a more abstract definition where the Lie bracket is even more central[54]

A Lie algebra is a vector space $\mathfrak{g}$ equipped with a binary operation $[,]: \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$. The binary operation satisfies the following axioms:

- Bilinearity: $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ and $[Z, aX + bY] = a[Z, X] + b[Z, Y]$, for arbitrary number a, b and $\forall X, Y, Z \in \mathfrak{g}$

- Anticommutativity: $[X, Y] = -[Y, X] \ \forall \ X, Y \in \mathfrak{g}$

- The Jacobi Identity: $[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] = 0$ $\forall X, Y, Z \in \mathfrak{g}$

You can check that the commutator of two matrices fulfills all these conditions and of course this standard commutator was used to motivate these axioms. Nevertheless, there are quite different binary operations that fulfill these axioms, for example, the famous Poisson bracket of classical mechanics.

[54] This definition will prove to be invaluable for the following sections and it will become clear in a moment in what sense the Lie bracket is "more central".

The important point is that this definition makes no reference to any Lie group. The definition of a Lie algebra stands on its own and we will see that this makes a lot of sense. In the next section we will have a look at the generators of $SU(2)$ and find that the basis generators, which is the set of generators we can use to construct all other generators by linear combinations, fulfill the same Lie bracket relation as the basis generators of $SO(3)$ (Eq. 3.72). This is interpreted as $SU(2)$ and $SO(3)$ having the **same** Lie algebra. This is an incredibly important result and it will tell us a lot about $SU(2)$ and $SO(3)$.

3.4.3 The Generators and Lie Algebra of $SU(2)$

We stumbled upon $SU(2)$ while trying to describe rotations in three dimensions and discovered that $SU(2)$ is the double cover[55] of $SO(3)$.

Remember that $SU(2)$ is the group of unitary 2×2 matrices with unit determinant[56] :

$$U^\dagger U = U U^\dagger = 1 \tag{3.74}$$

$$\det(U) = 1. \tag{3.75}$$

The first thing we now want to take a look at is the Lie algebra of this group. Writing the defining conditions of the group in terms of the generators $J_1, J_2, \ldots$ yields[57]

$$U^\dagger U = (e^{iJ_i})^\dagger e^{iJ_i} \overset{!}{=} 1 \tag{3.76}$$

$$\det(U) = \det(e^{iJ_i}) \overset{!}{=} 1 \tag{3.77}$$

The first condition tells us, using the Baker-Campbell-Hausdorff theorem (Eq. 3.55) and $[J_i, J_i] = 0$

$$(e^{iJ_i})^\dagger e^{iJ_i} = e^{-iJ_i^\dagger} e^{iJ_i} \overset{!}{=} 1$$

$$\rightarrow e^{-iJ_i^\dagger + iJ_i + \frac{1}{2}[J_i^\dagger, J_i] + \cdots} \overset{!}{=} 1$$

$$\underset{e^0 = 1}{\rightarrow} J_i^\dagger \overset{!}{=} J_i. \tag{3.78}$$

A matrix fulfilling the condition $J_i^\dagger = J_i$ is called Hermitian and we therefore learn here that the generators of $SU(2)$ must be Hermitian.

Using the identity $\det(e^A) = e^{\text{tr}(A)}$, we see that the second condition yields

$$\det(e^{iJ_i}) = e^{i\text{tr}(J_i)} = 1 \underset{e^0 = 1}{\rightarrow} \text{tr}(J_i) \overset{!}{=} 0. \tag{3.79}$$

We conclude that the generators of $SU(2)$ must be Hermitian *traceless* matrices. A basis for Hermitian traceless 2×2 matrices is given by the following 3 matrices[58]:

[58] A complex 2×2 matrix has 4 complex entries and therefore 8 degrees of freedom. Because of the two conditions the degrees of freedom are reduced to 3.

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \qquad (3.80)$$

This means every Hermitian traceless 2×2 matrix can be written as a linear combination of these matrices that are called **Pauli matrices**.

We can put these explicit matrices for the basis generators into the Lie bracket and this yields

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k, \qquad (3.81)$$

where ϵ_{ijk} is again the Levi-Civita symbol. To get rid of the nasty 2 it is conventional to define the generators of $SU(2)$ as $J_i \equiv \frac{1}{2}\sigma_i$. The Lie algebra then reads

$$[J_i, J_j] = i\epsilon_{ijk}J_k. \qquad (3.82)$$

Take note that this is exactly the same Lie bracket relation we derived for $SO(3)$ (Eq. 3.72)! Therefore one says that $SU(2)$ and $SO(3)$ have the same Lie algebra, because we define Lie algebras by their Lie bracket. We will use the abstract definition of this Lie algebra, to get different descriptions for the transformations described by $SU(2)$. We will learn that an $SU(2)$ transformation doesn't need to be described by 2×2 matrices. To make sense of things like this, we need a more abstract definition of a Lie group. At this point $SU(2)$ is defined as a set of 2×2 matrices, and a description of $SU(2)$ by, for example, 3×3 matrices, makes little sense. The abstract definition of a Lie group will enable us to see the connection between different descriptions of the same transformation. We will identify with each Lie group a geometrical object (a manifold) and use this abstract object to define a group. This may seem like a strange thought, but will make a lot of sense after taking a second look at two examples we already encountered in earlier sections.

3.4.4   The Abstract Definition of a Lie Group

One of the first Lie groups we discussed was $U(1)$, the unit complex numbers. These are defined as complex numbers that satisfy $z^\star z = 1$. If we write $z = a + ib$ this condition reads

$$z^\star z = (a + ib)^\star (a + ib) = (a - ib)(a + ib) = a^2 + b^2 = 1. \qquad (3.83)$$

This is exactly the defining condition of the unit circle[59]. The set

[59] The unit circle S^1 is the set of all points in two dimensions with distance 1 from the origin. In mathematical terms this means all points (x_1, x_2) fulfilling $x_1^2 + x_2^2 = 1$.

of unit-complex numbers **is** the unit circle in the complex plane. Furthermore, we found that there is a one-to-one map[60] between elements of $U(1)$ and $SO(2)$. Therefore, for these groups it is easy to identify them with a geometric object: the unit circle. Instead of talking about different descriptions for $SO(2)$ or $U(1)$, which are defined by objects of given dimension, it can help to think about this group as the unit-circle. Rotations in two-dimensions are, as a Lie group, the unit-circle and we can represent these transformations by elements of $SO(2)$, i.e. 2×2 matrices or elements of $U(1)$, i.e. unit-complex numbers.

The next groups we discussed were $SO(3)$ and $SU(2)$. Remember that we found a one-to-one map between $SU(2)$ and the unit quaternions. The unit quaternions are defined as those quaternions $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ that satisfy the condition (Eq. 3.28)

$$a^2 + b^2 + c^2 + d^2 \overset{!}{=} 1, \qquad (3.84)$$

which is the same condition that defines[61] the three sphere S^3! Therefore the quaternions provide us with a map between $SU(2)$ and the three sphere S^3. This map is an isomorphism (1-1 and onto) and therefore we can really think of $SU(2)$ as the three sphere S^3.

These observations motivate the modern definition of a Lie group[62]: *A Lie group is a group, which is also a differentiable manifold[63]. Furthermore, the group operation $\circ$ must induce a differentiable map of the manifold into itself. This is a compatibility requirement that ensures that the group property is compatible with the manifold property. Concretely this means that every group element, say A induces a map that takes any element of the group B to another element of the group C $=$ AB and this map must be differentiable. Using coordinates this means that the coordinates of AB must be differentiable functions of the coordinates of B.*

We can now understand the remark at the end of Section 3.4:

"... there is precisely one **distinguished** Lie group for each Lie algebra."

a bit better[64]. From the geometric perspective, the distinguished group has the property of being **simply connected**. This means that, if we use the modern definition of a Lie group as a manifold, any closed curve on this manifold can be shrunk smoothly to a point[65].

To emphasize this important point:[66]

There is precisely one simply-connected Lie group corresponding to each Lie algebra.

[60] To be precise: an isomorphism. To say two things are isomorphic is the mathematical way of saying that they are "the same thing". Two things are called isomorphic if there exists an isomorphism between them.

[61] Recall that the unit circle S^1 is defined as the set of points that satisfy the condition $x_1^2 + x_2^2 = 1$. Equally, the two-sphere S^2 is defined by the condition $x_1^2 + x_2^2 + x_3^2 = 1$ and analogously the three sphere S^3 by $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$. The number that follows the S denotes the dimension. In two dimensions, with one condition we get a one-dimensional object: S^1. Equally we get in four dimensions, with one condition $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$ a three dimensional object S^3. S^3 is the surface of the four-dimensional ball.

[62] The technical details that follow aren't important for what we want to do in this book. Especially, don't worry if the exact meaning of notions like "induces" or "differentiable" is not clear. The important message to take away is: Lie group = manifold.

[63] A manifold is a set of points, for example a sphere that looks locally like flat Euclidean space R^n. Another way of thinking about a n-dimensional manifold is that it's a set which can be given n independent coordinates in some neighborhood of any point. For some more information about manifolds, see the appendix in Section 3.11 at the end of this chapter.

[64] Recall that we already discovered that different groups can have the *same* Lie algebra. For example, using the abstract definition of a Lie algebra, we say that $SO(3)$ and $SU(2)$ have the same Lie algebra (Eq. 3.82).

[65] We will not discuss this any further, but you are encouraged to read about it, for example in the books recommended at the end of this chapter. For the purpose of this book it suffices to know that there is always **one** distinguished group.

[66] A proof can be found, for example, in Michael Spivak. *A Comprehensive Introduction to Differential Geometry, Vol. 1, 3rd Edition.* Publish or Perish, 3rd edition, 1 1999. ISBN 9780914098706

This simply-connected group can be thought of as the "mother" of all those groups having the same Lie algebra, because there are maps to all other groups with the same Lie algebra from the simply connected group, but not vice versa. We could call it the mother group of this particular Lie algebra, but mathematicians tend to be less dramatic and call it the covering group. All other groups having the same Lie algebra are said to be covered by the simply connected one. We already stumbled upon an example of this: $SU(2)$ is the double cover of $SO(3)$. This means there is a two-to-one map from $SU(2)$ to $SO(3)$.

Furthermore, $SU(2)$ is the three sphere, which is a simply connected manifold. Therefore, we have already found the "most important" group belonging to the Lie algebra in Eq. 3.82. We can get all other groups belonging to this Lie algebra through maps from $SU(2)$.

We can now understand what manifold $SO(3)$ is. The map from $SU(2)$ to $SO(3)$ identifies with **two** points of $SU(2)$, **one** point of $SO(3)$. Therefore, we can think of $SO(3)$ as the top half[67] of S^3.

We can see, from the point of view that Lie groups are manifolds that $SU(2)$ is a more complete object than $SO(3)$. $SO(3)$ is just "part" of the complete object.

In this book we take the view that to describe nature at the most fundamental level, we need to use the most fundamental groups. For rotations in three dimensions this group is $SU(2)$ and not $SO(3)$. We will discover something similar when considering the symmetry group of special relativity.

We will see that Nature agrees with such lines of thought! To describe elementary particles one uses the representations of the *covering group* of the Poincaré group, instead of just the usual representation one uses to transform four-vectors. To describe nature at the most fundamental level, we must use the covering group, instead of any of the other groups that one can map to from the covering group.

We are able to derive the representations[68] of the most fundamental group, belonging to a given Lie algebra, by deriving representations of the Lie algebra. We can then put the matrices representing the Lie algebra elements (the generators) into the exponential function to get matrices representing group elements.

Herein lies the strength of Lie theory. By using pure mathematics we are able to reveal something fundamental about nature. The standard symmetry group of special relativity hides something from

[67] This picture is a bit oversimplified. Strictly speaking $SO(3)$ as a manifold is still a sphere, but with antipodal points identified.
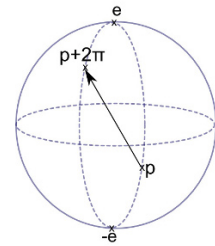
Fig. 3.7: Two-dimensional slice of the three Sphere S^3 (which is a three dimensional surface and therefore not drawable itself). We can see that the top half of the sphere is $SO(3)$, because to get from $SU(2)$ to $SO(3)$ we identify two points, for example, p and $p + 2\pi$, with each other.

[68] This notion will be made precise in the next section.

us, because it is not the most fundamental group belonging to this symmetry[69]. The covering group of the Poincaré group is the fundamental group and therefore we will use it to describe nature[70].

To summarize[71]

- $S^1 \hat{=} U(1) \underbrace{\leftrightarrow}_{\text{one-to-one}} SO(2)$

- $S^3 \hat{=} SU(2) \underbrace{\rightarrow}_{\text{two-to-one}} SO(3) \hat{=}$ "half" of S^3
  $\Rightarrow SU(2)$ is the distinguished group belonging to the Lie algebra $[J_i, J_j] = i\epsilon_{ijk}J_k$ (Eq. 3.82), because S^3 is simply connected.

Next, we will introduce another important branch of Lie theory, called representation theory. It is representation theory that enables us to derive from a given Lie group the tools that we need to describe nature at the most fundamental level.

3.5 Representation Theory

The important thing about group theory is that it is able to describe transformations without referring to any objects in the real world.

For theoretical considerations it is often useful to regard any group as an abstract group. This means defining the group by its manifold structure and the group operation. For example $SU(2)$ is the three sphere S^3, the elements of the group are points of the manifold and the rule associating a product point ab with any two points b and a satisfies the usual group axioms. In physical applications one is more interested in what the group actually does, i.e. the group action.

An important idea is that **one** group can act on **many** different kinds of objects[72]. This idea motivates the definition of a representation: A representation is a map[73] between any group element g of a group G and a linear transformation[74] $R(g)$ of some vector-space V

$$g \underbrace{\rightarrow}_{R} R(g) \tag{3.85}$$

in such a way that the group properties are preserved:

- $R(e) = I$ (The identity element of the group transforms nothing at all)

- $R(g^{-1}) = \left(R(g)\right)^{-1}$ (Every inverse element is mapped to the corresponding inverse transformation)

- $R(g) \circ R(h) = R(gh)$ (The combination of transformations corresponding to g and h is the same as the transformation corresponding to the point gh)

A representation[75] identifies with each point (abstract group element) of the group manifold (the abstract group) a linear transformation of a vector space. Although we define a representation as a map, most of the time we will call a set of matrices a representation. For example, the usual rotation matrices are a representation of the group $SO(3)$ on the vector space[76] R^3. The rotation matrices are linear transformations on R^3. However, the important thing here is that we can examine the group action on other vector spaces, too!

Using representation theory, we are able to investigate systematically how a given group acts on very different vector spaces and that is where things start to get really interesting.

One of the most important examples in physics is $SU(2)$. For example, we can examine how $SU(2)$ acts on the complex vector space of dimension one $\mathbb{C}^1$, which is especially easy, as we will see later. We can also investigate how $SU(2)$ acts on $\mathbb{C}^2$. The objects living in $\mathbb{C}^2$ are complex vectors of dimension two and therefore $SU(2)$ acts on them as 2×2 matrices. The matrices (=linear transformations) acting on $\mathbb{C}^2$ are just the "usual" $SU(2)$ matrices that we already know. In addition, we can examine how $SU(2)$ acts on $\mathbb{C}^3$ or even higher dimensional vector spaces. There is a well defined framework for constructing such representations and as a result, $SU(2)$ acts, for example, on complex vectors of dimension three as 3×3 matrices. A basis for the $SU(2)$ generators when they act on $\mathbb{C}^3$ is given by[77]

$$J_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad J_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \qquad J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

$$(3.86)$$

As usual, we can compute matrices that represent elements of the *group $SU(2)$* by putting linear combinations of these generators into the exponential function.

One can go on and inspect how $SU(2)$ acts on higher dimensional vectors. This can be quite confusing and it would be better to call[78] this group S^3 instead of $SU(2)$, because usually $SU(2)$ is defined as the set of complex 2×2 (!) matrices satisfying (Eq. 3.32)

$$U^\dagger U = 1 \quad \text{and} \quad \det(U) = 1 \qquad (3.87)$$

and now we write $SU(2)$ transformations as 3×3 matrices. Therefore

[75] This concept can be formulated more generally if one accepts arbitrary (not necessarily linear) transformations of an arbitrary (not necessarily a vector) space. Such a map is called a realization. In physics one is concerned most of the time with linear transformations of objects living in some vector space (for example Hilbert space in quantum mechanics or Minkowski space for special relativity), therefore the concept of a representation is more relevant to physics than the general concept called realization.

[76] R^3 denotes three dimensional Euclidean space, where elements are ordinary 3 component vectors, as we use them for example in Appendix A.1.

[77] We will learn later in this chapter how to derive these. At this point just take notice that it is possible.

[78] In an early draft version of this book the group was consequently called S^3. Unfortunately, such a non-standard name makes it hard for beginners to dive deeper into the subject using the standard textbooks.

one must always keep in mind that we mean the abstract group, instead of the 2×2 definition, when we talk about higher dimensional representation of $SU(2)$ or any other group.

Typically a group is defined in the first place with the help of an explicit representation. For example, we began our discussion of $SU(2)$ with explicit 2×2 matrices. This approach enables us to study the group properties concretely, as we did in the preceding sections. After this initial study it's often more helpful to regard the group as an abstract group[79], because it's possible to find other, useful representations of the group.

Before we move on to examples we need to define some abstract, but useful, notions. These notions will clarify the hierarchy of representations, because not every possible representation is equally fundamental.

The first notion we want to talk about is called **similarity transformation**. Given a matrix R and an invertible[80] matrix S then a transformation of the form

[80] A matrix S is called invertible, if there exists a matrix T, such that $ST = TS = 1$. The inverse matrix is usually denoted S^{-1}.

$$R \rightarrow R' = S^{-1}RS \tag{3.88}$$

is called a similarity transformation. The usefulness of this kind of transformation in this context lies in the fact that if we have a representation $R(G)$ of a group G, then $S^{-1}RS$ is also a representation. This follows directly from the definition of a representation: Suppose we have two group elements g_1, g_2 and a map $R: G \Rightarrow GL(V)$, i.e. $R(g_1)$ and $R(g_2)$. This is a representation if

$$R(g_1)R(g_2) = R(g_1g_2) \tag{3.89}$$

If we now look at the similarity transformation of the representation

$$S^{-1}R(g_1)\underbrace{SS^{-1}}_{=1}R(g_2)S = S^{-1}R(g_1)R(g_2)S = S^{-1}R(g_1g_2)S, \tag{3.90}$$

we see that this is a representation, too. Speaking colloquially, this means that if we have a representation, we can transform its elements wildly with literally any non-singular matrix S to get nicer matrices[81].

[81] The freedom to perform similarity transformations correspond to our freedom to choose a basis for the vector space our group acts on.

The next notion we want to introduce is called **invariant subspace**. When we have a representation R of a group G on a vector space V, we call $V' \subseteq V$ an invariant subspace if for[82] $v \in V'$ we have $R(g)v \in V'$ for all $g \in G$. This means, if we have a vector in the subspace V' and we act on it with arbitrary group elements, the

[82] Of course $v \in V$, too. The vector space V' must be part of the vector space V, which is mathematically denoted by $V' \subseteq V$. In other words this means that every element of V' is at the same time an element of V.

transformed vector will always be again part of the subspace V'. If we find such an invariant subspace we can define a representation R' of G on V', called a subrepresentation of R, by

$$R'(g)v = R(g)v \tag{3.91}$$

for all $v \in V'$. Therefore, one is led to the thought that the representation R, we talked about in the first place, is not fundamental, but a composite of smaller building blocks, called subrepresentations.

This leads us to the very important notion **irreducible representation**. An irreducible representation is a representation of a group G on a vector space V that has no invariant subspaces besides the zero space $\{0\}$ and V itself[83]. Such representations can be thought of as truly fundamental, because they are not made up by smaller representations. The irreducible representations of a group are the building blocks from which we can build up all other representations. There is another way to think about irreducible representation: An irreducible representation cannot be rewritten, using a similarity transformation, in block diagonal form[84]. In contrast, a reducible representation can be rewritten in block-diagonal form through similarity transformations. These notions are important because we use *irreducible* representations to describe elementary particles[85]. We will see later that the behavior of elementary particles under transformations is described by irreducible representations of the corresponding symmetry group.

There are many possible representations for each group[86], how do we know which one to choose to describe nature? There is an idea that is based on the Casimir elements. A Casimir element C is built from generators of the Lie algebra and its defining feature is that it commutes with every generator X of the group

$$[C, X] = 0. \tag{3.92}$$

What does this mean? A famous Lemma, called Schur's Lemma[87], tells us that if we have an irreducible representation $R : \mathfrak{g} \to GL(V)$, any linear operator $T : V \to V$ that commutes with all operators $R(X)$ must be a scalar multiple of the identity operator. Therefore, the Casimir elements give us linear operators with constant values for each representation. As we will see, these values provide us with a way of labelling representations naturally.[88] We can therefore start to investigate the irreducible representations, by starting with the representation with the lowest possible scalar value for the Casimir element.

Is there anything we can say about the vector space V mentioned

[83] The subspace consisting solely of the identity element is always, trivially an invariant subspace.

[84] An example for a matrix in block-diagonal form is $\begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & e \end{pmatrix}$.

[85] What else?

[86] For example we already know two different representations for rotations in two-dimensions. One using complex numbers and one using 2×2 matrices. Both are representations of S^1 as a group.

[87] A basic result of group theory, which you can look up in any book about group theory. For example, Schur's lemma is proven at page 239 in Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148.

[88] This will become much clearer as soon as we look at an example.

in the definition of a representation above? An important observation that helps us to make sense of the vector space, is that for any Lie group, one or several of the corresponding generators can be diagonalized using similarity transformations. In physics we use these diagonal generators to get labels for the basis vectors that span our vector space. We use the eigenvectors of these diagonal generators as basis for our vector space and the corresponding eigenvalues as labels. This idea is incredibly important to actually understand the physical implications of a given group. If there is just one generator that can be diagonalized simultaneously, each basis vector is labelled by just one number: the corresponding eigenvalue. If there are several generators that can be diagonalized simultaneously, we get several numbers as labels for each basic vector. Each such number is simply the eigenvalue of a given diagonal generator that belongs to this basic vector (=eigenvector). In particular this is where the "charge labels" for elementary particles: electric charge, weak charge, color charge, come from. This idea will make a lot more sense, once we are ready to look at some explicit examples[89].

In the next sections, we will derive the irreducible representations of the Lie algebra of $SU(2)$, because, as we will see, the Lie algebra of the Lorentz group can be thought of as two copies of the algebra $\mathfrak{su}(2)$[90]. The Lorentz group is part of the Poincaré group and therefore we will talk about these groups in this order.

3.6 $SU(2)$

We used in Section 3.4.3 specific matrices (=a specific representation) to identify how the generators of $SU(2)$ behave, when put into the Lie bracket[91]. We can use this knowledge to find further representations. Unsurprisingly, we will derive again the representation that we started with, i.e. the set of unitary 2×2 matrices with unit determinant. However, we are then able to see that this is just one special case. Before we tackle this task, we want to take a moment to think about what representations we can expect.

3.6.1 The Finite-dimensional Irreducible Representations of $SU(2)$

As noted earlier, there are special operators that we can build from a given set of generators that are useful to understand the representations of the Lie algebra in question. These operators are called Casimir operators and have the special property that they commute with all generators of the given Lie algebra. For the Lie algebra $\mathfrak{su}(2)$ there is one such operator[92]

[89] A bit more background information: The set of diagonal generators is called Cartan subalgebra, and the corresponding generators Cartan generators. As already mentioned, these generators play a big role in elementary particle physics, because the eigenvalues of the Cartan generators are used to give charge labels to elementary particles. For example, to derive quantum chromodynamics, we use the group $SU(3)$, as we will see later, and there are two $SU(3)$ Cartan generators. Therefore, each particle that interacts via chromodynamics, carries two charge labels. Conventionally instead of writing two numbers, one uses the words red, blue, green, and calls the corresponding charge colour. Analogous, the theory of weak interactions uses the group $SU(2)$, which has only one Cartan generator. Therefore, each particle is labelled by the corresponding eigenvalue of this Cartan generator.

[90] Technically it's the *complexification* of the Lie algebra of the Lorentz group that can be understood as two copies of the *complexification* of the Lie algebra $\mathfrak{su}(2)$.

[91] Recall that this is what we use to define the Lie algebra of a group in abstract terms. The final result was Eq. 3.82.

[92] We restrict ourselves here to quadratic Casimir operators, which means operators that are quadratic in the generators. There are also higher order Casimir operators, but here we always mean quadratic Casimir operators.

$$J^2 = J_1^2 + J_2^2 + J_3^3. \tag{3.93}$$

To illustrate this, let's consider again the 3-dimensional representation[93], as shown in Eq. 3.86. For, this representation the Casimir operator is

$$J^2_{\text{3-dim}} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \tag{3.94}$$

In other words, the Casimir operator is simply two times the 3×3 identity matrix. In contrast, for the 2-dimensional representation, as given by $J_i = \frac{1}{2}\sigma_i$, where σ_i are the Pauli matrices (Eq. 3.80), we get

$$J^2_{\text{2-dim}} = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{3}{4} \end{pmatrix}. \tag{3.95}$$

This illustrates that if the generators J_1, J_2, J_3 are given in some representation, we can calculate the corresponding Casimir operator explicitly. If we act with this Casimir operator on any element of the vector space that our generators act on[94], we get a number. This number is what we use to label different representations.

In addition to this label for different representations, we use additional special operators to label the different elements of the vector space that our groups acts on. The operators we use for this purpose are known as Cartan elements of the Lie algebra. In other words, while Casimir elements operators provide labels for different representations, the Cartan element provide labels within a given representation. The Cartan elements are all those generators that can be diagonalized simultaneously. For $\mathfrak{su}(2)$ there is only one such element and it is conventional to choose J_3 as diagonal generator[95]. In the 3-dimensional representation (Eq. 3.86) it is given by

$$J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix} \tag{3.96}$$

and in the 2-dimensional representation it is given by

$$J_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}. \tag{3.97}$$

We use for each representation the eigenvectors of the diagonal generator J_3 as basis vectors for the vector space that our representation acts on. This means that every such basis vector is labeled by two numbers[96]

$$J^2 |b, m\rangle = b |b, m\rangle$$
$$J_3 |b, m\rangle = m |b, m\rangle. \tag{3.98}$$

[93] So far, we have not discussed where this representation comes from. We will understand this in a moment. Here we only use the final result to explain the basic idea behind the use of the Casimir operator J^2.

[94] Recall that a representation is a map of the abstract group or Lie algebra elements to the linear operators that act on some vector space.

[95] As noted above, for other algebras like $\mathfrak{su}(3)$, there is more than one Cartan element and therefore we get multiple labels for each vector.

[96] We use here an abstract notation for the elements of the vector space that our generators act on. This notations emphasizes the labels that we use to distinguish different elements and is extremely popular in quantum mechanics, as will be discussed in Section 8.5.3.

The first label is the value of the quadratic Casimir operator J^2 for the representation in question and the second label corresponds to the value that we get when we act with the Cartan element J_3 on the vector. By looking at Eq. 3.95 and Eq. 3.97, we can see that the basis vectors for the 2-dimensional representation in this notation are[97] $|\frac{3}{4}, \frac{1}{2}\rangle$, $|\frac{3}{4}, -\frac{1}{2}\rangle$. Analogously, by looking at Eq. 3.94 and Eq. 3.96, we can see that the basis vectors for the 3-dimensional representation are[98] $|2, 1\rangle$, $|2, 0\rangle$, $|2, -1\rangle$.

After this preliminary discussion we are finally ready to understand the representations of $SU(2)$. To learn something about what finite-dimensional, irreducible representations of $SU(2)$ are possible, we define new operators from the ones we used in Section 3.4.3:[99]

$$J_+ = J_1 + iJ_2 \tag{3.99}$$
$$J_- = J_1 - iJ_2. \tag{3.100}$$

These new operators obey the following commutation relations, as you can check by using the commutator relations in Eq. 3.82[100]

$$[J_3, J_\pm] = \pm J_\pm \tag{3.101}$$
$$[J_+, J_-] = 2J_3. \tag{3.102}$$

If we now investigate how these operators act on an eigenvector $|b, m\rangle$ of J_3 with eigenvalue[101] m, we discover something remarkable:

$$J_3(J_\pm |b, m\rangle) = J_3(J_\pm |b, m\rangle) + \underbrace{J_\pm J_3 |b, m\rangle - J_\pm J_3 |b, m\rangle}_{=0}$$

$$= \underbrace{J_\pm J_3 |b, m\rangle}_{=J_\pm m |b,m\rangle} + \underbrace{J_3 J_\pm |b, m\rangle - J_\pm J_3 |b, m\rangle}_{=[J_3, J_\pm]|b,m\rangle}$$

$$\underbrace{=}_{\text{Eq. 3.101}} (m \pm 1) J_\pm |b, m\rangle. \tag{3.103}$$

We conclude that $J_\pm |b, m\rangle$ is again an eigenvector of J_3, but with eigenvalue $(m \pm 1)$ and thus we write[102] $J_\pm |b, m\rangle \equiv C |b, m \pm 1\rangle$:

$$J_3 |b, m \pm 1\rangle = (m \pm 1) |b, m \pm 1\rangle. \tag{3.104}$$

The operators J_- and J_+ are called raising and lowering operators or **ladder operators**. Starting from one J_3 eigenvector, we can construct more and more J_3 eigenvectors using the ladder operators $J_\pm$ repeatedly. This process must come to an end, because eigenvectors with different eigenvalues are linearly independent and we are dealing with finite-dimensional representations. This means that the corresponding vector space is finite-dimensional and therefore we can only find a finite number of linearly independent vectors.

We conclude that there must be an eigenvector with a maximum eigenvalue that we call j. For this maximum eigenvector, we must have

$$J_+ |b, j\rangle = 0, \tag{3.105}$$

because the only other possibility would be that, according to Eq. 3.103, we produce a new eigenvector with eigenvalue $j + 1$ and thus j wouldn't be the maximum eigenvalue. We can calculate a relationship between this maximum J_3 eigenvalue j and the label b for the whole representation by using[103]

$$\begin{aligned} J_- J_+ &= (J_1 - iJ_2)(J_1 + iJ_2) \\ &= = \underbrace{J_1^2 + J_2^2}_{=J^2 - J_3^2} + i \underbrace{(J_1 J_2 - J_2 J_1)}_{=[J_1, J_2] = iJ_3 \text{ (Eq. 3.82)}} = J^2 - J_3^2 - J_3 . \end{aligned} \tag{3.106}$$

> [103] This follows from $J^2 = J_1^2 + J_2^2 + J_3^2$ $\rightarrow J_1^2 + J_2^2 = J^2 - J_3^2$.

Using this expression, we can write

$$0 \underbrace{=}_{\text{Eq 3.105}} J_- J_+ |b, j\rangle \underbrace{=}_{\text{Eq 3.106 and Eq 3.98}} (b - j^2 - j) |b, j\rangle . \tag{3.107}$$

Therefore, we can conclude $b - j^2 - j = 0$ and thus

$$b = j(j + 1). \tag{3.108}$$

Completely analogous to how we concluded that there is an eigenvector of J_3 with maximum eigenvalue j, we can conclude that there is an eigenvector with minimal eigenvalue. We call this minimal eigenvalue k and we must have $J_- |b, k\rangle = 0$, because otherwise k would not be the minimal eigenvalue. Analogous to the calculation in Eq. 3.107, we can calculate

$$0 = J_+ J_- |b, k\rangle = (b - k^2 + k) |b, k\rangle . \tag{3.109}$$

Therefore, we have $(b - k^2 + k) = 0$ and can conclude

$$b = -k(-k + 1). \tag{3.110}$$

By comparing Eq. 3.108 with Eq. 3.110, we can conclude

$$k = -j. \tag{3.111}$$

Now it's time to recall our discussion from the beginning of this section. We have one label b coming from the quadratic Casimir operator J^2, which we use to distinguish different representations. In addition, we have another label m coming from the diagonal generator J_3, which we use to label different elements of the vector space our representation acts on. We have derived above that for finite dimensional representations, there is a maximal value for m, which

we called j. We introduced ladder operators $J_\pm$ that act on a given $|b,m\rangle$ and yield a new eigenvector of J_3 with eigenvalue $m \pm 1$. In practice, this means we can start with the eigenvector with maximum J_3 eigenvalue $|b,j\rangle$ for a given representation and then "move down the ladder" using J_-. The repeated application of J_- yields new eigenvectors of J_3. However, this process must come to an end at some eigenvector $|b,k\rangle$ with lowest J_3 eigenvalue k, because we are dealing with a finite dimensional representation and there can not be infinitely many eigenvectors. Acting with J_- on this eigenvector with minimal eigenvalue $|b,k\rangle$ yields simply zero. We calculated that this minimal eigenvalue is $k = -j$.

In summary, for each $SU(2)$ representation, labeled by $b = j(j+1)$ (Eq. 3.108), we get a set of J_3 eigenvectors with integer spaced eigenvalues in the range $-j \leq m \leq j$. As mentioned above, we can start climbing down the ladder at the vector $|b,j\rangle$ with maximum J_3 eigenvalue and get to the vector with minimum eigenvalue $|b,-j\rangle$ by repeated application of the operator J_-. After each application the eigenvalue is lowered by one. Therefore, the difference between the maximum J_3 eigenvalue j and the minimal value $-j$ is an integer: $j - (-j) =$ integer. From this we can conclude immediately

$$2j = \text{integer} \rightarrow j = \frac{\text{integer}}{2}. \tag{3.112}$$

This is an incredibly important insight, because it allows us to understand the possible finite-dimensional representations of $\mathfrak{su}(2)$. If we simply start plugging in the allowed values of $j = 0, 1/2, 1, 3/2, \ldots$, we get all possible $\mathfrak{su}(2)$ representations. In addition, we can use what learned above to understand and construct these representations explicitly. This explicit construction will be the topic of the next sections. It is particular instructive to have a look at the possible J_3 eigenvalues m for the different allowed values of j and the value of the quadratic Casimir operator J^2 for these representations[104]

$$
\begin{aligned}
&j = 0, &&m = 0, &&j(j+1) = 0 \\
&j = 1/2, &&m = 1/2, -1/2, &&j(j+1) = 3/4 \\
&j = 1, &&m = 1, 0, 1, &&j(j+1) = 2 \\
&j = 3/2, &&m = 3/2, 1/2, -1/2, -3/2, &&j(j+1) = 15/4 \\
&\ldots
\end{aligned} \tag{3.113}
$$

We learn here that the $j = 0$ representation is one-dimensional, the $j = 1/2$ representation is two-dimensional, the $j = 1$ representation is three-dimensional and the $j = 3/2$ representation is four-dimensional[105]. Take note that the values for the quadratic Casimir

[104] In Eq. 3.98 we called the eigenvalue that belongs to J^2 simply b. However in Eq. 3.108, we calculated that we express b through the maximum J_3 eigenvalue that we called j: $b = j(j+1)$.

[105] We use the eigenvectors of the Cartan generator, here J_3, as basis vectors for the vector space that our representation acts on. For example, if there are two possible eigenvalues of J_3, i.e. m values, we have two basis vectors and thus the corresponding representation is two-dimensional.

operator J^2 for the two-dimensional and the three-dimensional representation are exactly those that we calculated at the beginning of this section.

There is one more thing we need to discuss, before we can move on to the actual explicit construction of these various representations. We calculated in Eq. 3.103 that if we act with J_+ on a given J_3 eigenvector $|b, m\rangle$, we get a new J_3 eigenvector with a higher eigenvalue: $J_3 J_+ |b, m\rangle = (m + 1) J_+ |b, m\rangle$. However, we can not simply conclude that $J_+ |b, m\rangle = |b, m + 1\rangle$. Instead, we only know that $J_+ |b, m\rangle$ is **proportional** to $|b, m + 1\rangle$ and therefore we write

$$J_+ |b, m\rangle = C |b, m + 1\rangle \,, \tag{3.114}$$

where C is some constant. We now want to calculate this constant, because only then we have fully understood the action of the ladder operators and can actually use them to construct the representations explicitly. The thing is that we want to work with normalized basis vectors and hence want that $|b, m\rangle$ and $|b, m + 1\rangle$ are normalized:

$$|b, m\rangle^\dagger |b, m\rangle = 1$$
$$|b, m + 1\rangle^\dagger |b, m + 1\rangle = 1 \,. \tag{3.115}$$

If we act with J_+ or equally with J_- on a normalized basis vector like $|b, m\rangle$ the result will be, in general, not normalized. If we wouldn't care about normalization, we could write $J_+ |b, m\rangle = |b, m + 1\rangle$. However, for us a basis vector like $|b, m + 1\rangle$ means always a normalized basis vector and hence we need to include the additional constant C.

From the definition[106] of J_- and J_+ it follows that $J_-^\dagger = J_+$ and $J_+^\dagger = J_-$. Thus if we calculate the norm[107] of Eq. 3.114, we get

$$(J_+ |b, m\rangle)^\dagger J_+ |b, m\rangle = |C|^2 \underbrace{|b, m + 1\rangle^\dagger |b, m + 1\rangle}_{=1 \text{ (Eq. 3.115)}}$$

$$|b, m\rangle^\dagger J_+^\dagger J_+ |b, m\rangle = |C|^2$$

$$|b, m\rangle^\dagger \underbrace{J_- J_+}_{=J^2 - J_3^2 - J_3 \text{ (Eq. 3.106)}} |b, m\rangle = |C|^2$$

$$|b, m\rangle^\dagger \underbrace{(J^2 - J_3^2 - J_3)}_{=(b - m^2 - m)|b,m\rangle \text{ (Eq. 3.98)}} |b, m\rangle = |C|^2$$

$$|b, m\rangle^\dagger (\underbrace{b}_{=j(j+1) \text{ (Eq. 3.108)}} - m^2 - m) |b, m\rangle = |C|^2$$

[107] It will become clear in a second why this is a clever thing to do.

$$\underbrace{|b,m\rangle^{\dagger}\,|b,m\rangle}_{=1\ (\text{Eq. 3.115})}(j(j+1)-m^2-m)=|C|^2$$

$$(j(j+1)-m^2-m)=|C|^2. \qquad (3.116)$$

The final result of this calculation expresses the norm of the previously unknown constant C in terms of the labels j and m that we use to characterize our representation and the vectors within a representation. We can conclude[108] from Eq. 3.116:

$$\sqrt{(j(j+1)-m^2-m)}=C. \qquad (3.117)$$

Thus, for a given representation, which means a given j and a given basis vector, which means given m, we can use Eq. 3.117 to calculate the full result of what happens if we act with J_+ on a given eigenvector:

$$J_+\,|j(j+1),m\rangle = C\,|j(j+1),m+1\rangle$$

$$\underbrace{=}_{\text{Eq. 3.117}}\sqrt{(j(j+1)-m^2-m)}\,|j(j+1),m+1\rangle. \qquad (3.118)$$

We already argued above that if we act with J_+ on the eigenvector with maximum J_3 eigenvalue $|j(j+1),j\rangle$, we must get zero. If we now plug in $m=j$ in Eq. 3.117 we can see that this indeed happens:

$$\sqrt{(j(j+1)-j^2-j)}=C$$

$$\sqrt{j^2+j-j^2-j}=C$$

$$0=C. \qquad (3.119)$$

In addition, following exactly the same steps as above, we can calculate the constant $\tilde{C}$ that we get if we act with J_- on a given eigenvector. The result is

$$J_-\,|j(j+1),m\rangle = \tilde{C}\,|j(j+1),m+1\rangle$$

$$=\sqrt{(j(j+1)-m^2+m)}\,|j(j+1),m+1\rangle. \qquad (3.120)$$

We have now everything we need to calculate explicit matrix expressions for the various $SU(2)$ representations. In fact, it's possible to show that every irreducible representation of $SU(2)$ must be equivalent to one of these that we can construct by using the tools described above[109]. There is one small thing we need to discuss before we move on. The label $j(j+1)$ takes a lot of space and is somewhat redundant. It is therefore conventional to use simply j as a label instead. This means, we use $|j,m\rangle$ instead of $|j(j+1),m\rangle$.

Now we look at specific examples for the representations. We start, of course, with the lowest dimensional representation.

3.6.2 The Representation of $SU(2)$ in one Dimension

The lowest possible value for j is zero. As already mentioned above, this representation acts on a one-dimensional vector space. We can see that this representation is trivial, because the only 1×1 "matrix" that fulfills the commutation relations of the $SU(2)$ Lie algebra $[J_l, J_m] = i\epsilon_{lmn}J_n$, are the number 0. If we exponentiate the generator 0, we always get the transformation $U = e^0 = 1$, which changes nothing at all.

3.6.3 The Representation of $SU(2)$ in two Dimensions

We now take a look at the next possible value $j = \frac{1}{2}$. This representation is $2\frac{1}{2} + 1 = 2$ dimensional[110]. The generator J_3 has eigenvalues $\frac{1}{2}$ and $\frac{1}{2} - 1 = -\frac{1}{2}$, as can be seen from Eq. 3.113 and is therefore given by

$$J_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.121}$$

because we choose J_3 to be the diagonal generator[111]. The eigenvectors corresponding to the eigenvalues $+\frac{1}{2}$ and $-\frac{1}{2}$ are[112]

$$|1/2, 1/2\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |1/2, -1/2\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.122}$$

We can find the explicit matrix form of the other two $SU(2)$ generators J_1 and J_2 in this basis by rewriting them using the ladder operators

$$J_1 = \frac{1}{2}(J_- + J_+) \tag{3.123}$$

$$J_2 = \frac{i}{2}(J_- - J_+), \tag{3.124}$$

which we get directly from inverting the definitions of $J_\pm$ in Eq. 3.100 and Eq. 3.99. Recall that a basis for the vector space of this representation is given by the eigenvectors of J_3 and we therefore express the generators J_1 and J_2 in this basis. In other words, in this basis J_1 and J_2 are defined by their action on the eigenvectors of J_3. We compute

$$J_1 |1/2, 1/2\rangle = \frac{1}{2}(J_- + J_+) |1/2, 1/2\rangle$$

$$= \frac{1}{2}(J_- |1/2, 1/2\rangle + \underbrace{J_+ |1/2, 1/2\rangle}_{=0})$$

$$= \frac{1}{2}J_- |1/2, 1/2\rangle$$

$$\underbrace{=}_{\text{Eq. 3.120 with } \tilde{C}=1} \frac{1}{2} |1/2, -1/2\rangle, \tag{3.125}$$

[110] See Eq. 3.113.

[111] For $SU(2)$ only one generator is diagonal, because of the commutation relations. Furthermore, remember that we are able to transform the generators using similarity transformations and could therefore easily make another generator diagonal.

[112] As mentioned above, we use, for brevity, $|j, m\rangle$ instead of $|j(j+1), m\rangle$. Here $j = 1/2$.

where we used that $1/2$ is already the maximum J_3 eigenvalue and we cannot go higher. Similarly we get

$$J_1 \left|1/2, -1/2\right\rangle = \frac{1}{2}(J_- + J_+) \left|1/2, -1/2\right\rangle = \frac{1}{2} \left|1/2, 1/2\right\rangle . \quad (3.126)$$

Using Eq. 3.125 and Eq. 3.126, we can write J_1 in matrix form:

$$J_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \quad (3.127)$$

You can check that this matrix has the correct action on the basis vectors that we derived above[113]. In the same way, we find

$$J_2 = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} . \quad (3.128)$$

These are the same generators $J_i = \frac{1}{2}\sigma_i$, with the Pauli matrices σ_i, we found while investigating the Lie algebra of $SU(2)$ at the beginning of this chapter (Eq. 3.80). We can now see that the representation we used there was exactly this two dimensional representation. Nevertheless, there are many more, for example, in three-dimensions as we will see in the next section[114].

3.6.4   The Representation of $SU(2)$ in three Dimensions

Following the same procedure[115] as in two-dimensions, we find:

$$J_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$
$$(3.129)$$

This is the representation of the generators of $SU(2)$ in three dimensions. If you're interested, you can derive the corresponding representation for the group elements of $SU(2)$ in three dimensions, by putting these generators into the exponential function. We will not go any further and derive even higher dimensional representations, because at this point we already have everything we need to understand the most important representations of the Lorentz group.

3.7   The Lorentz Group $O(1,3)$

"To arrive at abstraction, it is always necessary to begin with a concrete reality ... You must always start with something. Afterward you can remove all traces of reality."

- Pablo Picasso[116]

---

[113] We derived in Eq. 3.125: $J_1 \left|1/2, 1/2\right\rangle = \frac{1}{2} \left|1/2, -1/2\right\rangle$. Using the explicit matrix form of J_1 we get
$J_1 \left|1/2, 1/2\right\rangle = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2} \left|1/2, -1/2\right\rangle$ ✓.

[114] Again, don't get confused by the name $SU(2)$, which we originally defined as the set of unitary 2×2 matrices with unit determinant. Here we mean the abstract group, defined by the corresponding manifold S^3 and we are going to talk about higher dimensional representations of this group, which result in, for example, a representation with 3×3 matrices. It would help if we could give this structure a different name (For example, using the name of the corresponding manifold S^3), but unfortunately $SU(2)$ is the conventional name.

[115] Again we start with the diagonal generator J_3, which we can write down immediately because we know its eigenvalues $(1, 0, -1)$ from Eq. 3.113. Afterwards, the other two generators J_1, J_2 can be derived by looking at how they act on the basis vectors, i.e. the eigenvectors of J_3. To calculate this, we use again that we can write J_1 and J_2 in terms of $J_\pm$.

[116] As quoted in Robert S. Root-Bernstein and Michele M. Root-Bernstein. *Sparks of Genius*. Mariner Books, 1st edition, 8 2001. ISBN 9780618127450

In this section we will derive **one** well-known representation of the Lorentz group. Then we will use this familiar representation to derive the Lie algebra of the Lorentz group. This is exactly the same route we followed for $SU(2)$. There we started with explicit 2×2 matrices to derive the corresponding Lie algebra. We will find that the complexified Lie algebra of the Lorentz group consists of two copies of the Lie algebra $\mathfrak{su}(2)$. This fact can be used to discover further representations of the Lorentz group, whereas the well-known vector representation, which is the representation of the Lorentz group by 4×4 matrices acting on four-vectors, will prove to be **one** of the representations. The new representations will provide us with tools to describe physical systems that cannot be described by the vector representation. This shows the power of Lie theory. Using Lie theory we are able to identify the hidden abstract structure of a symmetry and by using this knowledge, we are able to describe nature at the most fundamental level with the required tools.

We start with a characterisation of the Lorentz group and its subgroups. The Lorentz group is the set of all transformations that preserve the inner product of Minkowski space[117]

$$x^\mu x_\mu = x^\mu \eta_{\mu\nu} x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \qquad (3.130)$$

where $\eta_{\mu\nu}$ denotes the metric of Minkowski space

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \qquad (3.131)$$

[117] This was derived in Chapter 2. Recall that this definition is analogous to our definition of rotations and spatial reflections in Euclidean space, which preserve the inner product of Euclidean space.

This is the reason why we call the Lorentz group $O(1,3)$. The group $O(4)$ preserves $(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$. Let's see what restriction this imposes. The conventional name for a Lorentz transformation is Λ ("lambda"). For the moment, Λ is just a name and we will derive now how these transformations look like explicitly. If we transform $x^\mu \to x'^\mu = \Lambda^\mu_\nu x^\nu$, we get the product

$$x^\mu \eta_{\mu\nu} x^\nu \to x'^\sigma \eta_{\sigma\rho} x'^\rho = (x^\mu \Lambda^\sigma_\mu) \eta_{\sigma\rho} (\Lambda^\rho_\nu x^\nu) \overset{!}{=} x^\mu \eta_{\mu\nu} x^\nu \qquad (3.132)$$

and because this must hold for arbitrary x^μ we conclude

$$\Lambda^\sigma_\mu \eta_{\sigma\rho} \Lambda^\rho_\nu \overset{!}{=} \eta_{\mu\nu} \qquad (3.133)$$

or written in matrix form[118]

$$\Lambda^T \eta \Lambda \overset{!}{=} \eta. \qquad (3.134)$$

[118] Recall that in order to write the product of two vectors in matrix notation, the left vector is transposed. Therefore we get here Λ^T.

This is how the Lorentz transformations Λ are defined!

Starting from this definition, we will now derive a useful classification for all Lorentz transformations.

If we take the determinant of Eq. 3.133 and use $\det(AB) = \det(A)\det(B)$, we get the condition

$$\det(\Lambda)\underbrace{\det(\eta)}_{=-1}\det(\Lambda) \overset{!}{=} \underbrace{\det(\eta)}_{=-1} \rightarrow \det(\Lambda)^2 \overset{!}{=} 1 \qquad (3.135)$$

$$\rightarrow \det(\Lambda) \overset{!}{=} \pm 1 . \qquad (3.136)$$

Furthermore, if we look at the $\mu = \nu = 0$ component in Eq. 3.133[119], we get

$$\Lambda_0^\sigma \eta_{\sigma\rho} \Lambda_0^\rho \overset{!}{=} \underbrace{\eta_{00}}_{=1} \rightarrow \Lambda_0^\sigma \eta_{\sigma\rho} \Lambda_0^\rho = (\Lambda_0^0)^2 - \sum_i (\Lambda_0^i)^2 \overset{!}{=} 1 \qquad (3.137)$$

and thus conclude

$$\Lambda_0^0 \overset{!}{=} \pm\sqrt{1 + \sum_i (\Lambda_0^i)^2}. \qquad (3.138)$$

We now divide the Lorentz transformations into four "sub-categories", depending on the signs in Eq. 3.136 and Eq. 3.138:

$$
\begin{aligned}
L_+^\uparrow &: \det(\Lambda) &=& +1; \Lambda_0^0 \geq 1 \\
L_-^\uparrow &: \det(\Lambda) &=& -1; \Lambda_0^0 \geq 1 \\
L_+^\downarrow &: \det(\Lambda) &=& +1; \Lambda_0^0 \leq -1 \\
L_-^\downarrow &: \det(\Lambda) &=& -1; \Lambda_0^0 \leq -1 .
\end{aligned}
\qquad (3.139)
$$

This classification is useful, because, as we will see in a moment, only the transformations in one of these categories[120], $L_+^\uparrow$, can be generated through infinitesimal transformations. We, want to use the power of Lie theory and in particular learn as much as possible from the Lie algebra that belongs to a given group. Thus we focus on this sub-category $L_+^\uparrow$ and try to learn as much as possible using the corresponding generators. The transformations in the other categories are a combination of the transformations in this special category and one or both of two special transformations known as time-reversal Λ_T and space-inversion Λ_P, where[121]

$$
\Lambda_P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \qquad (3.140)
$$

[120] The technical term for this most important category $L_+^\uparrow$ is proper, orthochronous Lorentz group. "Proper" refers to the property $\det(\Lambda) = +1$, which in physical terms means that transformations in this category do not change the spatial orientation. For example, a right-handed coordinate system stays right-handed and does not become a left-handed one. The word "orthochronous" refers to the property $\Lambda_0^0 \geq 1$ and means in physical terms that transformations with this property do not change the direction of time.

[121] At least for one representation, these operators look like this. We will see later that for different representations, these operators look quite different. The subscript P here denotes "parity" which is another name for space-inversion. In physical terms a parity transformation is a reflection at the spatial axes, whereas a time-reversal transformation is a reflection at the time axis.

$$\Lambda_T = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.141}$$

This is illustrated in figure 3.8. In this sense, we can understand the complete Lorentz group as the set

$$O(1,3) = \{L_+^\uparrow, \Lambda_P L_+^\uparrow, \Lambda_T L_+^\uparrow, \Lambda_P \Lambda_T L_+^\uparrow\}. \tag{3.142}$$

So, why can only transformations in the $L_+^\uparrow$ category be built up by infinitesimal transformations? The transformations that can be built up by infinitesimal transformations are smoothly connected to the identity transformation, because the infinitesimal transformations are infinitesimally close to the identity element. Therefore, the identity element is in the same category as all the transformations that can be built up by repeating infinitesimal ones. We have

$$\det(\text{Id}) = +1; \ \text{Id}_0^0 = 1$$

and therefore the identity transformation and with it all transformations that can be generated through infinitesimal transformations, belong to the $L_+^\uparrow$ category.

Fig. 3.8: The four components of the Lorentz group are connected through the two discrete transformations known as time-reversal Λ_T and space-inversion Λ_P.

An important related observation is that the four categories that we defined above: $L_+^\uparrow, L_-^\uparrow, L_+^\downarrow, L_-^\downarrow$, are not smoothly connected to each other. For example, there is a "gap" between the transformations with $\det(\Lambda) = +1$ and those with $\det(\Lambda) = -1$. There are no transformations in between. Equally, there is a "gap" between $\Lambda_0^0 \geq 1$ and $\Lambda_0^0 \leq -1$ and also no transformations in between. The jump across this gap can only be achieved by making use of the discrete transformations Λ_P and Λ_T. Therefore, the transformations in the other categories are not smoothly connected to the identity element, which is part of the $L_+^\uparrow$ category. For this reason, we can not get the transformations in these other categories, by solely using infinitesimal transformations. Instead, we always need the discrete "jumps" provided by Λ_P and Λ_T.

To summarize: We concentrate in the following on the transformations in the $L_+^\uparrow$ subcategory, because it contains all transformations that can be built up by repeating infinitesimal ones. Such transformations are especially nice, because we can understand them through the corresponding Lie algebra. In practice this means that in the next step, we investigate the generators[122] that generate all transformations in the $L_+^\uparrow$ category. The transformations in the other categories are simply a combination of the transformations that we derive this way, and the discrete operations Λ_P and Λ_T.

[122] Recall: Generators are the elements of the Lie algebra.

3.7.1   One Representation of the Lorentz Group

Let's see how we can use the defining condition of the Lorentz group (Eq. 3.133) to construct an explicit matrix representation of the allowed transformations. First let's think a moment about what we are trying to find. The Lorentz group, when acting on 4-vectors[123], is given by real 4×4 matrices. The matrices must be real, because we want to know how they act on elements of the real Minkowski space $R^{(1,3)}$. A generic, real 4×4 matrix has 16 parameters. The defining condition of the Lorentz group, which is in fact 10 conditions[124], restricts this to 6 parameters. In other words, to describe a most general Lorentz transformation, 6 parameters are needed. Therefore, if we find 6 linearly independent generators, we have found a complete basis for the Lie algebra of this group. This means every other generator can be written as a linear combination of these basis generators. In addition, we are then able to compute how these basis generators behave when put into the Lie bracket and therefore to derive the abstract definition of this Lie algebra.

First note that the rotation matrices of 3-dimensional Euclidean space, which involve only space and leave the time unchanged, fulfil the condition in Eq. 3.133[125]. This follows because the spatial part[126] of the Minkowski metric is proportional to the 3×3 identity matrix[127] and therefore for transformations involving only space, we have from Eq. 3.133 the condition

$$-R^T I_{3\times3} R = -R^T R \stackrel{!}{=} -I_{3\times3}$$

$$\rightarrow R^T I_{3\times3} R = R^T R \stackrel{!}{=} I_{3\times3}.$$

This is exactly the defining condition of $O(3)$. Together with the condition (Eq. 3.139)

$$\det(\Lambda) \stackrel{!}{=} 1$$

these are the defining conditions of $SO(3)$. We conclude that the corresponding Lorentz transformations are given by

$$\Lambda_{\text{rot}} = \begin{pmatrix} 1 & \\ & R_{3\times3} \end{pmatrix}$$

with the rotation matrices $R_{3\times3}$ shown in Eq. 3.22 and derived in Section 3.4.1. The corresponding generators are therefore analogous to those we derived for three spatial dimensions in Section 3.4.1:

$$J_i = \begin{pmatrix} 0 & \\ & J_i^{3dim} \end{pmatrix}. \tag{3.143}$$

[123] The usual vector space of special relativity is the real, four-dimensional Minkowski space $R^{(1,3)}$. We will look at the representation on this vector space first, because the Lorentz group is defined there in the first place, i.e. as the set of transformations that preserve the 4×4 metric. Equivalently $SU(2)$ was defined as complex 2×2 matrices in the first place and we tried to learn as much as possible about $SU(2)$ from these matrices, in order to derive other representations later .

[124] You can see this, by putting a generic 4×4 matrix Λ, in $\Lambda^T \eta \Lambda = \eta$.

[125] $\Lambda_\mu^\sigma \eta_{\sigma\rho} \Lambda_\nu^\rho \stackrel{!}{=} \eta_{\mu\nu}$

[126] The spatial part are the components $\mu = 1, 2, 3$. Commonly this is denoted by η_{ij}, because Latin indices, like i, j always run from 1 to 3 and Greek indices, like μ and ν, run from 0 to 3.

[127] Recall $\eta_{11} = \eta_{22} = \eta_{33} = -1$ and $\eta_{ij} = 0$ for $i \neq j$.

For example, using Eq. 3.71 we now have

$$J_1 = \begin{pmatrix} 0 & \\ & J_1^{3dim} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix}. \tag{3.144}$$

To investigate transformations involving time **and** space we will start, as always in Lie theory, with an infinitesimal transformation[128]

$$\Lambda_\rho^\mu \approx \delta_\rho^\mu + \epsilon K_\rho^\mu. \tag{3.145}$$

We put this into the defining condition (Eq. 3.133)

$$\Lambda_\rho^\mu \eta_{\mu\nu} \Lambda_\sigma^\nu \overset{!}{=} \eta_{\rho\sigma}$$

$$\rightarrow (\delta_\rho^\mu + \epsilon K_\rho^\mu)\eta_{\mu\nu}(\delta_\sigma^\nu + \epsilon K_\sigma^\nu) \overset{!}{=} \eta_{\rho\sigma}$$

$$\rightarrow \cancel{\eta_{\rho\sigma}} + \epsilon K_\rho^\mu \eta_{\mu\sigma} + \epsilon K_\sigma^\nu \eta_{\rho\nu} + \underbrace{\epsilon^2 K_\rho^\mu \eta_{\mu\nu} K_\sigma^\nu}_{\approx 0 \text{ because } \epsilon \text{ is infinitesimal} \rightarrow \epsilon^2 \approx 0} = \cancel{\eta_{\rho\sigma}}$$

$$\rightarrow K_\rho^\mu \eta_{\mu\sigma} + K_\sigma^\nu \eta_{\rho\nu} = 0 \tag{3.146}$$

which reads in matrix form[129]

$$K^T \eta = -\eta K. \tag{3.147}$$

Now we have the condition for the generators of transformations involving time and space. A transformation generated by these generators is called a **boost**. A boost means a change into a coordinate system that moves with a different constant velocity compared with the original coordinate system. We can boost the description that we have, for example in a frame of reference where the object in question is at rest, into a frame of reference where it moves relative to the observer. Let's go back to the example used in Chapter 2.1: A boost along the x-axis. Because we know that $y' = y$ and $z' = z$ the generator is of the form

$$K_1 = \begin{pmatrix} \underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{\equiv k_1} & \\ & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} \tag{3.148}$$

and we only need to solve a 2×2 matrix equation. Equation 3.147 reduces to

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = -\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

[128] With the Kronecker delta defined by $\delta_\rho^\mu = 1$ for $\mu = \rho$ and $\delta_\rho^\mu = 0$ for $\mu \neq \rho$. This means writing the Kronecker delta in matrix form is just the identity matrix.

[129] Recall that the first index denotes the row and the second the column. So far we have been a little sloppy with first and second index, by writing them above each other. In fact, we have $K_\rho^\mu \equiv K^\mu_\rho \rightarrow (K^T)^\mu_\rho = K^{\ \mu}_\rho$. Matrix multiplication always works by multiplying rows with columns. Therefore $K_\sigma^\nu \eta_{\rho\nu} = \eta_{\rho\nu} K_\sigma^\nu$, where the ρ-row of η is multiplied with the σ-column of K. This term then is in matrix notation ηK. Furthermore, $K_\rho^\mu \eta_{\mu\sigma} = K^\mu_\rho \eta_{\mu\sigma} = (K^T)^{\ \mu}_\rho \eta_{\mu\sigma}$. In order to write this index term in matrix notation we need to use the transpose of K, because only then we get a product of the form row times column. The ρ-row of K^T is multiplied with the σ-column of η. Therefore, this term is $K^T \eta$ in matrix notation. In index notation we are free to move objects around, because for example K_ρ^μ is just one element of K, i.e. a number.

[130] $\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}$ and

$-\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = -\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$.

which is solved by[130]

$$k_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

The complete generator for boosts along the x-axis is therefore

$$K_1 = \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.149}$$

and equally we can find the generators for boosts along the y- and z-axis

$$K_2 = \begin{pmatrix} 0 & 0 & i & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad K_3 = \begin{pmatrix} 0 & 0 & 0 & i \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ i & 0 & 0 & 0 \end{pmatrix}. \tag{3.150}$$

Now, we already know from Lie theory how we get from the generators to finite transformations

$$\Lambda_1(\phi) = e^{i\phi K_1}.$$

For brevity let's focus again on the exciting part of the generator K_1, i.e. the upper left 2×2 matrix k_1, which is defined in Eq. 3.148. We can then evaluate the exponential function using its series expansion and that[131] $(ik_1)^2 = 1$

[131] As you can easily check: $(ik_1)^2 = i^2 \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, equally $k_1^4 = 1$ etc. for all even exponents and of course $(ik_1)^3 = ik_1$, $k_1^5 = ik_1$ etc. for all uneven exponents.

$$\Lambda_1(\phi) = e^{i\phi k_1} = \sum_{n=0}^{\infty} \frac{i^n \phi^n k_1^n}{n!} = \sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \underbrace{(ik_1)^{2n}}_{=1} + \sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \underbrace{(ik_1^{2n+1})}_{=ik_1}$$

$$= \left(\sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \right) I + i \left(\sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \right) k_1 = \cosh(\phi) I + i\sinh(\phi) k_1$$

$$= \begin{pmatrix} \cosh(\phi) & 0 \\ 0 & \cosh(\phi) \end{pmatrix} + \begin{pmatrix} 0 & -\sinh(\phi) \\ -\sinh(\phi) & 0 \end{pmatrix}$$

$$= \begin{pmatrix} \cosh(\phi) & -\sinh(\phi) \\ -\sinh(\phi) & \cosh(\phi) \end{pmatrix} \tag{3.151}$$

This computation is analogous to the computation in Section 3.4.1, but observe that the sums here have no factor $(-1)^n$ and therefore these sums are not $\sin(\phi)$ and $\cos(\phi)$, but different functions called hyperbolic sine $\sinh(\phi)$ and hyperbolic cosine $\cosh(\phi)$. The complete 4×4 transformation matrix for a boost along the x-axis is therefore

$$\Lambda_1 = \begin{pmatrix} \cosh(\phi) & -\sinh(\phi) & 0 & 0 \\ -\sinh(\phi) & \cosh(\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.152}$$

Analogously, we can derive the transformation matrices for boosts along the other axes:

$$\Lambda_2 = \begin{pmatrix} \cosh(\phi) & 0 & -\sinh(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ -\sinh(\phi) & 0 & \cosh(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.153}$$

$$\Lambda_3 = \begin{pmatrix} \cosh(\phi) & 0 & 0 & -\sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\sinh(\phi) & 0 & 0 & \cosh(\phi) \end{pmatrix}. \tag{3.154}$$

An arbitrary boost can be composed by multiplication of these 3 transformation matrices.

3.7.2 Generators of the Other Components of the Lorentz Group

To understand how the generators for the transformations of the other components[132] of the Lorentz Group look like, we simply have to act with the parity operation Λ_P and the time reversal operator Λ_T on the matrices J_i, K_i we just derived. In index notation we have[133]

$$(\Lambda_P)^{\alpha}_{\alpha'} (\Lambda_P)^{\beta}_{\beta'} (J_i)^{\alpha'\beta'} \underbrace{\hateq}_{\text{switching to matrix notation}} \Lambda_P J_i (\Lambda_P)^T = J_i \hateq (J_i)^{\alpha\beta} \tag{3.155}$$

$$(\Lambda_P)^{\alpha}_{\alpha'} (\Lambda_P)^{\beta}_{\beta'} (K_i)^{\alpha'\beta'} \underbrace{\hateq}_{\text{switching to matrix notation}} \Lambda_P K_i (\Lambda_P)^T = -K_i \hateq -(K_i)^{\alpha\beta}, \tag{3.156}$$

as we can check by a brute force computation, using the explicit matrices derived in the last section. For example,

$$J_1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} \rightarrow J_1' = \Lambda_P J_1 (\Lambda_P)^T = J_1, \tag{3.157}$$

because

$$\Lambda_P J_1 (\Lambda_P)^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^T$$

$$= \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \end{pmatrix} \tag{3.158}$$

[132] Recall that the Lorentz group is in fact $O(1,3) = \{L_+^\uparrow, \Lambda_P L_+^\uparrow, \Lambda_T L_+^\uparrow, \Lambda_P \Lambda_T L_+^\uparrow\}$ and we derived in the last section the generators of $L_+^\uparrow$.

[133] We need two matrices Λ_P, one for each index. This is just the ordinary transformation behavior of operators under changes of the coordinate system.

In contrast,

$$K_1 = \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow K_1' = \Lambda_P K_1 (\Lambda_P)^T = -K_1, \qquad (3.159)$$

because

$$\Lambda_P K_1 (\Lambda_P)^T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^T$$

$$= -\begin{pmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \qquad (3.160)$$

In conclusion, we have under parity transformations

$$J_i \underset{P}{\rightarrow} J_i \qquad K_i \underset{P}{\rightarrow} -K_i. \qquad (3.161)$$

This will become useful later, because for different representations the parity transformations will not be as obvious as in the vector representation. Equally we can investigate the time-reversed generators and the result will be the same, because time-reversal involves only the first component, which only changes something for the boost generators K_i

$$(\Lambda_T)^\alpha_{\alpha'} (\Lambda_T)^\beta_{\beta'} (J_i)^{\alpha'\beta'} \underset{\text{switching to matrix notation}}{\hat{=}} \Lambda_T J_i (\Lambda_T)^T = J_i \hat{=} (J_i)^{\alpha\beta} \qquad (3.162)$$

$$(\Lambda_T)^\alpha_{\alpha'} (\Lambda_T)^\beta_{\beta'} (K_i)^{\alpha'\beta'} \underset{\text{switching to matrix notation}}{\hat{=}} \Lambda_T K_i (\Lambda_T)^T = -(K_i)^{\alpha\beta}, \qquad (3.163)$$

Or shorter:

$$J_i \underset{T}{\rightarrow} J_i \qquad K_i \underset{T}{\rightarrow} -K_i. \qquad (3.164)$$

3.7.3 The Lie Algebra of the Proper Orthochronous Lorentz Group

[134] See Eq. 3.149 for the boost generators and Eq. 3.61 for the rotation generators

[135] The Levi-Civita symbol ϵ_{ijk}, is defined in Appendix B.5.5.

Now using the explicit matrix form of the generators[134] for $L^\uparrow_+$ we can derive the corresponding Lie algebra by brute force computation[135]

$$[J_i, J_j] = i\epsilon_{ijk}J_k \qquad (3.165)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \qquad (3.166)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k , \qquad (3.167)$$

where again J_i denotes the generators of rotations and K_i are the generators of boosts. A general Lorentz transformation is of the form

$$\Lambda = e^{i\vec{J}\cdot\vec{\theta} + i\vec{K}\cdot\vec{\Phi}} \qquad (3.168)$$

Equation 3.166 tells us that the two types of generator (J_i and K_i) do not commute with each other. While the rotation generators are closed under commutation[136], the boost generators are not[137]. We can now define two new operator types from the old ones that are closed under commutation and commute with each other[138]

$$N_i^{\pm} = \frac{1}{2}(J_i \pm iK_i). \qquad (3.169)$$

Working out the commutation relations yields

$$[N_i^+, N_j^+] = i\epsilon_{ijk}N_k^+ \qquad (3.170)$$

$$[N_i^-, N_j^-] = i\epsilon_{ijk}N_k^- \qquad (3.171)$$

$$[N_i^+, N_j^-] = 0. \qquad (3.172)$$

These are precisely the commutation relations for the Lie algebra of $SU(2)$ and we have therefore discovered that the complexified Lie algebra of $L_+^{\uparrow}$ consists of two copies of the Lie algebra $\mathfrak{su}(2)$[139].

This is great news, because we already know how to construct all irreducible representations of the Lie algebra of $\mathfrak{su}(2)$. However we must be careful. The Lorentz group is, like $SO(3)$, not simply-connected[140] and Lie theory tells us that, for groups that aren't simply connected, there is no one-to-one correspondence between the irreducible representations of the Lie algebra and representations of the corresponding group[141]. Instead, **by deriving the irreducible representations of the complexified Lie algebra of the Lorentz group, we find the irreducible representations of the covering group of the Lorentz group**, if we put the corresponding generators into the exponential function. Some of these representations will be representations of the Lorentz group, but we will find more than that. This is good, because we need these additional representations to describe certain elementary particles.

[136] Closed under commutation means that the commutator $[J_i, J_j] = J_iJ_j - J_jJ_i$, is again a rotation generator. From Eq. 3.165 we can see that this is the case.

[137] Eq. 3.167 tells us that the commutator of two boost generators K_i and K_j isn't another boost generator, but a generator of rotations.

[138] Again, take note that we use a complex linear combination here and recall that this process of considering a complex linear combination, instead of the original generators is called a complexification. Usually we only allow real linear combinations of the generators. So from here on, we consider the complexification of the Lie algebra of the Lorentz group.

[139] Recall that when we discussed the representations of the Lie algebra $\mathfrak{su}(2)$, we also used the corresponding complexification. The complexification of $\mathfrak{su}(2)$ is $\mathfrak{sl}(2,\mathbb{C})$ and therefore technically we have $\mathfrak{so}(1,3)_{\mathbb{C}} \cong \mathfrak{sl}(2,\mathbb{C}) \oplus \mathfrak{sl}(2,\mathbb{C})$, where $\mathfrak{so}(1,3)_{\mathbb{C}}$ denotes the complexification of the Lorentz group Lie algebra.

[140] We will use this simply as a fact here, because a proof would lead us too far apart.

[141] This can be quite confusing, but remember that there **is** always one distinguished group that belongs to a Lie algebra. This group is distinguished because it is simply connected. If we derive the irreducible representation of a Lie algebra, we get, by putting those Lie algebra elements (= the generators) in the exponential function, representations of the simply connected (= covering) group. Only for the simply connected group there is a one-to-one correspondence.

For brevity, we will continue to call the representations we will derive, representations of the Lorentz group instead of representations of the Lie algebra of the Lorentz group or representations of the double cover of the Lorentz group.

Each irreducible representation of the Lie algebra $\mathfrak{su}(2)$ can be labeled by the scalar value j of the $\mathfrak{su}(2)$ Casimir operator[142]. Therefore, we now know that we can label the irreducible representations of the covering group[143] of the Lorentz group by two integer or half integer numbers: j_1 and j_2. This means we will look at the (j_1, j_2) representations and use the $j_1, j_2 = 0, \frac{1}{2}, 1 \ldots$ representations for the two $\mathfrak{su}(2)$ copies, which we derived earlier.

It is conventional to write the Lorentz algebra in a more compact way by introducing a new symbol $M_{\mu\nu}$, which is defined through the equations

$$J_i = \frac{1}{2}\epsilon_{ijk}M_{jk}. \tag{3.173}$$

$$K_i = M_{0i}. \tag{3.174}$$

With this new definition the Lorentz algebra reads

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho}). \tag{3.175}$$

Next, we investigate different representations of the complexified Lie algebra of the Lorentz group and hence of the double cover of the Lorentz group in detail.

3.7.4   The $(0,0)$ Representation

The lowest dimensional representation is, as it is for $SU(2)$, trivial, because the vector space is one-dimensional for both copies of $\mathfrak{su}(2)$. Our generators must therefore be 1×1 matrices and the only 1×1 "matrix" that fulfills the commutation relations is the number 0:

$$N_i^+ = N_i^- = 0 \rightarrow e^{iN_i^+} = e^{iN_i^-} = e^0 = 1 \tag{3.176}$$

Therefore we conclude that the $(0,0)$ representation of the Lorentz group acts on objects that do **not** change under Lorentz transformations. This representation is called the **scalar representation**.

3.7.5   The $\left(\frac{1}{2}, 0\right)$ Representation

In this representation we use the[144] 2 dimensional representation for one copy of the $SU(2)$ Lie algebra N_i^+, i.e. $N_i^+ = \frac{\sigma_i}{2}$ and the 1

[142] To be precise: the scalar value that we get from the *quadratic* Casimir operator is $j(j+1)$. However, for brevity, we simply use j instead.

[143] The covering group of the proper orthochronous Lorentz group $L_+^\uparrow$ is $SL(2,C)$, which is defined as the set of 2×2 matrices with unit determinant and complex entries. The relationship $SL(2,C) \rightarrow L_+^\uparrow \equiv SO(1,3)_+^\uparrow$ is similar to the relationship $SU(2) \rightarrow SO(3)$ we discovered earlier in this text.

[144] Recall that the dimension of our vector space is given by $2j + 1$. Therefore we have here $2\frac{1}{2} + 1 = 2$ dimensions.

dimensional representation for the other N_i^-, i.e. $N_i^- = 0$. From the definition of N^- in Eq. 3.169 we conclude

$$N_i^- = \frac{1}{2}(J_i - iK_i) = 0 \qquad (3.177)$$

$$\rightarrow J_i = iK_i. \qquad (3.178)$$

Furthermore, we can use that we already derived in Section 3.6.3 the two dimensional representation of $SU(2)$:

$$N_i^+ = \frac{\sigma_i}{2}, \qquad (3.179)$$

where σ_i denotes once more the Pauli matrices, which were defined in Eq. 3.80. On the other hand, we have

$$N_i^+ \underbrace{=}_{\text{Eq. 3.169}} \frac{1}{2}(J_i + iK_i) \underbrace{=}_{\text{Eq. 3.178}} \frac{1}{2}(iK_i + iK_i) = iK_i \qquad (3.180)$$

Comparing Eq. 3.179 with Eq. 3.180 tells us that

$$iK_i = \frac{\sigma_i}{2} \rightarrow K_i = \frac{\sigma_i}{2i} = \frac{i\sigma_i}{2i^2} = \frac{-i}{2}\sigma_i \qquad (3.181)$$

$$\text{Eq. 3.178} \rightarrow J_i = iK_i = \frac{-i^2}{2}\sigma_i = \frac{1}{2}\sigma_i. \qquad (3.182)$$

We conclude that a Lorentz rotation in this representation is given by

$$R_\theta = e^{i\vec{\theta}\cdot\vec{J}} = e^{i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}} \qquad (3.183)$$

and a Lorentz boost by

$$B_\phi = e^{i\vec{\phi}\cdot\vec{K}} = e^{\vec{\phi}\cdot\frac{\vec{\sigma}}{2}}. \qquad (3.184)$$

By writing out the exponential function as series expansion we can easily get the representation of the double cover of the Lorentz group from the representation of the generators. For example, rotations about the x-axis e.g. are given by

$$R_x(\theta) = e^{i\theta J_1} = e^{i\theta \frac{1}{2}\sigma_1} = 1 + \frac{i}{2}\theta\sigma_1 + \frac{1}{2}\left(\frac{i}{2}\theta\sigma_1\right)^2 + \ldots \qquad (3.185)$$

And if we use the explicit matrix form of σ_1 (Eq. 3.80), together with the fact that $\sigma_1^2 = 1$, we get[145]

[145] The steps are completely analogous to what we did in Section 3.4.1

$$R_x(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{i}{2}\theta\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \frac{1}{2}\left(\frac{\theta}{2}\right)^2\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \ldots$$

$$= \begin{pmatrix} \cos(\frac{\theta}{2}) & i\sin(\frac{\theta}{2}) \\ i\sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{pmatrix}. \qquad (3.186)$$

[146] We will learn later that these two components correspond to **spin-up** and **spin-down** states.

[147] This name will make more sense after the definition of right-chiral spinors. Then we can see that parity transformations transform a left-chiral spinor transformation into a right-chiral spinor transformation and vice versa. These spinors are often called left-handed and right-handed, but this can be confusing, because these terms correspond originally to a concept called helicity, which is not the same as chirality. Recall what the parity operator does: changing a left-handed coordinate system into a right-handed coordinate system and vice versa. Hence the name.

[148] There is much more one can say about spinors. See, for example, chapter 3.2 in J. J. Sakurai. *Modern Quantum Mechanics*. Addison Wesley, 1st edition, 9 1993. ISBN 9780201539295

[149] Andrew M. Steane. An introduction to spinors. *ArXiv e-prints*, December 2013

Analogously, we can compute the transformation matrix for rotations around other axes or boosts. One important thing to notice is that we have here complex 2×2 matrices, representing the Lorentz transformations. These transformations certainly do not act on the four-vectors of Minkowski space, because these have 4 components. The **two-component**[146] objects this representation acts on are called **left-chiral spinors**[147]:

$$\chi_L = \begin{pmatrix} (\chi_L)_1 \\ (\chi_L)_2 \end{pmatrix} \tag{3.187}$$

Spinors in this context are two component objects. A possible definition for left-chiral spinors is that they are objects that transform under Lorentz transformations according to the $(\frac{1}{2}, 0)$ representation of the Lorentz group. Take note that this is not just another way to describe the same thing, because spinors have properties that usual vectors do not have. For instance, the factor $\frac{1}{2}$ in the exponent. This factor shows us that a spinor is after a rotation by 2π not the same, but gets a minus sign[148]. This is a pretty crazy property, because we usually expect that objects are exactly the same after a rotation by $360° = 2\pi$.

In the last section, we saw that the lowest-dimensional representation is trivial. The next representation in the "hierarchy" of representations is the spinor representation that we discovered in this section. In this sense, we can say that

"... a spinor is the most basic sort of mathematical object that can be Lorentz-transformed."

- **A. M. Steane**[149]

3.7.6   The $(0, \frac{1}{2})$ Representation

This representation can be constructed analogous to the $(\frac{1}{2}, 0)$ representation but this time we use the 1 dimensional representation for N_i^+, i.e. $N_i^+ = 0$ and the two dimensional representation for N_i^-, i.e. $N_i^- = \frac{1}{2}\sigma_i$. A first guess could be that this representation looks exactly like the $(\frac{1}{2}, 0)$ representation, but this is not the case! This time we get from the definition of N^+ in Eq. 3.169

$$N_i^+ = \frac{1}{2}(J_i + iK_i) = 0 \tag{3.188}$$

$$\rightarrow J_i = -iK_i. \tag{3.189}$$

Take notice of the minus sign. Using the two-dimensional representation of $\mathfrak{su}(2)$ for N^+, which was derived in Section 3.6.3, yields

$$N_i^- = \frac{1}{2}\sigma_i = \frac{1}{2}(J_i - iK_i) \underbrace{=}_{\text{Eq. 3.189}} \frac{1}{2}(-iK_i - iK_i) = -iK_i. \qquad (3.190)$$

Using this we can deduce the $(0, \frac{1}{2})$ representation of the boost generators

$$-iK_i = \frac{1}{2}\sigma_i \rightarrow K_i = \frac{-1}{2i}\sigma_i = \frac{-i}{2i^2}\sigma_i = \frac{i}{2}\sigma_i. \qquad (3.191)$$

In addition, from Eq. 3.189 we get

$$J_i = -iK_i = \frac{1}{2}\sigma_i. \qquad (3.192)$$

We conclude that in this representation a Lorentz rotation is given by

$$R_\theta = e^{i\vec{\theta}\cdot\vec{J}} = e^{i\vec{\theta}\cdot\frac{\vec{\sigma}}{2}} \qquad (3.193)$$

and a Lorentz boost by

$$B_\phi = e^{i\vec{\phi}\cdot\vec{K}} = e^{-\vec{\phi}\cdot\frac{\vec{\sigma}}{2}}. \qquad (3.194)$$

Therefore, rotations are the same as in the $(\frac{1}{2}, 0)$ representation, but boosts differ by a minus sign in the exponent. We conclude both representations act on objects that are **similar but not the same**. We call the objects the $(0, \frac{1}{2})$ representation of the Lorentz group acts on **right-chiral spinors**:

$$\chi_R = \begin{pmatrix} (\chi_R)^1 \\ (\chi_R)^2 \end{pmatrix} \qquad (3.195)$$

The generic name for left- and right-chiral spinors is **Weyl spinors**.

3.7.7 Van der Waerden Notation

Now we introduce a notation that makes working with spinors very convenient. We know that we have two kinds of objects that transform differently and therefore must be distinguished. In the last section we learned that they are different, but not too different. In a moment we will learn that there is a connection between the objects transforming according to the $(\frac{1}{2}, 0)$ representation (left-chiral spinors) and the objects transforming according to the $(0, \frac{1}{2})$ representation (right-chiral spinors). To be able to describe these different objects using one notation we introduce the notions of dotted and undotted indices, sometimes called Van der Waerden notation, after their inventor. This will help us to keep track of which object transforms in what way. This will become much clearer in a minute, as soon as we have set up the full formalism.

Let's define that a left-chiral spinor χ_L has a lower, undotted index

$$\chi_L = \chi_a \tag{3.196}$$

and a right-chiral spinor χ_R has an upper, dotted index

$$\chi_R = \chi^{\dot{a}}. \tag{3.197}$$

Next, we introduce the "spinor metric". The spinor metric enables us to transform a right-chiral spinor into a left-chiral and vice versa, but not alone as we will see. We define the spinor metric[150] as

$$\epsilon^{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{3.198}$$

and show that it has the desired properties. Furthermore, we define[151]

$$\chi_L^C \equiv \epsilon \chi_L^\star \tag{3.199}$$

where the $\star$ denotes complex conjugation. We will now inspect how χ_L^C transforms under Lorentz transformations and see that it transforms precisely as a right-chiral spinor. The defining feature of a right-chiral spinor is its transformation behavior and therefore we will conclude that χ_L^C *is a right-chiral spinor*. Let us have a look at how χ_L^C transforms under boosts, where we use

$$(-\epsilon)(\epsilon) = 1 \tag{3.200}$$

and

$$(\epsilon)\sigma_i^\star(-\epsilon) = -\sigma_i \tag{3.201}$$

for each Pauli matrix σ_i, as you can check by using Eq. 3.198. Transforming χ_L^C yields[152]

$$
\begin{aligned}
\chi_L^C \to \chi_L'^C &= \epsilon(\chi')_L^\star \\
&= \epsilon\left(\left(e^{\frac{\vec{\phi}}{2}\vec{\sigma}} \chi_L \right) \right)^\star \\
&= \epsilon\left(e^{\frac{\vec{\phi}}{2}\vec{\sigma}} \underbrace{(-\epsilon)(\epsilon)}_{=1 \ (\text{Eq. 3.200})} \chi_L \right)^\star \\
&= \underbrace{\epsilon e^{\frac{\vec{\phi}}{2}\vec{\sigma}^\star}(-\epsilon)}_{=e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \ (\text{Eq. 3.201})} (\epsilon)\chi_L^\star \\
&= e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \underbrace{\epsilon\chi_L^\star}_{=\chi_L^C} \\
&= e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \chi_L^C ,
\end{aligned}
\tag{3.202}
$$

[150] Take note that this is the Levi-Civita symbol in two dimensions as defined in Appendix B.5.5.

[151] Maybe a short comment on the strange notation χ_L^C is in order. The superscript C denotes charge conjugation, as will be explained in Section 3.7.10 in more detail. Here we see that this operation flips one label, i.e. a left-chiral spinor becomes right-chiral. Later we will see this operation flips all labels, including, for example, the electric charge.

[152] We use the notation $\vec{\phi}\vec{\sigma} = \sum_i \phi_i \sigma_i \underset{\substack{\text{summation convention}}}{=} \phi_i \sigma_i$. The "vector" $\vec{\sigma}$ shouldn't be taken too seriously, because it's just a shorthand, conventional notation.

which is exactly the transformation behavior of a *right-chiral* spinor[153].

To get to the fifth line, we use the series expansion of $e^{\frac{\vec{\phi}}{2}\vec{\sigma}}$ and Eq. 3.201 on every term. You can check in the same way that the behavior under rotations is not changed by complex conjugation and multiplication with ϵ, as it should be, because χ_L and χ_R transform in the same way under rotations:

$$\chi_L^C \to \chi_L'^C = \epsilon(\chi')_L^\star = \epsilon(e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\chi)_L^\star = e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\epsilon(\chi_L)^\star = e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\chi_L^C. \quad (3.203)$$

Furthermore, you can check that ϵ is invariant under all transformations and that if we want to go the other way round, i.e. transform a right-chiral spinor into a left-chiral spinor, we have to use $(-\epsilon)$.

Therefore, we define in analogy with the tensor notation of special relativity that our "metric" raises and lowers indices

$$\underbrace{\epsilon\chi_L =}_{\text{written in index notation}} \epsilon^{ac}\chi_c = \chi^a, \quad (3.204)$$

where summation over identical indices is implicitly assumed (Einstein summation convention). Furthermore, we know that if we want to get χ_R from χ_L we need to use complex conjugation as well

$$\chi_R = \epsilon\chi_L^\star. \quad (3.205)$$

This means that complex conjugation transforms an undotted index into a dotted index:

$$\chi_R = \epsilon\chi_L^\star = \chi^{\dot{a}}. \quad (3.206)$$

Therefore, we can get a lower, dotted index by complex conjugating χ_L:

$$\chi_L^\star = \chi_a^\star = \chi_{\dot{a}} \quad (3.207)$$

and an upper, undotted index, by complex conjugating χ_R

$$\chi_R^\star = (\chi^{\dot{a}})^\star = \chi^a. \quad (3.208)$$

It is instructive to investigate how $\chi_{\dot{a}}$ and χ^a transform, because these objects are needed to construct terms from spinors, which do not change under Lorentz transformations. Terms like this are incredibly important, because we need them to derive physical laws that are the same in all frames of reference. This will be made explicit in a moment. From the transformation behavior of a left-chiral spinor

$$\chi_L = \chi_a \to \chi_a' = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^b \chi_b, \quad (3.209)$$

we can derive how a spinor with lower, *dotted* index transforms:

$$\chi_L^\star = \chi_a^\star = \chi_{\dot a} \rightarrow \chi'_{\dot a} = (\chi'_a)^\star = \left(\left(e^{i\vec\theta\frac{\vec\sigma}{2} + \vec\phi\frac{\vec\sigma}{2}} \right)_a^{\ b} \right)^\star \chi_{\dot b}^\star$$

$$= \left(e^{-i\vec\theta\frac{\vec\sigma^\star}{2} + \vec\phi\frac{\vec\sigma^\star}{2}} \right)_{\dot a}^{\ \dot b} \chi_{\dot b} \qquad (3.210)$$

Analogously, we use that we know how a right-chiral spinor transforms:

$$\chi_R \rightarrow \chi'_R = \chi'^{\dot a} = \left(e^{i\vec\theta\frac{\vec\sigma}{2} - \vec\phi\frac{\vec\sigma}{2}} \right)^{\dot a}_{\ \dot b} \chi^{\dot b} \qquad (3.211)$$

to derive how a spinor with upper, *undotted* index transforms:

$$\chi_R^\star = (\chi^{\dot a})^\star = \chi^a \rightarrow \chi'^a = (\chi'^{\dot a})^\star = \left(\left(e^{i\vec\theta\frac{\vec\sigma}{2} - \vec\phi\frac{\vec\sigma}{2}} \right)^{\dot a}_{\ \dot b} \right)^\star (\chi^{\dot b})^\star$$

$$= \left(e^{-i\vec\theta\frac{\vec\sigma^\star}{2} - \vec\phi\frac{\vec\sigma^\star}{2}} \right)^a_{\ b} \chi^b. \qquad (3.212)$$

To be able to write products of spinors that do not change under Lorentz transformations, we need one more ingredient. Recall how the scalar product of two vectors is defined: $\vec a \cdot \vec b = \vec a^T \vec b$. In the same spirit we should transpose one of the spinors in a spinor product. We can see this, because at the moment we have the complex conjugate of the Pauli matrices $\sigma_i^\star$ in the exponent, for example, $e^{-i\vec\theta\frac{\vec\sigma^\star}{2}}$. Together with transposing this becomes the Hermitian conjugate: $\sigma_i^\dagger = (\sigma_i^\star)^T$, where the symbol $\dagger$ is called "dagger". The Hermitian conjugate of every Pauli matrix, is again the same Pauli matrix

$$\sigma_i^\dagger = (\sigma_i^\star)^T = \sigma_i, \qquad (3.213)$$

as you can easily check by looking at the explicit form of the Pauli matrices (Eq. 3.80).

By comparing Eq. 3.209 with Eq. 3.212 and using Eq. 3.213, we see that the transformation behavior of a transposed spinor with *lower*, undotted index is exactly the opposite of a spinor with *upper*, undotted index. This means a term of the form $(\chi^a)^T \chi_a$ is invariant (=does not change) under Lorentz transformations, because[154]

[154] As explained in Appendix B.5.5, the symbol δ_b^c is called Kronecker symbol and denotes the unit matrix in index notation. This means $\delta_b^c = 1$ for $b = c$ and $\delta_b^c = 0$ for $b \neq c$.

$$(\chi^a)^T \chi_a \rightarrow (\chi'^a)^T \chi'_a = \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\star}{2} - \vec{\phi}\frac{\vec{\sigma}^\star}{2}} \right)_b^a \chi^b \right)^T \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c \chi_c$$

$$= (\chi^b)^T \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^{\star T}}{2} - \vec{\phi}\frac{\vec{\sigma}^{\star T}}{2}} \right)_b^a \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c \chi_c$$

$$\underbrace{=}_{\text{Eq. 3.213}} (\chi^b)^T \underbrace{\left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^a \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c}_{=\delta_b^c} \chi_c$$

$$= (\chi^c)^T \chi_c. \qquad (3.214)$$

In the same way we can combine an *upper*, dotted index with a *lower*, dotted index as you can verify by comparing Eq. 3.210 with Eq. 3.211. In contrast, a term of the form $(\chi^{\dot{a}})^T \chi_a \hat{=} \chi_R^T \chi_L$ isn't invariant under Lorentz transformations, because

$$\chi_R^T \chi_L = (\chi^{\dot{a}})^T \chi_a \rightarrow (\chi'^{\dot{a}})^T \chi'_a = \chi^{\dot{b}} \underbrace{\left(e^{i\vec{\theta}\frac{\vec{\sigma}^T}{2} - \vec{\phi}\frac{\vec{\sigma}^T}{2}} \right)_{\dot{b}}^{\dot{a}} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c}_{\neq \delta_b^c} \chi_c$$

$$(3.215)$$

Therefore a term combining a left-chiral with a right-chiral spinor is **not** Lorentz invariant. We conclude, we must always combine an upper with a lower index of the same type[155] in order to get Lorentz invariant terms. Or formulated differently, we must combine the Hermitian conjugate of a right-chiral spinor with a left-chiral spinor $\chi_R^\dagger \chi_L = (\chi_R^\star)^T \chi_L \hat{=} (\chi^a)^T \chi_a$, or the Hermitian conjugate of a left-chiral spinor with a right-chiral spinor $\chi_L^\dagger \chi_R = (\chi_L^\star)^T \chi_R = (\chi_{\dot{a}})^T \chi^{\dot{a}}$ to get Lorentz invariant terms. We will use this later, when need invariant terms that we can use to formulate our laws of nature.

In addition, we have now another justification for calling ϵ^{ab} the spinor metric, because the invariant spinor product in Eq. 3.214, can be written as

$$\chi_a^T \chi^a \underbrace{=}_{\text{Eq. 3.204}} \chi_a^T \epsilon^{ab} \chi_b. \qquad (3.216)$$

Compare this to how we defined in Eq. 2.31 the invariant product of Minkowsi space by using the Minkowski metric $\eta^{\mu\nu}$:

$$x_\mu y^\mu = x_\mu \eta^{\mu\nu} y_\nu. \qquad (3.217)$$

The spinor metric is for spinors indeed what the Minkowski metric is for four-vectors[156].

After setting up this notation, we can now write the spinor "met-

[155] In this context dotted $^{\dot{a}}{}_{\dot{a}}$ or undotted $^a{}_a$.

[156] Don't get confused why we have no transposition for the four-vectors here. These equations can be read in two ways. On the one hand as vector equations and on the other hand as component equations. It's conventional and sometimes confusing to use the same symbol x_μ for a four-vector and its components. If we read the equation as a component equation we need no transposition. The same is of course true for our spinor products. Nevertheless, we have seen above that we mustn't forget to transpose and in order to avoid errors we included the explicit superscript T, although the spinor equation here can be read as component equation that do not need it. In contrast, for three component vectors there is a clear distinction using the little arrow: $\vec{a}$ has components a_i.

ric" with lowered indices

$$\epsilon_{ab} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad (3.218)$$

because we need[157] $(-\epsilon)$ to get from χ_R to χ_L. In addition, we can now write the two transformation operators as one object Λ. For example, when it has dotted indices we know it multiplies with a right-chiral spinor and we know which transformation operator to choose:

$$\chi_R \to \chi_R' = \chi'^{\dot{a}} = \Lambda^{\dot{a}}_{\ \dot{b}}\chi^{\dot{b}} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}}\right)^{\dot{a}}_{\ \dot{b}}\chi^{\dot{b}} \qquad (3.219)$$

and analogously for left-chiral spinors

$$\chi_L \to \chi_L' = \chi_a' = \Lambda_a^{\ b}\chi_b = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ b}\chi_b. \qquad (3.220)$$

Therefore:

$$\Lambda_{(\frac{1}{2},0)} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right) \hat{=} \Lambda_a^{\ b} \qquad (3.221)$$

and

$$\Lambda_{(0,\frac{1}{2})} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}}\right) \hat{=} \Lambda^{\dot{a}}_{\ \dot{b}} \qquad (3.222)$$

This notation is useful, because, as we have seen, the two different objects χ_L and χ_R aren't so different after all. In fact we can transform them into each other and a unified notation is the logical result.

Now we move on to the next irreducible representation, which will turn out to be an old acquaintance.

3.7.8   The $\left(\frac{1}{2}, \frac{1}{2}\right)$ Representation

For this representation we use the 2-dimensional representation for both copies of the $SU(2)$ Lie algebra[158] N_i^+ and N_i^-. This time let's have a look at what kind of object our representation is going to act on first. The copies will not interfere with each other, because N_i^+ and N_i^- commute, i.e. $[N_i^+, N_j^-] = 0$ (Eq. 3.172). Therefore, our objects will transform separately under both copies. Let's name the object we want to examine v. This object will have 2 indices $v_a^{\dot{b}}$, each transforming under a separate two-dimensional copy of $\mathfrak{su}(2)$. Here the notation we introduced in the last section comes in handy.

We know that our object v will have 4 components, because each representation is 2 dimensional and this means that both indices can take on two values ($\frac{1}{2}$ and $-\frac{1}{2}$). Therefore, the objects can be 2×2 matrices, but it's also possible to enforce a four component vector form, as we will see[159].

First, let's look at the complex matrix choice. A general 2×2 matrix has 4 complex entries and therefore 8 free parameters. As noted above, we only need 4. We can write every complex matrix M as a sum of a Hermitian ($H^\dagger = H$) and an anti-Hermitian ($A^\dagger = -A$) matrix: $M = H + A$. Both Hermitian and anti-Hermitian matrices have 4 free parameters. In addition, we will see in a moment that our transformations in this representation always transform a Hermitian 2×2 matrix into another Hermitian 2×2 matrix and equivalently an anti-Hermitian matrix into another anti-Hermitian matrix. This means Hermitian and anti-Hermitian matrices are invariant subsets. As explained in Section 3.5 this means that working with a general matrix here, corresponds to having a *reducible* representation. Putting these observations together, we conclude that we can assume that our irreducible representation acts on Hermitian 2×2 matrices. A basis[160] for Hermitian 2×2 matrices is given by the Pauli matrices together with the identity matrix.

Instead of examining $v_a^{\dot{b}}$, we will have a look at $v_{a\dot{b}}$, because then we can use the Pauli matrices as defined in Eq. 3.80. Take note that $v_a^{\dot{b}}$ and $v_{a\dot{b}}$ can be transformed into each other by multiplication with $\epsilon^{\dot{b}\dot{c}}$ and therefore if you want to work with $v_a^{\dot{b}}$, you simply have to use the Pauli matrices that have been multiplied with ϵ.

If we define $\sigma^0 = I_{2\times 2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, we can write

$$v_{a\dot{b}} = v_\nu \sigma^\nu_{a\dot{b}} = v^0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + v^1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + v^2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + v^3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$
$$(3.223)$$

As explained above, we could use[161] $v_a^{\dot{b}} = v_\mu \sigma^\mu_{a\dot{c}} \epsilon^{\dot{b}\dot{c}}$ instead, which means we would use the basis $(\tilde{\sigma}_a^{\dot{b}})^\mu = \sigma^\mu_{a\dot{c}} \epsilon^{\dot{b}\dot{c}}$. We therefore write a general Hermitian matrix as

$$v_{a\dot{b}} = \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix}.$$
$$(3.224)$$

Remember that we have learned in the last section that different indices transform differently. For example, a lower dotted index transforms differently than a lower undotted index.

Now we have a look at how $v_{a\dot{b}}$ transforms and use the transforma-

[160] This means that an arbitrary Hermitian 2×2 matrix can be written as a linear combination of the form: $a_0 1 + a_i \sigma_i$

[161] This is really just a basis choice and here we choose the basis that gives us with our definition of the Pauli matrices, the transformation behavior we derived earlier for vectors.

tion operators that we derived in the last sections

$$
v \to v' = v'_{a\dot{b}} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\star}{2}+\vec{\phi}\frac{\vec{\sigma}^\star}{2}}\right)_{\dot{b}}^{\ \dot{d}}\right)^T
$$
$$
= \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2}+\vec{\phi}\frac{\vec{\sigma}^\dagger}{2}}\right)_{\dot{b}}^{\ \dot{d}}
$$
$$
\underbrace{=}_{\sigma_i^\dagger=\sigma_i} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_{\dot{b}}^{\ \dot{d}}. \qquad (3.225)
$$

We can now see that a Hermitian matrix is after such a transformation still Hermitian, as promised above[162]

$$
\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_{\dot{b}}^{\ \dot{d}} \to \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_{\dot{b}}^{\ \dot{d}}\right)^\dagger
$$
$$
\underbrace{=}_{(ABC)^\dagger=C^\dagger B^\dagger A^\dagger} \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_{\dot{b}}^{\ \dot{d}}\right)^\dagger v_{c\dot{d}}^\dagger \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c}\right)^\dagger
$$
$$
= \left(e^{i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2}+\vec{\phi}\frac{\vec{\sigma}^\dagger}{2}}\right)_{\dot{b}}^{\ \dot{d}} v_{c\dot{d}}^\dagger \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2}+\vec{\phi}\frac{\vec{\sigma}^\dagger}{2}}\right)_a^{\ c}
$$
$$
\underbrace{=}_{\text{if } v_{c\dot{d}}^\dagger = v_{c\dot{d}}} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2}+\vec{\phi}\frac{\vec{\sigma}}{2}}\right)_{\dot{b}}^{\ \dot{d}} \quad \checkmark
$$
$$
(3.226)
$$

The explicit computation for an arbitrary transformation is long and tedious[163] so instead, we will look at one specific example. Let's boost v along the z-axis[164]

$$
v_{a\dot{b}} \to v'_{a\dot{b}} = \left(e^{\phi\frac{\sigma_3}{2}}\right)_a^{\ c} v_{c\dot{d}} \left(e^{\phi\frac{\sigma_3}{2}}\right)_{\dot{b}}^{\ \dot{d}}
$$
$$
= \begin{pmatrix} e^{\frac{\phi}{2}} & 0 \\ 0 & e^{-\frac{\phi}{2}} \end{pmatrix} \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix} \begin{pmatrix} e^{\frac{\phi}{2}} & 0 \\ 0 & e^{-\frac{\phi}{2}} \end{pmatrix}
$$
$$
= \begin{pmatrix} e^{\phi}(v_0 + v_3) & v_1 - iv_2 \\ v_1 + iv_2 & e^{-\phi}(v_0 - v_3) \end{pmatrix}, \qquad (3.227)
$$

where we have used the fact that σ_3 is diagonal[165] and that $e^A = \begin{pmatrix} e^{A_{11}} & 0 \\ 0 & e^{A_{22}} \end{pmatrix}$ holds for every diagonal matrix. We now compare the transformed object that we calculated in Eq. 3.227 with a generic object v':

$$
v'_{a\dot{b}} = \begin{pmatrix} v'_0 + v'_3 & v'_1 - iv'_2 \\ v'_1 + iv'_2 & v'_0 - v'_3 \end{pmatrix} = \begin{pmatrix} e^{\phi}(v_0 + v_3) & v_1 - iv_2 \\ v_1 + iv_2 & e^{-\phi}(v_0 - v_3) \end{pmatrix}.
$$

This tells us how the components of the transformed object are related to the untransformed components[166]

---

[162] Exactly the same computation shows that an anti-Hermitian matrix is still anti-Hermitian after such a transformation. To see this, use in the last step instead of $v_{c\dot{d}}^\dagger = v_{c\dot{d}}$ that $v_{c\dot{d}}^\dagger = -v_{c\dot{d}}$.

[163] See, for example, page 128 in Matthew Robinson. *Symmetry and the Standard Model.* Springer, 1st edition, August 2011. ISBN 978-1-4419-8267-4

[164] This means $\vec{\phi} = (0,0,\phi)^T$. Such a boost is the easiest because σ_3 is diagonal. For boosts along other axes the exponential series must be evaluated in detail.

[165] $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

[166] We rewrite the equations using the connection between the hyperbolic sine, the hyperbolic cosine function and the exponential function $e^{-\phi} = (\cosh(\phi) - \sinh(\phi))$ and $e^{\phi} = (\cosh(\phi) + \sinh(\phi))$, which is conventional in this context. If you are unfamiliar with these functions you can either take notice of their definitions: $\cosh(\phi) \equiv \frac{1}{2}\left(e^{\phi} + e^{-\phi}\right)$ and $\sinh(\phi) \equiv \frac{1}{2}\left(e^{\phi} - e^{-\phi}\right)$ or rewrite the few equations here in terms of e^{ϕ} and $e^{-\phi}$, which is equally good.

$$\rightarrow v'_0 + v'_3 = e^\phi (v_0 + v_3) = (\cosh(\phi) + \sinh(\phi))(v_0 + v_3)$$

$$\rightarrow v'_0 - v'_3 = e^{-\phi}(v_0 - v_3) = (\cosh(\phi) - \sinh(\phi))(v_0 - v_3).$$

The addition and subtraction of both equations yields

$$\rightarrow v'_0 = \cosh(\phi)v_0 + \sinh(\phi)v_3$$
$$\rightarrow v'_3 = \sinh(\phi)v_0 + \cosh(\phi)v_3\,. \qquad (3.228)$$

This is exactly what we get using the 4-vector formalism[167]

[167] See 3.154 for the explicit form of the matrix for a boost along the z-axis.

$$\begin{pmatrix} v'_0 \\ v'_1 \\ v'_2 \\ v'_3 \end{pmatrix} = \begin{pmatrix} \cosh(\phi) & 0 & 0 & \sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(\phi) & 0 & 0 & \cosh(\phi) \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

$$= \begin{pmatrix} \cosh(\phi)v_0 + \sinh(\phi)v_3 \\ v_1 \\ v_2 \\ \sinh(\phi)v_0 + \cosh(\phi)v_3 \end{pmatrix}. \qquad (3.229)$$

This is true for arbitrary Lorentz transformations, as you can check by computing the other possibilities. What we have shown here is that the $(\frac{1}{2}, \frac{1}{2})$ representation *is* the vector representation. We can simplify our transformation laws by using the enforced vector form, because multiplying a matrix with a vector is simpler than the multiplication of three matrices. Nevertheless, we have seen how the familiar 4-vector is related to the more fundamental spinors. A 4-vector **is a rank-2 spinor**, which means a spinor with 2 indices that transforms according to the $(\frac{1}{2}, \frac{1}{2})$ representation of the Lorentz group. Furthermore, **we can now see that 4-vectors aren't appropriate to describe every physical system on a fundamental level, because they aren't fundamental**. There are physical systems they cannot describe.

We can now understand why some people say that "spinors are the square root of vectors". This is meant in the same way as vectors are the square root of rank-2 tensors[168]. A rank-2 tensor has two vector indices and a vector has two spinor indices. Therefore, the most basic object that can be Lorentz transformed is indeed a spinor.

[168] A rank-2 tensor is simply a matrix $M_{\mu\nu}$.

When we started our studies of the Lorentz group, we noted that it consists of four components[169]. These components are connected by the parity and the time-reversal operator[170]. Therefore, to be able to describe all transformations that preserve the speed of light, we need to find the parity and time-reversal transformation for each representation. In this text we restrict ourselves to parity transformations,

[169] Eq. 3.142: $O(1,3) = \{L_+^\uparrow, \Lambda_P L_+^\uparrow, \Lambda_T L_+^\uparrow, \Lambda_P \Lambda_T L_+^\uparrow\}$.

[170] See Eq. 3.142

because the discussion for the time-reversal transformation is very similar. We will discuss in in a later section that nature isn't always symmetric under parity transformations[171].

3.7.9   Spinors and Parity

Up to this point, there is no justification why we call the objects transforming according to the $(\frac{1}{2}, 0)$ representation left-chiral and the objects transforming according to the $(0, \frac{1}{2})$ representation right-chiral. After talking a bit about parity transformation, this will make sense.

Recall that we already know the behavior of the generators of the Lorentz group under parity transformations. The result was Eq. 3.161, which we recite here for convenience

$$J_i \underset{P}{\to} J_i \qquad K_i \underset{P}{\to} -K_i \,. \qquad (3.230)$$

By looking at the definition of the generators $N^{\pm}$ in Eq. 3.169, which we also recite here

$$N_i^{\pm} = \frac{1}{2}(J_i \pm i K_i) \,. \qquad (3.231)$$

we can see that under parity transformations $N^+ \leftrightarrow N^-$. Therefore, the $(0, \frac{1}{2})$ representation of a transformation, becomes the $(\frac{1}{2}, 0)$ representation of this transformation and vice versa under parity transformations. This is the reason for talking about left- and right-chiral spinors[172]. Just as a right-handed coordinate system changes into a left-handed coordinate system under parity transformations, these two representations change into each other.

Rotations are the same for both representations, but boost transformations differ by a sign and it is easy to make the above statement explicit:

$$(\Lambda_{\vec{K}})_{(\frac{1}{2},0)} = e^{\vec{\phi}\vec{K}} \underset{P}{\to} e^{-\vec{\phi}\vec{K}} = (\Lambda_{\vec{K}})_{(0,\frac{1}{2})} \qquad (3.232)$$

$$(\Lambda_{\vec{K}})_{(0,\frac{1}{2})} = e^{-\vec{\phi}\vec{K}} \underset{P}{\to} e^{\vec{\phi}\vec{K}} = (\Lambda_{\vec{K}})_{(\frac{1}{2},0)} \,. \qquad (3.233)$$

We learn here that if we want to describe a physical system that is invariant under parity transformations, we will always need right-chiral **and** left-chiral spinors. The easiest thing to do is to write them below each other into a single object called **Dirac spinor**

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_a \\ \xi^{\dot{a}} \end{pmatrix} \,. \qquad (3.234)$$

Recalling that the generic name for left- and right-chiral spinors is
Weyl spinors, we can say that a Dirac spinor Ψ consists of two Weyl
spinors χ_L and ξ_R. Note that we want to stay general here and don't
assume any a priori connection between χ and ξ. A Dirac spinor of
the form

$$\Psi_M = \begin{pmatrix} \chi_L \\ \chi_R \end{pmatrix} \tag{3.235}$$

is a special case, called Majorana spinor. A Dirac is *not a four-vector*,
because it transforms completely different[173]. A Dirac spinor trans-
forms according to the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation[174] of the Lorentz
group, which means nothing more than writing the corresponding
transformations in block-diagonal form into one big matrix:

$$\Psi \to \Psi' = \Lambda_{(\frac{1}{2},0)\oplus(0,\frac{1}{2})}\Psi = \begin{pmatrix} \Lambda_{(\frac{1}{2},0)} & 0 \\ 0 & \Lambda_{(0,\frac{1}{2})} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \tag{3.236}$$

For example, a boost transformation is in this representation

$$\Psi \to \Psi' = \begin{pmatrix} e^{\frac{\vec{\phi}}{2}\vec{\sigma}} & 0 \\ 0 & e^{\frac{-\vec{\phi}}{2}\vec{\sigma}} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \tag{3.237}$$

It is instructive to investigate how Dirac spinors behave under
parity transformations, because once we know how Dirac spinors
transform under parity transformations, we can check if a given
theory is invariant under such transformations. We can't expect that
a Dirac spinor is after a parity transformation still a Dirac spinor (an
object transforming according to the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation),
because we know that under parity transformations $N^+ \leftrightarrow N^-$ and
therefore

$$\left(0, \frac{1}{2}\right) \underset{P}{\leftrightarrow} \left(\frac{1}{2}, 0\right). \tag{3.238}$$

We conclude that if a Dirac spinor transforms according to the
$(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation, the parity transformed object trans-
forms according to the $(0, \frac{1}{2}) \oplus (\frac{1}{2}, 0)$ representation.

$$\Psi^P \to (\Psi^P)' = \Lambda_{(0,\frac{1}{2})\oplus(\frac{1}{2},0)}\Psi^P = \begin{pmatrix} \Lambda_{(0,\frac{1}{2})} & 0 \\ 0 & \Lambda_{(\frac{1}{2},0)} \end{pmatrix} \begin{pmatrix} \xi_R \\ \chi_L \end{pmatrix}. \tag{3.239}$$

Therefore

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \quad \to \quad \Psi^P = \begin{pmatrix} \xi_R \\ \chi_L \end{pmatrix}. \tag{3.240}$$

[173] Equally, a Majorana spinor is not a
vector.

[174] This a *reducible* representation,
which is obvious because of the
block-diagonal form of the trans-
formation matrix. In contrast, four-
vectors transform according to the
$(\frac{1}{2}, \frac{1}{2}) = (\frac{1}{2}, 0) \otimes (0, \frac{1}{2})$ representation.

A parity transformed Dirac spinor contains the same objects ξ_R, χ_L as the untransformed Dirac spinor, only written differently. A parity transformation does nothing like $\xi_L \to \xi_R$, which is a different kind of transformation we will talk about in the next section.

3.7.10   Spinors and Charge Conjugation

In Section 3.7.7 we stumbled upon a transformation, which yields $\chi_L \to \chi_R$ and $\xi_R \to \xi_L$. The transformation is achieved by $\chi_L \to \chi_L^C = \epsilon \chi_L^\star = \chi_R$ and analogously for a right-chiral spinor $\xi_R \to \xi_R^C = (-\epsilon)\xi_R^\star = \xi_L$. We are now able to understand it from a quite different perspective.

Up to this point, we used this transformation merely as a computational trick in order to raise and lower indices. Now, how does a Dirac spinor transform under such a transformation? Naively we get:

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \to \tilde{\Psi} = \begin{pmatrix} \chi_L^C \\ \xi_R^C \end{pmatrix} = \begin{pmatrix} \chi_R \\ \xi_L \end{pmatrix}. \tag{3.241}$$

Unfortunately, this object does not transform like a Dirac spinor[175], which transform under boosts as follows

$$\Psi \to \Psi' = \begin{pmatrix} e^{\frac{\vec{\theta}}{2}\vec{\sigma}} & 0 \\ 0 & e^{\frac{-\vec{\theta}}{2}\vec{\sigma}} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \tag{3.242}$$

In contrast, the object $\tilde{\Psi}$ we get from the naive operation, transforms as

$$\tilde{\Psi} \to \tilde{\Psi}' = \begin{pmatrix} e^{-\frac{\vec{\theta}}{2}\vec{\sigma}} & 0 \\ 0 & e^{\frac{\vec{\theta}}{2}\vec{\sigma}} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \tag{3.243}$$

This is a different kind of object, because it transforms according to a different representation of the Lorentz group. Therefore we write

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \to \Psi^C = \begin{pmatrix} \xi_R^C \\ \chi_L^C \end{pmatrix} = \begin{pmatrix} \xi_L \\ \chi_R \end{pmatrix}, \tag{3.244}$$

which incorporates the transformation behavior we observed earlier and transforms like a Dirac spinor. This operation is commonly called **charge conjugation**, which can be a little misleading. We know that this transformation transforms a left-chiral spinor into a right-chiral, i.e. flips one label we use to describe our elementary particles[176]. Later we will learn that this operator flips not only one, but all labels we use to describe fundamental particles. One such label is electric charge, hence the name charge conjugation, but before we are able to show this, we need of course to understand first what electric

[175] Unlike for parity transformations, we have a choice here and we prefer to keep working with the same kind of object. The object $\tilde{\Psi}$ can then be seen as a Dirac spinor that has been parity transformed and charge conjugated.

[176] For the more advanced reader: Recall that each Weyl spinor we are talking about here, is in fact a two component object. Later we will define a physical measurable quantity, called spin, that is described by $\frac{1}{2}\sigma_3$. The matrix ϵ, flips an object with eigenvalue $+\frac{1}{2}$ for the spin operator $\frac{1}{2}\sigma_3$ into an object with eigenvalue $-\frac{1}{2}$. This is commonly interpreted as spin flip, which means an object with spin $\frac{1}{2}$, becomes an object with spin $-\frac{1}{2}$.

charge is. Nevertheless, it's always important to remember that all labels get flipped, not only electric charge.

We could now go on and derive higher-dimensional representations of the Lorentz group, but at this point we already have every finite-dimensional irreducible representation we need for the purpose of this text. Nevertheless, there is another representation, the infinite-dimensional representation, that is especially interesting, because we need it to transform fields, like, for example, the electromagnetic field.

3.7.11 Infinite-Dimensional Representations

In the last sections, we talked about finite-dimensional representations of the Lorentz group and learned how we can classify them. These finite-dimensional representations acted on constant one-, two- or four-component objects. In physics the objects we are dealing with are dynamically changing in space and time, so we need to understand how such objects transform. So far, we have dealt with transformations of the form

$$\Phi_a \to \Phi'_a = M_{ab}(\Lambda)\Phi_b \,, \tag{3.245}$$

where $M_{ab}(\Lambda)$ denotes the matrix of the particular finite-dimensional representation of the Lorentz transformation Λ. This means $M_{ab}(\Lambda)$ is a matrix that acts, for example, on a two-component object like a Weyl spinor. The result of the multiplication with this matrix is simply that the components of the object in question get mixed and are multiplied with constant factors. If our object Φ changes in space and time, it is a function of coordinates[177] $\Phi = \Phi(x)$ and these coordinates are affected by the Lorentz transformations, too. In general we have

$$x^\mu \to \Lambda^\mu_\nu x^\nu, \tag{3.246}$$

where Λ^μ_ν denotes the vector representation (= $(\frac{1}{2}, \frac{1}{2})$ representation) of the Lorentz transformation in question. We have in this case[178]

$$\Phi_a(x) \to M_{ab}(\Lambda)\Phi_b(\Lambda x). \tag{3.247}$$

Our transformation will therefore consist of two parts. One part, represented by a finite-dimensional representation, acting on Φ_a and a second part that transforms the spacetime coordinates x. This second part will act on an infinite-dimensional[179] vector space and we therefore need an infinite-dimensional representation. The infinite-dimensional representation of the Lorentz group is given by differential operators[180]

$$M^{\text{inf}}_{\mu\nu} = i(x^\mu \partial^\nu - x^\nu \partial^\mu) \,. \tag{3.248}$$

[177] Here x is a shorthand notation for all spacetime coordinates t, x, y, z

[178] Most books use the Wigner convention for symmetry operators: $\Phi_a(x) \to M_{ab}(\Lambda)\Phi_b(\Lambda^{-1}x)$, but unfortunately there is at this point no way to motivate this convention.

[179] Each component of Φ is now a function of x. The corresponding operators act on $\Phi_a(x)$, i.e. functions of the coordinates and the space of functions is in this context infinite-dimensional. The reason that the space of functions is infinite-dimensional is that we need an infinite number of basis functions. The expansion of an arbitrary function in terms of such an infinite number of basis functions is the idea behind the Fourier transform as explained in Appendix D.1.

[180] The symbols ∂^ν are a shorthand notation for the partial derivative $\frac{\partial}{\partial_\nu}$.

You can check by a straightforward computation that $M_{\mu\nu}^{\text{inf}}$ satisfies the Lorentz algebra (Eq. 3.175) and transforms the coordinates as desired.

The transformation of the coordinates is now given by[181]

$$\Phi'(\Lambda x) = e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}^{\text{inf}}}\Phi(x), \tag{3.249}$$

where the exponential function is, as usual, understood in terms of its series expansion. The complete transformation is then a combination of a transformation generated by the finite-dimensional representation $M_{\mu\nu}^{\text{fin}}$ and a transformation generated by the infinite-dimensional representation $M_{\mu\nu}^{\text{inf}}$ of the generators:

$$\Phi_a(x) \rightarrow \left(e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}^{\text{fin}}}\right)_a^b e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}^{\text{inf}}}\Phi_b(x). \tag{3.250}$$

Because our matrices $M_{\mu\nu}^{\text{fin}}$ are finite-dimensional and constant we can put the two exponents together

$$\Phi_a(x) \rightarrow \left(e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}}\right)_a^b \Phi_b(x) \tag{3.251}$$

with $M_{\mu\nu} = M_{\mu\nu}^{\text{fin}} + M_{\mu\nu}^{\text{inf}}$. This representation of the generators of the Lorentz group is called **field representation**.

We can now talk about a different kind of transformation: **translations**, which means transformations to another location in spacetime. Translations do not mix the components and therefore, we need no finite-dimensional representation. However, it's quite easy to find the infinite-dimensional representation for translations. These are not part of the Lorentz group, but the laws of nature should be location independent. The Lorentz group (boosts and rotations) plus translations is called the **Poincaré group**, which is the topic of the next section. Nevertheless, we will introduce the infinite-dimensional representation of translations already here. For simplicity, we restrict ourselves to one dimension. In this case an **infinitesimal** translation of a function, along the x-axis is given by[182]

$$\Phi(x) \rightarrow \Phi(x+\epsilon) = \Phi(x) + \underbrace{\partial_x\Phi(x)}_{\text{"rate of change" along the x-axis}}\epsilon,$$

which is, of course, the first term of the Taylor series expansion. It is conventional in physics to add an extra $-i$ to the generator and we therefore define

$$P_i \equiv -i\partial_i. \tag{3.252}$$

With this definition an arbitrary, finite translation is[183]

$$\Phi(x) \to \Phi(x+a) = e^{-ia^i P_i} \Phi(x) = e^{-a^i \partial_i} \Phi(x) \, ,$$

where a^i denotes the amount we want to translate in each direction. If we write the exponential function as Taylor series[184], this equation can simply be seen as the Taylor expansion[185] for $\Phi(x+a)$. If we want to transform to another point in time we use $P_0 = i\partial_0$, for a different location we use $P_i = -i\partial_i$.

[184] This is derived in Appendix B.4.1.

[185] This is derived in Appendix B.3.

Now, let's move on to the full spacetime symmetry group of nature: the Poincaré group.

3.8   The Poincaré Group

The Lorentz group includes rotations and boosts. Further transformations that leave the speed of light invariant are translations in space and time, because measuring the speed of light at a different point in spacetime does not change its value. If we add these symmetries to the Lorentz group we get the Poincaré group[186]

[186] The Poincaré group is not the direct, but the semi-direct, sum of the Lorentz group and translations, but for the purpose of this text we can neglect this technical detail.

Poincaré group = Lorentz group **plus** translations

$$= \text{Rotations } \textbf{plus} \text{ boosts } \textbf{plus} \text{ translations} \quad (3.253)$$

The generators of the Poincaré group are the generators of the Lorentz group J_i, K_i plus the generators of translations P_μ.

In terms of J_i, K_i and P_μ the Lie algebra reads[187]

[187] This is not very enlightening, but included for completeness.

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.254)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \quad (3.255)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k \quad (3.256)$$

$$[J_i, P_j] = i\epsilon_{ijk}P_k \quad (3.257)$$

$$[J_i, P_0] = 0 \quad (3.258)$$

$$[K_i, P_j] = i\delta_{ij}P_0 \quad (3.259)$$

$$[K_i, P_0] = -iP_i \, . \quad (3.260)$$

Because this looks like a huge mess it is conventional to write this in terms of $M_{\mu\nu}$, which was defined by

$$J_i = \frac{1}{2}\epsilon_{ijk}M_{jk} \tag{3.261}$$

$$K_i = M_{0i}. \tag{3.262}$$

With $M_{\mu\nu}$ the **Poincaré algebra** reads

$$[P_\mu, P_\nu] = 0 \tag{3.263}$$

$$[M_{\mu\nu}, P_\rho] = i(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu) \tag{3.264}$$

and of course still have

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho}). \tag{3.265}$$

For this quite complicated group it is very useful to label the representations by using the fixed scalar values of the Casimir operators. The Poincaré group has two Casimir operators[188] . The first one is:

$$P_\mu P^\mu =: m^2. \tag{3.266}$$

We give the scalar value the suggestive name m^2, because we will learn later that it coincides with the mass of particles[189].

The second Casimir operator is $W_\mu W^\mu$, where[190]

$$W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}P_\nu M_{\rho\sigma} \tag{3.267}$$

which is called the **Pauli-Lubanski four-vector**. In a lengthy computation it can be justified, that in addition to m, we use the number $j \equiv j_1 + j_2$ as a label. This label is commonly called **spin**. For the moment this is just a name. Later we will understand why the name spin is appropriate. Exactly as for the Lorentz group, we have one j_i for each of the two[191] representations of the $SU(2)$ algebra.

For example, the $(j_1, j_2) = (0, 0)$ representation is called spin 0 representation[192]. The $(j_1, j_2) = (\frac{1}{2}, 0)$ and $(j_1, j_2) = (0, \frac{1}{2})$ are both called spin $\frac{1}{2}$ representations[193] and analogously the $(j_1, j_2) = (\frac{1}{2}, \frac{1}{2})$ representation is called spin 1 representation[194].

The message to take away is that each representation is labelled by two scalar values: m and j. While m can take on arbitrary values, j is restricted to half-integer or integer values.

[188] Recall that a Casimir operator is defined as an operator, constructed from the generators, that commutes with all other generators.

[189] Don't worry, this will make much more sense later.

[190] $\epsilon^{\mu\nu\rho\sigma}$ is the four-dimensional Levi-Civita symbol, which is defined in Appendix B.5.5.

[191] Recall that the complexified Lie algebra of the Lorentz group could be seen to consist of two copies of the Lie algebra of $SU(2)$. The representations of $SU(2)$ could be labelled by a number j and consequently we used for representations of the double cover of the Lorentz group two numbers (j_1, j_2).

[192] $j_1 + j_2 = 0 + 0 = 0$

[193] $j_1 + j_2 = \frac{1}{2} + 0 = 0 + \frac{1}{2} = \frac{1}{2}$

[194] $j_1 + j_2 = \frac{1}{2} + \frac{1}{2} = 1$

3.9   Elementary Particles

The labels for the irreducible representations of double cover of the Poincaré group, **mass** m and by their **spin** $(= j$ here), are how elementary particles are **labeled** in physics[195]. An elementary particle with given labels m and spin, say $j = \frac{1}{2}$, is described by an object, which transforms according to the m, spin $\frac{1}{2}$ representation of the Poincaré group.

More labels, called charges, will follow later from internal symmetries. These labels are used to define an elementary particle. For example, an electron is defined by the following labels

- mass: $9,109 \cdot 10^{-31}$ kg,

- spin: $\frac{1}{2}$,

- electric charge: $1,602 \cdot 10^{-19}$ C,

- weak charge, called weak isospin: $-\frac{1}{2}$,

- strong charge, called color charge: 0.

These labels determine how a given elementary particle behaves in experiments. The representations we derived in this chapter define how we can describe them mathematically. An elementary particle with[196]

- spin 0 is described by an object Φ, called **scalar**, that transforms according to the $(0,0)$, called **spin** 0 **representation** or scalar representation. For example, the Higgs particle is described by a scalar field.

- spin $\frac{1}{2}$ is described by an object Ψ, called **spinor**, that transforms according to the $(\frac{1}{2},0) \oplus (0,\frac{1}{2})$ representation, called **spin** $\frac{1}{2}$ **representation** or spinor representation. For example, electrons and quarks are described by spinors.

- spin 1 is described by an object A, called **vector**, that transforms according to the $(\frac{1}{2},\frac{1}{2})$, called **spin** 1 **representation** or vector representation. For example photons are described by vectors.

This is an incredibly important, deep and beautiful insight, so again:

What we get from deriving the irreducible representations of the Poincaré group are the **mathematical tools we need to describe all elementary particles**. To describe spin 0 particles, like the Higgs Boson, we use mathematical objects, called **scalars**, that transform

[195] Some physicists go even farther and say that elementary particles **are** the irreducible representations of the Poincaré group.

[196] Remember that in the introductory remarks about what we can't derive, it was said there is no real reason to stop here after three representations. We could go on to higher dimensional representations, but there are no elementary particles, for example, with spin $\frac{3}{2}$. Nevertheless, such representations can be used to describe composite objects. In addition, there are many physicists that believe the fundamental particle mediating gravity, called graviton, has spin 2 and therefore a corresponding higher dimensional representation must be used to describe it.

according to the spin 0 representation. To describe **spin $\frac{1}{2}$ particles**
like electrons, neutrinos, quarks etc. we use mathematical objects,
called **spinors**, that transform according to the spin $\frac{1}{2}$ representation.
To describe photons or other particles with **spin** 1 we use objects,
called **vectors**, transforming according to the spin 1 representation.

An explanation for the very suggestive name spin, will be given
in Section 8.5.5, after we talked about how and what we measure in
experiments. We first have to know how we are able to find out if
something is spinning, before we can justify the name spin. At this
point, spin is merely a label.

Further Reading Tips

[197] John Stillwell. *Naive Lie Theory*. Springer, 1st edition, August 2008. ISBN 978-0387782140

[198] Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

[199] Howard Georgi. *Lie Algebras In Particle Physics: from Isospin To Unified Theories (Frontiers in Physics)*. Westview Press, 2 edition, 10 1999. ISBN 9780738202334

- **John Stillwell - Naive Lie Theory**[197] is a very readable, math orientated introduction to Lie Theory.
- **Nadir Jeevanjee - An Introduction to Tensors and Group Theory for Physicists**[198] is a very good introduction, with strong focus on the usage of group theory in physics.
- **Howard Georgi - Lie Algebras In Particle Physics**[199] is a great book to learn more about the parts of Lie group theory that are relevant for particle physics.

3.10   Appendix: Rotations in a Complex Vector Space

The concept of transformations that preserve the inner product can
also be used in complex vector spaces. We want the inner product
of a vector with itself to be a real number, because by definition this
should result in the squared length of the vector. A complex number
would make little sense as the length of the vector. Therefore, the
inner product of complex vector spaces is defined with additional
complex conjugation[200]

[200] Because for $z = a + ib$ we have $z^\star = a - ib$ and therefore $z^\star z = (a + ib)(a - ib) = a^2 + b^2$, which is real.

$$a \cdot a = (a^T)^\star a = a^\dagger a. \tag{3.268}$$

The symbol †, called dagger, denotes Hermitian conjugation, which
means complex conjugation and transposing. We see that a trans-
formation that preserves this inner product must fulfil the condition
$U^\dagger U = 1$:

$$(Ua) \cdot (Ua) = a^\dagger U^\dagger U a = a^\dagger a. \tag{3.269}$$

Transformations like these form groups that are called $U(n)$, where n
denotes the dimensions of the complex vector space and "U" stands

for unitary. Again the groups $SU(n)$ are called "special", because their elements fulfil the additional condition $\det(U) = 1$.

3.11   Appendix: Manifolds

A manifold M is a set of points with a continuous 1-1 map from each open neighborhood onto an open set of R^n. In easy words this means that a manifold M looks **locally** like the standard R^n. This map from each open neighborhood of M onto R^n associates with each point P of M an n-tupel $(x_1(P), ... x_n(P))$ where the numbers $x_1(P), ... x_n(P)$ are called the coordinates of the point P. Therefore another way of thinking about a n-dimensional manifold is that it's a set, which can be given n independent coordinates ins some neighborhood of any point.

An example for a manifold is the surface of a ball. The surface of the three-dimensional ball is called two-sphere S^2 and is defined as the set of points in R^3 for which $x^2 + y^2 + z^2 = r$ holds, where r is the radius of the sphere. Take note that the surface of the three-dimensional ball is two-dimensional, because the definition involves 3 coordinates and one condition, which eliminates one degree of freedom. That is why it's called mathematically two-sphere. To see that the sphere is a manifold we need a map onto R^2. This map is given by the usual spherical coordinates.
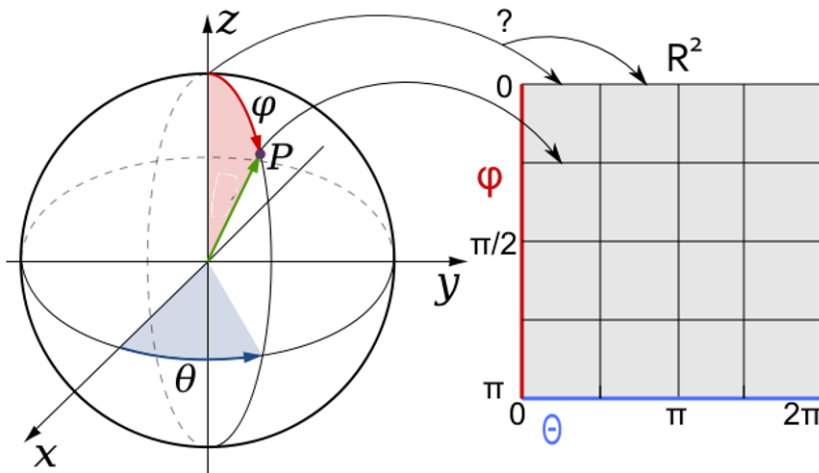
Fig. 3.9: Illustration of the map from one neighborhood of the sphere on to R^n.

Almost all points on the surface of the sphere can be identified unambiguously with a coordinate combination of the form (φ, θ). Almost all! Where is the pole $\varphi = 0$ mapped to? There is no one-to-one identification possible, because the pole is mapped to a whole line,

as indicated in the image. Therefore this map does not work for the complete sphere and we need another map in the neighborhood of the pole to describe things there. A similar problem occurs for the map on the semicircle $\theta = 0$. Each point can be mapped in the R^2 to $\theta = 0$ and $\theta = 2\pi$, which is again not a one-to-one map. This illustrates the fact that for manifolds there is in general not one coordinate system for all points of the manifold, only local coordinates, which are valid in some neighborhood. This is no problem because the defining feature of a manifold is that it looks locally like R^n.

The spherical coordinate map is only valid in the open neighborhood $0 < \varphi < \pi, 0 < \theta < 2\pi$ and we need a second map to cover the whole sphere. We can use, for example, a second spherical coordinate system with different orientation, such that the problematic poles lie at different points for this map and no longer at $\varphi = 0$. With this second map every point of the sphere has a map onto R^2 and the two-sphere can be seen to be a manifold.

A trivial example for a manifold is of course R^n.

4

The Framework

The basic idea of this chapter is that we get the correct equations of nature, by minimizing *something*. What could this *something* be? One thing is for sure: the object shouldn't change under Lorentz transformations, because otherwise we get different laws of nature for different frames of reference. In mathematical terms this means the object we are searching for must be a scalar, which is an object transforming according to the $(0,0)$ representation of the Lorentz group. Together with the restriction to the simplest possible choice this will be enough to derive the correct equations of nature. Nature likes it simple.

Starting with this idea, we will introduce a framework called the **Lagrangian formalism**. By minimizing the central object of the theory we get the equations of motion that describe the physical system in question. The result of this minimization procedure is called the **Euler-Lagrange equations**.

The Lagrangian formalism enables us to derive one of the most-important theorems of modern physics: **Noether's Theorem**. This theorem reveals the deep connection between symmetries and conserved quantities[1]. We will use this connection in the next chapter to understand how the quantities that we measure in experiments can be described by the theory.

4.1 Lagrangian Formalism

The Lagrangian formalism is an incredibly powerful framework[2] that is used in most parts of fundamental physics. It is relatively simple, because the fundamental object, the **Lagrangian**, is a scalar[3]. The formalism is extremely useful if we want to use symmetry considerations. If we demand the **action**, the integral over the Lagrangian, to

[1] A conserved quantity is a quantity that does not change in time. Famous examples are the energy or the momentum of a given system. In mathematical terms this means $\partial_t Q = 0 \to Q =$ const.

[2] There are of course other frameworks, e.g. the Hamiltonian formalism, which has the Hamiltonian as its central object. The problem with the Hamiltonian is that it is not Lorentz invariant, because the energy, it represents, is just one component of the covariant energy-impulse vector.

[3] A scalar is an object transforming according to the $(0,0)$ representation of the Lorentz group. This means that it does not change at all under Lorentz transformations.

be invariant under some symmetry transformation, we ensure that the dynamics of the system in question respect this symmetry.

4.1.1   Fermat's Principle

"Whenever any action occurs in nature, the quantity of action employed by this change is the least possible."

- Pierre de Maupertius[4]

[4] "Recherche des loix du mouvement" (1746)

The basic idea of the Lagrangian formalism emerged from Fermat's principle, which states that light always chooses the path $q(t)$ between two points in space that minimizes the time it takes to travel between the points. To put this into mathematical terms, we define the **action** of a given path $q(t)$ as

$$S_{\text{light}}[\mathbf{q}(t)] = \int dt$$

and our task is then to find the specific path $q(t)$ that minimizes the action[5]. To find the minimum[6] of a given **function** $f(x)$, we take the derivative and set it to zero: $f'(x) \stackrel{!}{=} 0$. The solutions of this equation are the extrema of the functions. Here we want to find the minimum of a **functional** $S[\mathbf{q}(t)]$, which is a function S of a function $\mathbf{q}(t)$. To find the minimum of a functional, we need a new mathematical idea, called **variational calculus**.

[5] The action is here simply the integral over the time for a specific path, i.e. the total time. However, in general the action will be a bit more complicated, as we will see in a moment.

[6] In general, we want to find extrema, which means minimums **and** maximums. The idea outlined in the next section is capable of finding both. Nevertheless, we will continue to talk about minimums.

4.1.2   Variational Calculus - the Basic Idea

If we want to develop a new theory capable of finding the minima of functionals, we need to take a step back and think about what characterises a mathematical minimum. The answer of variational calculus is that a minimum is characterised by its immediate neighbourhood. For example, let's find the minimum x_{min} of an ordinary function $f(x) = 3x^2 + x$. We start by looking at one specific $x = a$ and take a close look at its neighborhood. Mathematically this means $a + \epsilon$, where ϵ denotes an infinitesimal (positive or negative) variation. We put this variation around a into our function $f(x)$:

Fig. 4.1: Variations of a path with fixed starting- and end-point.

$$f(a + \epsilon) = 3(a + \epsilon)^2 + (a + \epsilon) = 3(a^2 + 2a\epsilon + \epsilon^2) + a + \epsilon.$$

Now the crucial observation is that if a is a minimum, first order variations in ϵ must vanish. Otherwise we can choose ϵ to be negative $\epsilon < 0$ and then $f(a + \epsilon)$ is smaller than $f(a)$. Therefore, we collect all terms linear in ϵ and demand that they vanish:

$$3 \cdot 2a\epsilon + \epsilon \stackrel{!}{=} 0 \rightarrow 6a + 1 \stackrel{!}{=} 0.$$

This equation is solved by

$$x_{\min} = a = \frac{-1}{6}\,.$$

This is, of course, exactly the same result that we get if we take the derivative $f(x) = 3x^2 + x \rightarrow f'(x) = 6x + 1$ and demand this to be zero. In terms of ordinary functions this is just another way of doing the same thing. However, variational calculus is additionally able to find the extrema of functionals. We will see in a moment how this works for a general action functional.

The idea of the Lagrangian formalism is that a principle like the Fermat principle for light exists for massive objects, too. Unfortunately, massive objects do not simply obey Fermat's principle. Yet, we can try a more general ansatz

$$S[q(t)] = \int \mathcal{L}\,dt,$$

where $\mathcal{L}$ is, in general, a non-constant function called the Lagrangian. This function happens to be constant for light, but a general Lagrangian is a function of the position $q(t)$ of the object in question. In addition, the Lagrangian can depend on the velocity of the object: $\mathcal{L} = \mathcal{L}(q(t), \frac{\partial}{\partial t}q(t))$. This will be discussed in more detail in the next section[7]. Before we take a closer look at the usage of the variational calculus idea for a functional like this, we need to talk about two small things.

4.2   Restrictions

As already noted in Chapter 1.1 there are restrictions to our present theories, that, so far, we can't motivate from first principles. We only know that we must respect them to get a sensible theory.

One important restriction is that we are only allowed to use the lowest, non-trivial derivatives in the Lagrangian. Trivial in this context means with no influence of the dynamics of the system, i.e. on the equations of motion. For some theories this will be first order and for others second order derivatives. The lowest order of a given theory is determined by the condition that the Lagrangian must be Lorentz invariant[8], because otherwise we would get different equations of motions for different frames of reference. For some theories we can't get an invariant term with first order derivatives and therefore second order derivatives are the lowest possible order.

We simply do not know how to work with theories including higher order derivatives and there are deep systematic problems

[7] Our task will be to find the path $q(t)$ with lowest possible action for a given Lagrangian and given initial conditions. Before we are able to do that, we need to find the correct Lagrangian, describing the physical system in question. Here is where the symmetries we talked about in the last chapters come in handy. Demanding that the Lagrangian is invariant under all transformations of the Lorentz group, will lead us to the correct Lagrangians.

[8] In fact, the action must be Lorentz invariant, but if the Lagrangian is Lorentz invariant, the action certainly is, too.

with such theories[9]. In addition, higher order derivatives in the La-
grangian lead to higher order derivatives in the equations of motion
and therefore more initial conditions would be required.

It is sometimes claimed that the constraint to first order derivatives
is a consequence of our demand to get a local[10] theory, but this only
rules out an **infinite** number of derivatives. A non-local interaction is
of the form[11]

$$\Phi(x-h)\Phi(x), \tag{4.1}$$

that is, two fields interacting with each other at two different points
in spacetime with arbitrary distance h. Using the Taylor expansion
we can write

$$\Phi(x-h) = \sum_{k=0}^{\infty} \left(\left(\frac{\partial}{\partial x} \right)^k \Phi(x) \Big|_{x=h} \right) \frac{(h)^k}{k!}, \tag{4.2}$$

which shows that allowing an infinite number of derivatives would
result in a non-local interaction theory.

Another restriction is that in order to get a theory describing free
(=non-interacting) fields/particles we must stop at second order in
the field $\Phi(x)$. This means we only include the terms[12]

$$\Phi^0, \Phi^1, \Phi^2$$

into our considerations. For example, a term of the form $\Phi^2 \partial_\mu \Phi$ is of
third order in Φ and therefore not included in the Lagrangian for our
free theory.

4.3 Particle Theories vs. Field Theories

Currently we have two frameworks to describe nature. On the one
hand, we have **particle theories** that describe physical systems in
terms of positions of particles depending on time, i.e. $\vec{q} = \vec{q}(t)$. We
use the symbol q instead of x, because there is no need to describe
nature in terms of Cartesian coordinates[13]. For such theories the
Lagrangian depends on the position $\vec{q}$, the velocity $\partial_t \vec{q}$ and the time t:

$$\mathcal{L} = \mathcal{L}(\vec{q}, \partial_t \vec{q}, t). \tag{4.3}$$

A famous example is the Lagrangian $\mathcal{L} = \frac{1}{2}m\dot{\vec{q}}^2$, from which we
can derive Newton's equation of motion for classical mechanics. This
will be discussed later in great detail.

On the other hand, we have **field theories** that do not use the
location $\vec{q}(t)$ of individual particles to describe nature, but fields[14]. In

such theories space and time form the stage the fields $\Phi(\vec{x},t)$ act on. Using the restrictions we get[15]

$$\mathscr{L} = \mathscr{L}(\Phi(\vec{x},t), \partial_\mu \Phi(\vec{x},t), \vec{x},t) \tag{4.4}$$

A famous example is the Lagrangian $\mathscr{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)$, from which we will derive the Klein-Gordon equation.

The big advantage of field theories is that they treat space and time on an equal footing. In a particle theory we use the space coordinates $\vec{q}(t)$ to describe our particle as a function of time. Notably, there is no term like $\partial_{\vec{q}} t$ in the Lagrangian and if it would be there, how should we interpret it? It is clear what we mean when we talk about the location of a particle, but there is no obvious way to make sense of a similar sentence for time.

Having discussed all this, we are able to return to the original minimizing problem, proposed at the beginning of this chapter. We want to find the minimum of some functional

$$S[q(t)] = \int \mathcal{L} dt,$$

which will give us the correct equations of motion.

The solutions of these equations are for particles the correct paths that minimize the functional. For a field theory the solutions are the correct field configurations.

For the moment, do not worry about the object $\mathcal{L}$, because we will describe in the following chapters in great detail how we can derive the correct Lagrangian $\mathcal{L}$ for the systems in question. Here, we work with a general $\mathcal{L}$ and use the machinery of variational calculus, introduced above, to derive the minimum of the functional $S[q(t)]$. This minimization procedure will give us the equations of motion for the system.

4.4   Euler-Lagrange Equation

We start with particle theory and this means, we want to find out how a particle moves between two fixed points. The mathematical problem we have to solve is to find the **function** $q(t)$ for which the **action**

$$S = \int_{t_1}^{t_2} L\left(q(t), \frac{dq(t)}{dt}, t\right) dt$$

is an **extremum** (maximum or minimum). We use the notation

$$\dot{q}(t) \equiv \frac{dq(t)}{dt}.$$

[15] The usage of a different symbol $\mathscr{L}$ here is intentional, because in field theories we will work most of the time with the Lagrangian density $\mathscr{L}$. The Lagrangian density is related to the Lagrangian via $\mathcal{L} = \int d^3 x \mathscr{L}$.

Analogous to the example above, we fix $q(t) = a(t)$ and vary this particle path a little bit

$$a(t) + \epsilon(t)$$

where ϵ is again infinitesimal. Considering a particle, we are not only able to vary the path, but in the same way we can vary the velocity $\dot{a}(t) + \dot{\epsilon}(t)$ with $\dot{\epsilon}(t) = \frac{d\epsilon(t)}{dt}$.

At the boundary, the transformed path must be equivalent to the untransformed path:

$$0 = \epsilon(t_1) = \epsilon(t_2), \tag{4.5}$$

because we are searching for the path **between two fixed points** that extremizes the action integral.

This variation of the fixed function results in the functional

$$S = \int_{t_1}^{t_2} dt \mathcal{L}(q + \epsilon, \dot{q} + \dot{\epsilon}, t).$$

Analogous to the example above, where we searched for the minimum of a function, we will demand here that the terms linear in the variation ϵ vanish. Because we work with a general $\mathcal{L}$, we expand it as Taylor series[16] and demand that the first order terms vanish

$$\int_{t_1}^{t_2} dt \left[\epsilon(t) \frac{\partial \mathcal{L}}{\partial q} + \left(\frac{d}{dt} \epsilon(t) \right) \frac{\partial \mathcal{L}}{\partial \dot{q}} \right] \overset{!}{=} 0. \tag{4.6}$$

To proceed further, we need to find a way to bracket $\epsilon(t)$ out. This is necessary, because $\epsilon(t)$ is an arbitrary variation and when we bracket it out, we know that everything inside the brackets has to vanish if the total result is zero. So far we have one term that gets multiplied with $\epsilon(t)$ and a second term that gets multiplied with $\frac{d}{dt}\epsilon(t)$. To bracket $\epsilon(t)$ out, we need get rid of the derivative. To achieve this, we integrate the last term by parts[17], which yields

$$\int_{t_1}^{t_2} dt \left(\frac{d}{dt} \epsilon(t) \right) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}}$$

$$= \epsilon(t) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} dt \epsilon(t) \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right).$$

Using Eq. 4.5 we have

$$\epsilon(t) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \Big|_{t_1}^{t_2} = 0.$$

So, we can rewrite Eq. 4.6 as

$$\int_{t_1}^{t_2} dt \epsilon(t) \left[\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial q} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right) \right] \overset{!}{=} 0$$

[16] We are using a generalization of the formula, derived in Appendix B.3 for functions with more than one variable:
$\mathcal{L}(q + \epsilon, \dot{q} + \dot{\epsilon}, t) = \mathcal{L}(q) + (q + \epsilon - q) \frac{\partial \mathcal{L}}{\partial q}$
$+ (\dot{q} + \dot{\epsilon} - \dot{q}) \frac{\partial \mathcal{L}}{\partial \dot{q}} + \dots$
$= \mathcal{L}(q) + \epsilon \frac{\partial \mathcal{L}}{\partial q} + \dot{\epsilon} \frac{\partial \mathcal{L}}{\partial \dot{q}} + \dots$

[17] Integration by parts is a direct consequence of the product rule and derived in Appendix B.2.

and we can see that this only vanishes for arbitrary variations $\epsilon(t)$ if the expression in the square bracket [ ] is identically zero. This yields[18]

$$\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial q} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}}\right) = 0\,, \tag{4.7}$$

which is the famous **Euler-Lagrange equation**.

For a field theory we can follow a similar route. First notice that in this case we treat time and space equally. Therefore, we introduce the Lagrangian density $\mathscr{L}$:

$$\mathcal{L} = \int d^3x \, \mathscr{L}(\Phi^i, \partial_\mu \Phi^i) \tag{4.8}$$

and we can write the action in terms of this Lagrangian density

$$S = \int dt \mathcal{L} = \int d^4x \, \mathscr{L}(\Phi^i, \partial_\mu \Phi^i)\,. \tag{4.9}$$

Following the same steps as above, we get the equations[19] of motion for a field theory:

$$\frac{\partial \mathscr{L}}{\partial \Phi^i} - \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi^i)}\right) = 0\,. \tag{4.10}$$

In the next section we will derive, using the Lagrangian formalism, one of the most important theorems of modern physics. We need this theorem to see the deep connection between symmetries and conserved quantities. The conserved quantities are the appropriate quantities[20] to describe nature and from this theorem we learn how we can work with them in a theoretical context.

4.5   Noether's Theorem

Noether's theorem shows that each symmetry of the Lagrangian is directly related to one conserved quantity. Or formulated slightly different: The notions physicists commonly use to describe nature (the conserved quantities) are directly connected to symmetries. This is one of the most beautiful insights in the history of science.

4.5.1   Noether's Theorem for Particle Theories

Let's have a look at what we can say about conserved quantities in particle theories. We will restrict to continuous symmetries, because then we can look at infinitesimal changes. As we have seen in earlier chapters, we can build up finite changes by repetition of infinitesimal

[18] Maybe you wonder about the two different symbols for derivatives. $\frac{d}{dt}$ is called the total derivative with respect to t, whereas $\frac{\partial}{\partial t}$ is called the partial derivative. The total derivative gives us the total change, which is for a function f given by sum of the change rates, also known as partial derivatives, times the change of the quantity itself. For example, the total change of a function $f(x(t), y(t), z(t))$ in three dimensional space is given by $\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial t} + \frac{\partial f}{\partial t}\frac{\partial t}{\partial t}$. The change rate times the distance it is changed. In contrast the partial derivative is just one part of this total change. For a function that does not explicitly depend on t the partial derivative is zero. For example, for $f(x(t), y(t)) = x^2y + y^3$, we have $\frac{\partial f}{\partial t} = 0$, but $\frac{\partial f}{\partial x} = 2xy \neq 0$ and $\frac{\partial f}{\partial y} = x^2 + 3y^2 \neq 0$. Therefore, $\frac{df}{dt} = 2xy\frac{\partial x}{\partial t} + (x^2 + 3y^2)\frac{\partial y}{\partial t}$. In contrast for another function $g(x(t), y(t), t) = x^2t + y$ we have $\frac{\partial g}{\partial t} = x^2$.

[19] Plural because we get one equation for each field component, i.e. $\Phi_1, \Phi_2, \ldots$.

[20] We can see them as anchors in an otherwise extremely complicated world. While everything changes, the conserved quantities stay the same.

changes. The invariance of the Lagrangian under an infinitesimal transformation[21] $q \to q' = q + \delta q$ can be expressed mathematically

$$\delta \mathcal{L} = \mathcal{L}\left(q, \frac{dq}{dt}, t\right) - \mathcal{L}\left(q + \delta q, \frac{d(q + \delta q)}{dt}, t\right)$$

$$= \mathcal{L}\left(q, \frac{dq}{dt}, t\right) - \mathcal{L}\left(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t\right) \overset{!}{=} 0. \qquad (4.11)$$

> [21] The symbol for a small variation δ may not be confused with the symbol for partial derivatives ∂.

Demanding that the Lagrangian is invariant can be too restrictive. What really needs to be invariant for the dynamics to stay the same is the action and not the Lagrangian. Of course, if the Lagrangian is invariant, the action is automatically invariant:

$$\delta S = \int dt \mathcal{L}\left(q, \frac{dq}{dt}, t\right) - \int dt \mathcal{L}\left(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t\right) = \int dt \delta \mathcal{L} \underbrace{=}_{\text{if } \delta\mathcal{L}=0} 0.$$
$$(4.12)$$

How can the Lagrangian change, while the action stays the same? It turns out we can always add the total time derivative of a function G to the Lagrangian

$$\mathcal{L} \to \mathcal{L} + \frac{dG}{dt}$$

without changing the action, because[22]

> [22] Using $\delta G = \frac{\partial G}{\partial q}\delta q$, because $G = G(q)$ and we vary q. Therefore the variation of G is given by the rate of change $\frac{\partial G}{\partial q}$ times the variation of q.

$$\delta S \to \delta S' = \delta S + \int_{t1}^{t2} dt \frac{d}{dt}\delta G = \delta S + \underbrace{\frac{\partial G}{\partial q}\delta q\Big|_{t1}^{t2}}_{=0 \text{ because } \delta q(t_1)=\delta q(t_2)=0}.$$

In the last step, we use that the variation δq vanishes at the initial and final times (t_1, t_2). We conclude, there is no need for us to demand that the variation of the Lagrangian $\delta\mathcal{L}$ vanishes, rather we have the less restrictive condition

$$\delta\mathcal{L} \overset{!}{=} \frac{dG}{dt}. \qquad (4.13)$$

This means the Lagrangian can change in this way without changing the action. Therefore the equation of motion stay the same even though the Lagrangian changes, as long as the change can be written as total derivative of some function: $\frac{dG}{dt}$. Rewriting Eq. 4.11, now with $\frac{dG}{dt}$ instead of 0 on the right hand side yields

$$\delta\mathcal{L} = \mathcal{L}\left(q, \frac{dq}{dt}, t\right) - \mathcal{L}\left(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t\right) \overset{!}{=} \frac{dG}{dt} \qquad (4.14)$$

We expand the second term as Taylor series and keep only terms that are first order in δq, because δq is infinitesimal, and use the notation $\frac{dq}{dt} = \dot{q}$:

$$\to \delta\mathcal{L} = \mathcal{L} - \mathcal{L} - \frac{\partial\mathcal{L}}{\partial q}\delta q - \frac{\partial\mathcal{L}}{\partial\dot{q}}\delta\dot{q} \overset{!}{=} \frac{dG}{dt}$$

$$\rightarrow \delta \mathcal{L} = -\frac{\partial \mathcal{L}}{\partial q} \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \stackrel{!}{=} \frac{dG}{dt} . \qquad (4.15)$$

We can rewrite Eq. 4.15 by using the Euler-Lagrange equation[23]. This yields

[23] Eq. 4.7: $\frac{\partial \mathcal{L}}{\partial q} = \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right)$

$$\rightarrow \delta \mathcal{L} = -\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} = \frac{dG}{dt} .$$

This can be rewritten, using the product rule[24]

[24] If you're unsure about the product rule, have a look at Appendix B.1.

$$\rightarrow \delta \mathcal{L} = -\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q \right) = \frac{dG}{dt}$$

$$\rightarrow \frac{d}{dt} \underbrace{\left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q + G \right)}_{\equiv J} = 0. \qquad (4.16)$$

Therefore, we have found a quantity J that is conserved in time:

$$J = \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q + G, \qquad (4.17)$$

because we have

$$\frac{d}{dt} J = 0 \rightarrow J = \text{const.}$$

To illustrate this, we will use one later result from Section 10.2: Newton's second law for a free particle with constant mass[25]

$$m \ddot{\vec{q}} = 0. \qquad (4.18)$$

[25] If q denotes the position of some object, $\frac{dq}{dt} = \dot{q}$ is the velocity of the object and $\frac{d}{dt} \frac{dq}{dt} = \frac{d^2 q}{dt^2} = \ddot{q}$ the acceleration.

The Lagrangian that reproduces this famous equation of motion is[26]

$$\mathcal{L} = \frac{1}{2} m \dot{\vec{q}}^2 \qquad (4.19)$$

[26] We work now with more than one spatial dimension, which means instead of q and a, we use vectors $\vec{q}$ and $\vec{a}$.

as you can check, by putting it into the Euler-Lagrange equation (Eq. 4.7).

Let us compute for different symmetries of this Lagrangian the corresponding conserved quantities. Our Lagrangian is invariant ($\delta \mathcal{L} = 0$) under spatial translations $\vec{q} \rightarrow \vec{q}' = \vec{q} + \vec{a}$, where $\vec{a}$ denotes some constant vector, because $\mathcal{L} = \frac{1}{2} m \dot{\vec{q}}^2$ does not depend on $\vec{q}$. The corresponding conserved quantity reads[27]

[27] Here $G = 0$, which certainly fulfils the condition in Eq. 4.13.

$$J_{\text{trans}} = \frac{\partial \mathcal{L}}{\partial \dot{\vec{q}}} \vec{a} = m \dot{\vec{q}} \vec{a} = \vec{p} \vec{a}, \qquad (4.20)$$

where $\vec{p} = m \dot{\vec{q}}$ is what we usually call **momentum** in classical mechanics. The equation $\frac{d}{dt} J = 0$ holds for arbitrary $\vec{a}$ and therefore **momentum is conserved, because the Lagrangian is invariant under spatial translations.**

We now want to look at rotations and therefore need more than one dimension, because a rotation in one dimension makes no sense. It is useful to use an index notation instead of the usual vector notation here, because then we can simply rewrite all equations with $q \to q_i$ to be able to incorporate changes in more than one dimension. Let's have a look at an infinitesimal rotation[28] $q_i \to q_i' = q_i + \epsilon_{ijk} q_j a_k$ and therefore $\delta q_i = \epsilon_{ijk} q_j a_k$. Our Lagrangian is invariant under such transformations, because again, $\mathcal{L} = \frac{1}{2} m \dot{\vec{q}}^2$ does not depend on q, and the corresponding conserved quantity is[29]

[28] Here we write the generator of rotations by using the Levi-Civita symbol. This was explained in the text below Eq. 3.62.

[29] This was derived in Eq. 4.16, again we work here with $G = 0$.

$$
\begin{aligned}
J_{\text{rot}} &= \frac{\partial \mathcal{L}(q_i, \dot{q}_i, t)}{\partial \dot{q}_i} \delta q_i = \frac{\partial \mathcal{L}(q_i, \dot{q}_i, t)}{\partial \dot{q}_i} \epsilon_{ijk} q_j a_k \\
&= m \dot{q}_i \epsilon_{ijk} q_j a_k = p_i \epsilon_{ijk} q_j a_k \\
&\to J_{\text{rot}} = (\vec{p} \times \vec{q}) \cdot \vec{a} \equiv \vec{L} \cdot \vec{a} .
\end{aligned}
\tag{4.21}
$$

In the last step we rewrite our term in vector notation, where $\times$ is called the cross product and $\vec{L}$ is what we usually call **angular momentum** in classical mechanics. **Therefore, invariance under rotations leads us to conservation of angular momentum.**

Next let's have a look at invariance under time translations[30]. An infinitesimal time displacement $t \to t' = t + \epsilon$ has the effect

[30] This means that physics doesn't care about if we perform an experiment yesterday, today or in 50 years, given the same initial conditions, the physical laws stay the same.

$$
\delta \mathcal{L} = \mathcal{L}\left(q(t), \frac{dq(t)}{dt}, t\right) - \mathcal{L}\left(q(t+\epsilon), \frac{dq(t+\epsilon)}{dt}, t+\epsilon\right)
$$

$$
= -\frac{\partial \mathcal{L}}{\partial q} \frac{\partial q}{\partial t} \epsilon - \frac{\partial \mathcal{L}}{\partial \dot{q}} \frac{\partial \dot{q}}{\partial t} \epsilon - \frac{\partial \mathcal{L}}{\partial t} \epsilon \underbrace{=}_{\text{this is exactly the total derivative}} -\frac{d\mathcal{L}}{dt} \epsilon,
\tag{4.22}
$$

which tells us that in general $\delta \mathcal{L} \neq 0$, but the condition in Eq. 4.13 is fulfilled anyway with $G = -\mathcal{L}$.

We can put this into Eq. 4.16, which yields

$$
\frac{d}{dt} \underbrace{\left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - \mathcal{L} \right)}_{\equiv \mathcal{H}} = 0 .
\tag{4.23}
$$

The conserved quantity $\mathcal{H}$ is called the **Hamiltonian** and represents the total energy of the system. For our example Lagrangian, we have

$$
\mathcal{H} = \frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - \mathcal{L} = \underbrace{\left(\frac{\partial}{\partial \dot{q}} \frac{1}{2} m \dot{q}^2 \right)}_{= m\dot{q}} \dot{q} - \frac{1}{2} m \dot{q}^2
$$

$$
= m\dot{q}^2 - \frac{1}{2} m \dot{q}^2 = \frac{1}{2} m \dot{q}^2,
\tag{4.24}
$$

which is exactly the kinetic energy of our system. The kinetic energy is the total energy, because we worked without a potential/external

force. The Lagrangian that leads to Newton's second law for a particle in an external potential: $m\ddot{q} = \frac{dV}{dq}$ is

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2 - V(q).$$

The Hamiltonian is then

$$\mathcal{H} = \frac{1}{2}m\dot{q}^2 + V$$

which is the correct total energy=kinetic energy+potential energy.

The conserved quantity that follows from **boost invariance** is rather strange and the corresponding computation can be found in the appendix Section 4.6. The resulting conserved quantity is

$$\tilde{J}_{\text{boost}} = \underbrace{pt - \frac{1}{2}mvt}_{\equiv \tilde{p}t} - mq = \tilde{p}t - mq. \tag{4.25}$$

We can see that this quantity depends on the starting time and by choosing the starting time appropriately we can make it zero. Because this quantity is conserved, this conservation law tells us that zero stays zero for all times.

To summarize for particle theories we have the following connections:

- translational invariance in space $\Rightarrow$ conservation of momentum,

- boost invariance[31] $\Rightarrow$ conservation of $\tilde{p}t - mq$,

- rotational invariance $\Rightarrow$ conservation of angular momentum,

- translational invariance in time $\Rightarrow$ conservation of energy.

Noether's theorem shows us why these notions[32] are used in every physical theory of nature in one or another form. As long as we have the usual spacetime symmetries of our physical laws, we have momentum, energy and angular momentum as conserved quantities.

It is instructive to repeat the above discussion for field theories. In field theories we have two kinds of symmetries. On the one hand, our Lagrangian can be invariant under spacetime transformations, which means a transformation like a rotation. On the other hand, we can have invariance under transformations of the field itself, which are called **internal symmetries**.

4.5.2   Noether's Theorem for Field Theories - Spacetime Symmetries

For fields one has to distinguish between different kinds of changes that can happen under spacetime transformations. Observer S' sees

[31] Another name for a boost is translation in momentum space, because the transformation $q \to q + vt$, changes the momentum to $m\dot{q} \to m(\dot{q} + v)$.

[32] Except for the conserved quantity following from boost invariance.

[33] Remember for example, that a Weyl
spinor has two- and a vector field
has four-components. If we look at

[33] Remember for example, that a Weyl
spinor has two- and a vector field
has four-components. If we look at
the vector field $A_\mu = \begin{pmatrix} A_0 \\ A_1 \\ A_2 \\ A_3 \end{pmatrix}$ from a
different perspective, i.e. describe it in
a rotated coordinate system it can look
like $A'_\mu = \begin{pmatrix} A'_0 \\ A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} A_0 \\ -A_2 \\ A_1 \\ A_3 \end{pmatrix}$. A'_μ and
A_μ describe the same field in coordinate
systems that are rotated by 90° around
the z-axis relative to each other.

[34] If this is new to you: This is often
called the total derivative. The total
change is given by the sum of the
change rates, also known as derivatives,
times the change of the quantity itself.
For example, the total change of a
function $f(x, y, z)$ in three dimensional
space is given by $\frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial z}\delta z$.
The change rate times the distance it
is changed. We consider infinitesimal
changes and therefore this can be
seen as the first terms in the Taylor
expansion, where we can neglect higher
order terms.

[35] See Eq. 4.10: $\frac{\partial \mathscr{L}}{\partial \Phi} = \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Phi)} \right)$

the field $\Psi'(x')$ whereas observer S sees the field $\Psi(x)$. This is the
same field, just from another perspective and the two observers do
not see the same numerical field components. The two different
descriptions are related by the appropriate transformations of the
Lorentz group. We introduced in Section 3.7.11 the field representa-
tion, which we will be using now. The (infinite-dimensional) differ-
ential operator representation changes $x \to x'$. This means by using
this representation we can compute the field components at a differ-
ent point in spacetime or in a rotated frame. The finite-dimensional
representation of the Lorentz group changes $\Psi \to \Psi'$, i.e. mixes the
field components[33].

A complete transformation, for a field that depends on spacetime,
needs to consist of both parts. We will look at these parts separately,
starting with the change $x \to x'$. For rotations the conserved quantity
that follows is not really conserved, because we neglected the second
part of the transformation, i.e. the mixing of the field components.
Only the sum of the two conserved quantities that follow from $x \to$
x' and $\Psi \to \Psi'$ is conserved.

To make this more concrete consider a general Lagrangian density
$\mathscr{L}((\Phi(x_\mu), \partial_\mu\Phi(x_\mu), x_\mu)$. Symmetry means we have

$$\mathscr{L}((\Phi(x_\mu), \partial_\mu\Phi(x_\mu), x_\mu) = \mathscr{L}((\Phi'(x'_\mu), \partial_\mu\Phi'(x'_\mu), x'_\mu). \tag{4.26}$$

In general, the total change of a function-of-a-function, when the
independent functions are changed and the point at which they are
evaluated is also changed, is given by[34]

$$df(g(x), h(x), ...) = \frac{\partial f}{\partial g}\delta g + \frac{\partial f}{\partial h}\delta h + ... + \frac{\partial f}{\partial x}\delta x. \tag{4.27}$$

Applying this to the Lagrangian yields

$$\delta\mathscr{L} = \frac{\partial \mathscr{L}}{\partial \Phi}\delta\Phi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu\Phi)}\delta(\partial_\mu\Phi) + \frac{\partial \mathscr{L}}{\partial x_\mu}\delta x_\mu, \tag{4.28}$$

which we can rewrite using the Euler-Lagrange equations[35]

$$\delta\mathscr{L} = \partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu\Phi)}\right)\delta\Phi + \frac{\partial \mathscr{L}}{\partial(\partial_\mu\Phi)}\underbrace{\delta(\partial_\mu\Phi)}_{=\partial_\mu\delta\Phi} + \frac{\partial \mathscr{L}}{\partial x_\mu}\delta x_\mu$$

$$\underbrace{=}_{\text{Product rule}}\partial_\mu\left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu\Phi)}\delta\Phi\right) + \frac{\partial \mathscr{L}}{\partial x_\mu}\delta x_\mu. \tag{4.29}$$

The variation $\delta\Phi$ has now two parts

$$\delta\Phi = \epsilon_{\mu\nu}S^{\mu\nu}\Phi(x) - \frac{\partial\Phi(x)}{\partial x_\mu}\delta x_\mu, \tag{4.30}$$

with the transformation parameters $\epsilon_{\mu\nu}$, the transformation operator $S^{\mu\nu}$ in the corresponding finite-dimensional representation and a conventional minus sign. $S_{\mu\nu}$ is related to the generators of rotations by $J_i = \frac{1}{2}\epsilon_{ijk}S_{jk}$ and to the generators of boosts by $K_i = S_{0i}$, analogous to the definition of $M_{\mu\nu}$ in Eq. 3.173. This definition of the quantity $S_{\mu\nu}$ enables us to work with the generators of rotations and boosts at the same time.

The first part is only important for rotations and boosts, because translations do not mix the field components. For boosts the conserved quantity will not be very enlightening, just as in the particle case, so in fact this term will become only relevant for rotational symmetry.

Let's start with the simplest field transformation: a translation in spacetime

$$x_\mu \to x'_\mu = x_\mu + \delta x_\mu = x_\mu + a_\mu. \tag{4.31}$$

For translations we have $\epsilon_{\mu\nu} = 0$, because field components do not mix under translations and therefore Eq. 4.30 reads

$$\delta\Phi = -\frac{\partial\Phi}{\partial x_\mu}\delta x_\mu\,.$$

Thus if we want to investigate the consequences of invariance ($\delta\mathscr{L} = 0$), we get from Eq. 4.29

$$-\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial\Phi}{\partial x_\mu}\delta x_\mu\right) + \frac{\partial\mathscr{L}}{\partial x_\mu}\delta x_\mu = 0 \tag{4.32}$$

$$\to -\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial\Phi}{\partial x_\mu} - \delta_\mu^\nu\mathscr{L}\right)\delta x^\mu = 0\,. \tag{4.33}$$

From Eq. 4.31 we have $\delta x_\mu = a_\mu$, which we now put into Eq. 4.33. This yields

$$-\partial_\nu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial(\Phi)}{\partial x_\mu} - \delta_\mu^\nu\mathscr{L}\right)a^\mu = 0, \tag{4.34}$$

and this motivates us to define the energy-momentum tensor

$$T_\mu^\nu := \frac{\partial\mathscr{L}}{\partial(\partial_\nu\Phi)}\frac{\partial(\Phi)}{\partial x_\mu} - \delta_\mu^\nu\mathscr{L}. \tag{4.35}$$

Equation 4.34 tells us that each component μ of T_μ^ν fulfils a continuity equation, because a^μ is arbitrary[36]

$$\partial_\nu T_\mu^\nu = 0 \tag{4.36}$$

$$\to \partial_\nu T_\mu^\nu = \partial_0 T_\mu^0 + \partial_i T_\mu^i = 0\,. \tag{4.37}$$

[36] Take note that we have here a plus sign $\partial_\nu T_\mu^\nu = \partial_0 T_\mu^0 + \partial_i T_\mu^i$, although we have $A_\mu B^\mu = A_\mu \eta^{\mu\nu} B_\nu = A_0 B_0 - A_i B_i$. We defined a four-vector with a lower index without minus signs $A_\mu = (A_0, A_1, A_2, A_3)^T$ and a four-vector with an upper index as $B^\mu \equiv \eta^{\mu\nu}B_\nu = (B_0, -B_1, -B_2, -B_3)^T$. The minus signs here are a result of the minus signs in the Minkowski metric $\eta^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$. The reason for the plus sign in Eq. 4.37 is that ∂_μ is defined as $\partial_\mu \equiv \frac{\partial}{\partial x^\mu} = (\frac{\partial}{\partial x^0}, -\frac{\partial}{\partial x^1}, -\frac{\partial}{\partial x^2}, -\frac{\partial}{\partial x^3})^T$. The minus signs here cancel with the minus signs that appear through the upper index of T_μ^ν and we get an overall plus sign.

[37] Using $\partial_0 = \partial_t$ and $\partial^i T_i^0 = \nabla \vec{T}$
and the famous divergence theorem
$\int_V d^3 x \nabla A = \int_{\delta V} d^2 x A$, which enables
us to rewrite the integral over some
volume V, as an integral over the cor-
responding surface δV. A very illumi-
nating proof of the divergence theorem,
there called Gauss' theorem, can be
found at http://www.feynmanlectures.
caltech.edu/II_03.html which is
Chapter 3 of the freely online avail-
able Richard P. Feynman, Robert B.
Leighton, and Matthew Sands. *The
Feynman Lectures on Physics: Volume 2.*
Addison-Wesley, 1st edition, 2 1977.
ISBN 9780201021172

[38] Take note that strictly speaking this is
only true for spin 0 and spin 1/2 fields.
For spin 1 fields, only the field-strength
tensor $F_{\mu\nu}$ must vanish to guarantee
finite energy, while the "basic fields" A_μ
(with $F_{\mu\nu} := \partial_\mu A_\nu - \partial_\nu A_\mu$) can be non-
zero "at infinity". (The field strength
tensor will be defined in Eq. 6.23).
This observation leads to the non-
trivial vacuum picture in non-abelian
gauge theories. Unfortunately, a proper
discussion lies far beyond the scope of
this book.

[39] Because $\partial_0 T_\mu^0 a^\mu = \partial_0 T_0^0 a^0 - \partial_0 T_i^0 a^i = 0$, with arbitrary a^μ we get a separate
continuity equation for each compo-
nent.

This tells us directly that we have conserved quantities, because, for example, for $\mu = 0$ we get[37]

$$\partial_0 T_0^0 + \partial_i T_0^i = 0 \rightarrow \partial_0 T_0^0 = -\partial_i T_0^i$$

$$\partial_t T_0^0 = -\nabla \vec{T} \underbrace{\rightarrow}_{\text{Integrating over some infinite volume } V} \int_V d^3 x \partial_t T_0^0 = -\int_V d^3 x \nabla \vec{T}$$

$$\rightarrow \partial_t \int_V d^3 x T_0^0 = -\int_V d^3 x \nabla \vec{T} \overbrace{=}^{\text{Divergence theorem; } \delta V \text{ denotes the surface of the volume } V} -\int_{\delta V} d^2 x \vec{T} \underbrace{=}_{\text{Because fields vanish at infinity}} 0 \qquad (4.38)$$

$$\rightarrow \partial_t \int_V d^3 x T_0^0 = 0. \qquad (4.39)$$

In the last step we use that if we have an infinite volume V, like a sphere with infinite radius r, and we have to integrate over the surface of this volume δV, we need to evaluate our fields at $r = \infty$. Fields must vanish at infinity, because otherwise the total field energy would be infinite and field configurations with infinite energy are non-physical[38].

We conclude: the invariance under translations in spacetime leads us to 4 conserved quantities[39], one for each component μ. Equation 4.39 tells us these are

$$E = \int d^3 x T_0^0 \qquad (4.40)$$

$$P_i = \int d^3 x T_i^0, \qquad (4.41)$$

where as always $i = 1, 2, 3$. These quantities are called the **total energy** E of the system, which is conserved because we have **invariance under time-translations** $x_0 \rightarrow x_0 + a_0$ and the **total momentum** of the field configuration P^i, which is conserved because we have **invariance under spatial-translations** $x_i \rightarrow x_i + a_i$.

4.5.3   Rotations and Boosts

Next, we take a look at invariance under rotations and boosts. We will start with the second part of Eq. 4.30 and then look afterwards at the implications of the first part. We will get a quantity from the first part and a quantity from the second part, which are **together** conserved. A scalar field has no components that could mix, hence for a scalar field the first part is zero. In other words, for scalars we use the 1-dimensional representation of the generators of the Lorentz group $S_{\mu\nu} = 0$, which was derived in Section 3.7.4. Therefore what we derive in this section is the complete conserved quantity for a scalar field.

THE FRAMEWORK 109

The second part of the transformation is given by

$$x_\mu \to x'_\mu = x_\mu + \delta x_\mu = x_\mu + M_\mu^\sigma x_\sigma \qquad (4.42)$$

where M_μ^σ are the generators of rotations and boost in their infinite-dimensional representation. This means they are represented by differential operators, as defined in Eq. 3.248.

From Eq. 4.42, we see that the change of the coordinates is given by $\delta x_\mu = M_\mu^\sigma x_\sigma$. Putting this into Eq. 4.33 yields

$$\partial_\nu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\nu \Phi)} \frac{\partial(\Phi)}{\partial x_\mu} - \delta_\mu^\nu \mathscr{L} \right) M^{\mu\sigma} x_\sigma = 0 \qquad (4.43)$$

$$\underbrace{\to}_{\text{Eq. 4.35}} \partial_\nu T_\mu^\nu M^{\mu\sigma} x_\sigma = 0 \,.$$

We can write this a little differently to bring it into the conventional form

$$= \frac{1}{2}(\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma + \partial^\nu T_\nu^\mu \underbrace{M_{\mu\sigma}}_{=-M_{\sigma\mu}} x^\sigma) = \frac{1}{2}(\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma - \partial^\nu T_\nu^\mu M_{\sigma\mu} x^\sigma)$$

$$\underbrace{=}_{\text{Renaming dummy indices}} \frac{1}{2}(\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma - \partial^\nu T_\nu^\sigma M_{\mu\sigma} x^\mu) = \frac{1}{2}(\partial^\nu T_\nu^\mu x^\sigma - \partial^\nu T_\nu^\sigma x^\mu) M_{\mu\sigma}$$

$$= \frac{1}{2}\partial_\nu(T^{\mu\nu} x^\sigma - T^{\sigma\nu} x^\mu) M_{\mu\sigma} = 0$$

$$\to \text{ with } (J^\nu)^{\sigma\mu} \equiv T^{\mu\nu} x^\sigma - T^{\sigma\nu} x^\mu \quad \to \quad \partial_\nu(J^\nu)^{\sigma\mu} = 0. \qquad (4.44)$$

The quantity J^ν we introduced in the last step is called **Noether current**. Note that, because of the anti-symmetry of $M_{\mu\sigma} = -M_{\sigma\mu}$, which follows from the definition in Eq. 3.173, we have $M_{\mu\mu} = 0$. Therefore, we found 6 different **continuity equations** $\partial_\nu(J^\nu) = 0$, one for each non-vanishing $M_{\mu\sigma}$. With the same arguments from Eq. 4.39 we conclude that the quantities conserved in time are

$$Q^{\mu\sigma} = \int d^3x (T^{\mu 0} x^\sigma - T^{\sigma 0} x^\mu) \,. \qquad (4.45)$$

For rotational invariance[40], we therefore have the conserved quantities

$$L_{\text{orbit}}^i = \frac{1}{2}\epsilon^{ijk} Q^{jk} = \frac{1}{2}\epsilon^{ijk} \int d^3x (T^{j0} x^k - T^{k0} x^j), \qquad (4.46)$$

which we call the **orbital angular momentum** of our field[41].

Equivalently, we have the conserved quantity for boosts[42]

$$Q^{0i} = \int d^3x (T^{00} x^i - T^{i0} x^0), \qquad (4.47)$$

which is discussed a bit more in the appendix Section 4.7.

[40] Recall the connection between the generator of rotations J_i and $M_{\mu\nu}$ from Eq. 3.173: $J_i = \frac{1}{2}\epsilon_{ijk} M_{jk}$, which means we now restrict to the spatial components $i = j = \{1, 2, 3\}$ here and write i, j instead of $\mu = \nu = \{0, 1, 2, 3\}$.

[41] The prefix "orbital" will make sense in a moment, because actually this is just one part of the complete angular momentum. We will have a look at the missing part in a moment.

[42] Recall the connection between the boost generators K_i and $M_{\mu\nu}$: $K_i = M_{0i}$.

4.5.4   Spin

Next, we want to take a look at the implications of the first part of Eq. 4.30, which is what we missed in our previous derivation of the conserved quantity that follows from rotations. We now consider

$$\delta \Phi = \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x), \tag{4.48}$$

where $S^{\mu\nu}$ is the appropriate **finite-dimensional** representation of the transformation in question[43]. This gives us in Eq. 4.29 the extra term

$$\partial_\rho \left(\frac{\partial \mathscr{L}}{\partial(\partial_\rho \Phi)} \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x) \right), \tag{4.49}$$

which leads to an additional term in Eq. 4.46. The complete conserved quantity then reads

$$L^i = \frac{1}{2} \epsilon^{ijk} Q^{jk} = \frac{1}{2} \epsilon^{ijk} \int d^3x \left(\frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} S^{jk} \Phi(x) + (T^{j0} x^k - T^{k0} x^j) \right) \tag{4.50}$$

and therefore we write

$$L^i = L^i_{\text{spin}} + L^i_{\text{orbit}}. \tag{4.51}$$

The first part is something new, but needs to be similar to the usual orbital angular momentum we previously considered, because the two terms are added and appear when considering the same invariance. The standard point of view is that the first part of this conserved quantity is some-kind of internal angular momentum[44].

Recall that we label the representations of the Poincaré group by something we call spin, too. Now this notion appears again. We will learn in the next chapter and in great detail in Section 8.5.5, how we are able to measure this new kind of angular momentum, called spin and see later that it indeed coincides with the label we use for the representations of the Poincaré group.

4.5.5   Noether's Theorem for Field Theories - Internal Symmetries

We now take a look at internal symmetries[45]. The invariance of a Lagrangian ($\delta\mathscr{L} = 0$) under some infinitesimal transformation of the field **itself**

$$\Phi_i \rightarrow \Phi'_i = \Phi_i + \delta\Phi_i \tag{4.52}$$

can be written mathematically

$$\delta\mathscr{L} = \mathscr{L}(\Phi^i, \partial_\mu \Phi^i) - \mathscr{L}(\Phi^i + \delta\Phi^i, \partial_\mu(\Phi^i + \delta\Phi^i))$$

[43] The finite-dimensional representations are responsible for the mixing of the field components. For example, the two dimensional representation of the rotation generators: $J_i = \frac{1}{2}\sigma_i$, mix the components of Weyl spinors.

[44] We will see later that fields create and destroy particles. A spin $\frac{1}{2}$ field creates spin $\frac{1}{2}$ particles, which is an unchangeable property of an elementary particle. Hence the usage of the word "internal". Orbital angular momentum is a quantity that describes how two or more particles revolve around each other.

[45] Internal symmetries will be incredibly important for interacting field theories. In addition, one conserved quantity, following from one of the easiest internal symmetries will provide us with the starting point for quantum field theory.

$$= -\frac{\partial \mathscr{L}(\Phi^i, \partial_\mu \Phi^i)}{\partial \Phi^i} \delta \Phi^i - \frac{\partial \mathscr{L}(\Phi^i, \partial_\mu \Phi^i)}{\partial (\partial_\mu \Phi^i)} \partial_\mu \delta \Phi^i \overset{!}{=} 0, \qquad (4.53)$$

where we used the Taylor expansion to get to the second line and only keep terms of first order in $\delta \Phi_i$, because the transformation is infinitesimal. From the Euler-Lagrange equation (Eq. 4.10) for fields we know

$$\frac{\partial \mathscr{L}}{\partial \Phi^i} = \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial (\partial_\mu \Phi^i)} \right).$$

Putting this into Eq. 4.53, we get

$$\delta \mathscr{L} = -\partial_\mu \left(\frac{\partial \mathscr{L}}{\partial (\partial_\mu \Phi^i)} \right) \delta \Phi^i - \frac{\partial \mathscr{L}}{\partial (\partial_\mu \Phi^i)} \partial_\mu \delta \Phi^i = 0$$

which is, using the product rule, the same as

$$\partial_\mu \left(\frac{\partial \mathscr{L}}{\partial (\partial_\mu \Phi^i)} \delta \Phi^i \right) = 0. \qquad (4.54)$$

This means, if the Lagrangian is invariant under the transformation $\Phi_i \rightarrow \Phi'^i = \Phi_i + \delta \Phi^i$, we get again a quantity, called Noether current

$$J^\mu \equiv \frac{\partial \mathscr{L}}{\partial (\partial_\mu \Phi^i)} \delta \Phi^i, \qquad (4.55)$$

which fulfils the continuity equation

$$\text{Eq. 4.54} \rightarrow \partial_\mu J^\mu = 0. \qquad (4.56)$$

Furthermore, following the same steps as in Eq. 4.39, we are able to derive a quantity that is conserved in time

$$\partial_t \int d^3 x J^0 = 0. \qquad (4.57)$$

Let's have a look at one very important example: invariance under displacements of the field itself[46], which means $\delta \Phi = -i\epsilon$ in Eq. 4.52:

$$\Phi \rightarrow \Phi' = \Phi - i\epsilon \qquad (4.58)$$

or for more than one field component

$$\Phi_i \rightarrow \Phi'_i = \Phi_i - i\epsilon_i. \qquad (4.59)$$

The generator of this transformation is $-i\frac{\partial}{\partial \Phi}$, because

$$\Phi' = e^{-i\epsilon \frac{\partial}{\partial \Phi}} \Phi = \left(1 - i\epsilon \frac{\partial}{\partial \Phi} + \dots \right) \Phi \approx \Phi - i\epsilon \qquad (4.60)$$

The corresponding conserved quantity is called **conjugate** momentum Π. With Eq. 4.54 and Eq. 4.57 we get[47]

[46] Take note that at this point we made no assumptions about the field being complex or real. The constant ϵ is arbitrary and could be entirely imaginary if we want to move our field by a real value. The i is included at this point to make the generator Hermitian (which is by no means obvious at this point) and therefore the eigenvalues real, as will be shown later. Don't get confused at this point. You could simply see this extra i as a convention, because we could just as well absorb it into the constant by defining $\epsilon' := i\epsilon$.

[47] The $\delta \Phi$ in Eq. 4.55 becomes a constant ϵ and because the quantity is conserved for arbitrary ϵ we define our conserved quantity without it.

$$\partial_t \Pi = 0 \rightarrow \partial_t \Pi = \partial_t \int d^3x J^0 = \partial_t \int d^3x \pi = 0$$

$$\text{with } J^0 = \pi = \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \tag{4.61}$$

[48] This rather abstract quantity is one of the most important notions in quantum field theory, as we will see in Chapter 6.

The quantity π here is called the **conjugate momentum density**[48]. Take note that this is different from the *physical* momentum density of the field, which arises from the invariance under *spatial* translations $\Phi(x) \rightarrow \Phi(x') = \Phi(x + \epsilon)$ and is given by

$$p^i = T^{0i} = \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_i},$$

which we derived in Eq. 4.41. In later chapters we will look at many more internal symmetries, which will prove to be invaluable when we develop a theory that describes interactions. Now let's summarize what we learned about symmetries and conserved quantities in field theories:

- translational invariance in space $\Rightarrow$ conservation of physical momentum $P^i = \int d^3x T^{0i} = \int d^3x \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_i}$,

[49] Another name for a boost is a translation in momentum space.

- boost invariance[49] $\Rightarrow$ constant velocity of the center of energy,

- rotational invariance $\Rightarrow$ conservation of the total angular momentum $L^i = \frac{1}{2}\epsilon^{ijk} \int d^3x \left(\frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} S^{jk}\Phi(x) + (T^{k0}x^j - T^{j0}x^k) \right) = L^i_{\text{spin}} + L^i_{\text{orbit}}$ consisting of spin and orbital angular momentum,

- translational invariance in time $\Rightarrow$ conservation of energy $E = \int d^3x T^{00} = \int d^3x \left(\frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_0} - \mathscr{L} \right)$,

- displacement invariance of the field itself $\Rightarrow$ conservation of the conjugate momentum $\Pi = \int d^3x \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} = \int d^3x \pi$.

Further Reading Tips

[50] Cornelius Lanczos. *The Variational Principles of Mechanics*. Dover Publications, 4th edition, 3 1986. ISBN 9780486650678

[51] David Morin. *Introduction to Classical Mechanics: With Problems and Solutions*. Cambridge University Press, 1 edition, 2 2008. ISBN 9780521876223

[52] Herbert Goldstein, Charles P. Poole Jr., and John L. Safko. *Classical Mechanics*. Addison-Wesley, 3rd edition, 6 2001. ISBN 9780201657029

- **Cornelius Lanczos - The Variational Principles of Mechanics**[50] is a brilliant book about the usage of the Lagrangian formalism in classical mechanics.

- **David Morin - Introduction to Classical Mechanics**[51] is an extremely student friendly book that carefully explains how the Lagrangian framework is used in classical mechanics.

- **Herbert Goldstein - Classical Mechanics**[52] is *the* standard book about classical mechanics with many great explanations regarding the Lagrangian formalism.

4.6   Appendix: Conserved Quantity from Boost Invariance for Particle Theories

In this appendix we want to find the conserved quantity that follows for our example Lagrangian $\mathcal{L} = \frac{1}{2}m\dot{\vec{q}}^2$ from boost invariance. A boost transformation is $q \to q' = q + \delta q = q + vt$, where v denotes a constant velocity. Our Lagrangian is not invariant under boosts, but this is no problem at all, as we have seen above:

For our equations of motion to be unchanged under transformations the Lagrangian is allowed to change up to the total derivative of an arbitrary function. This condition is clearly fulfilled for our boost transformation $q \to q' = q + vt \underbrace{\Rightarrow}_{v=\text{const!}} \dot{q}' = \dot{q} + v$ and therefore our

Lagrangian

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2$$

is changed by

$$\delta\mathcal{L} = \mathcal{L}(q,\dot{q},t) - \mathcal{L}(q',\dot{q}',t) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m\dot{q}'^2$$

$$= \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m(\dot{q}+v)^2 = -m\dot{q}v - \frac{1}{2}mv^2,$$

which is the total derivative of the function[53]

$$-G \equiv -mqv - \frac{1}{2}mv^2 t.$$

We conclude that our equations of motion stay unchanged and we can find a conserved quantity[54].

For boosts, the conserved quantity is (Eq. 4.16)

$$J_{\text{boost}} = \underbrace{\frac{\partial \mathcal{L}}{\partial \dot{q}}}_{\equiv p} \underbrace{\delta q}_{=vt} -G = pvt - mvq - \frac{1}{2}mv^2 t,$$

where we can cancel one v from every term, because in Eq. 4.16 we have a zero on the right hand side. This gives us

$$\tilde{J}_{\text{boost}} = \underbrace{pt - \frac{1}{2}mvt}_{\equiv \tilde{p}t} -mq = \tilde{p}t - mq. \qquad (4.62)$$

This is a rather strange conserved quantity because its value depends on the starting time. We can choose a suitable zero time, which makes this quantity zero. Because it is conserved, this zero stays zero for all times.

[53] $\frac{dG}{dt} = \frac{d}{dt}(-mqv - \frac{1}{2}mv^2 t) = -m\dot{q}v - \frac{1}{2}mv^2$, because v and m are constant.

[54] Another way to see this is by looking directly at the equation of motion: $m\ddot{q} = 0$. We have the transformation $q \to q' = q + vt$, with $v =$const and therefore $\dot{q} \to \dot{q}' = \dot{q} + v$ and $\ddot{q} \to \ddot{q}' = \ddot{q}$.

4.7   Appendix: Conserved Quantity from Boost Invariance for Field Theories

For invariance under boosts in a field theory we have the conserved quantities

$$Q^{0i} = \int d^3x (T^{00}x^i - T^{i0}x^0), \qquad (4.63)$$

which we derived in Eq. 4.47, using $K_i = M_{0i}$.

How can we interpret this quantity? Q^{0i} is conserved, which means

$$0 = \frac{\partial Q^{0i}}{\partial t} = \underbrace{\int d^3x \frac{\partial T^{i0}}{\partial t} x^0}_{=x^0 \frac{\partial}{\partial t} \int d^3x T^{i0}} + \int d^3x T^{i0} \underbrace{\frac{\partial x_0}{\partial t}}_{=1 \text{ because } x_0 = t} - \frac{\partial}{\partial t} \int d^3x \frac{\partial T^{00}}{\partial t} x^i$$

$$= t\frac{\partial P^i}{\partial t} + P^i - \frac{\partial}{\partial t} \int d^3x T^{00}x^i.$$

From Eq. 4.20 we know that P^i is conserved $\frac{\partial P^i}{\partial t} = 0 \rightarrow P^i = \text{const.}$ and we can conclude

$$\frac{\partial}{\partial t} \int d^3x T^{00}x^i = P^i = \text{const.} \qquad (4.64)$$

[55] Don't worry about the meaning of this term too much, because this isn't really a useful conserved quantity like the others we derived so far.

Therefore this conservation law tells us[55] that the center of energy travels with constant velocity. Or from a different perspective: The momentum equals the total energy times the velocity of the centre-of-mass energy.

Nevertheless, the conserved quantity derived in Eq. 4.47 remains strange. By a suitable choice of the starting time the quantity can be chosen to be zero. Because it is conserved, this zero stays zero in all instances.

Part III

The Equations of Nature

"If one is working from the point of view of getting beauty into one's equation, ...one is on a sure line of progress."

5

Measuring Nature

Now that we have discovered the connection between symmetries and conserved quantities, we can utilize this connection. In more technical terms, Noether's Theorem establishes a connection between the generator of a symmetry transformation and a conserved quantity. In this chapter we will utilize this connection.

Conserved quantities are what physicists commonly use to describe physical systems, because no matter how complicated the system changes, the conserved quantities stay the same. For example, to describe what is happening in an experiment physicists use the momentum, energy and angular momentum. Noether's theorem hints us towards an incredibly important idea:

We identify the quantities that we use to describe nature with the corresponding generators:

$$\text{physical quantity} \Rightarrow \text{generator of the corresponding symmetry.} \quad (5.1)$$

As we will see, this identification naturally guide us towards quantum theory. Let's make this concrete by considering a particle theory.

5.1 The Operators of Quantum Mechanics

The invariance of the Lagrangian under the action of the generator of spatial-translations leads us to the conservation of momentum. Therefore, we make the identification

$$\text{momentum } \hat{p}_i \to \text{ generator of spatial-translations } -i\partial_i$$

It is conventional to use a hat: $\hat{O}$ to denote an operator.

Analogously, the invariance under the action of the generator of time-translations leads us to the conservation of energy. Conse-

quently

$$\text{energy } \hat{E} \rightarrow \text{ generator of time-translations } i\partial_0.$$

There is no symmetry corresponding to the "conservation of location" and therefore the location is not identified with a generator. We simply have[1]

$$\text{location } \hat{x}_i \rightarrow x_i.$$

The physical quantities we want to use in our theory to describe nature are now given by (differential) operators. The logical next thing to ask is: What do they act on and where is the connection to things we can measure in experiments? We will discuss this in a later chapter in detail. At this point it is sufficient to note that our physical quantities, now operators, need to act on *something*. Here we want to move on with an abstract thing that the operators act on. Let's name it Ψ. We will explore later what this *something* is.

At this point, we are able to derive one of the most important[2] equations of quantum mechanics. As explained above, we assume that our operators act on something abstract Ψ. Then we have

$$[\hat{p}_i, \hat{x}_j]\Psi = (\hat{p}_i\hat{x}_j - \hat{x}_j\hat{p}_i)\Psi = (-i\partial_i\hat{x}_j + \hat{x}_j i\partial_i)\Psi$$
$$\underbrace{= -(i\partial_i\hat{x}_j)\Psi - \hat{x}_j(i\partial_i\Psi)}_{\text{product rule}} + \hat{x}_j i\partial_i\Psi \underbrace{= -i\delta_{ij}\Psi.}_{\text{because } \partial_i\hat{x}_j=\frac{\partial x_j}{\partial x_i}} \qquad (5.2)$$

This equation holds for arbitrary Ψ, because we made no assumptions about it and therefore we can write the equation without Ψ[3]:

$$[\hat{p}_i, \hat{x}_j] = -i\delta_{ij}. \qquad (5.3)$$

5.1.1 Spin and Angular Momentum

In the last chapter (Section 4.5.4) we saw that the conserved quantity that resulted from rotational invariance has two parts. The second part was identified with the orbital angular momentum and we therefore make the identification with the **infinite** dimensional representation of the generator

$$\text{orb. angular mom. } \hat{L}_i \rightarrow \text{ gen. of rot. (inf. dim. rep.) } i\frac{1}{2}\epsilon_{ijk}(x^j\partial^k - x^k\partial^j)$$

Analogously we identify the first part, called spin, with the corresponding **finite** dimensional representation of the generators[4]

$$\text{spin } \hat{S}_i \rightarrow \text{ generators of rotations (fin. dim. rep.) } S_i.$$

[1] Alternatively, we can see this by looking at the conserved quantity corresponding to invariance under boosts. Take note that we derived in Section 4.6 the conserved quantity corresponding to a non-relativistic Galilei boost, because we started with a non-relativistic Lagrangian. Nevertheless, we can do the same for the relativistic Lagrangian and end up with the conserved quantity $t\vec{p} - \vec{x}E$. The relativistic energy is given by $E = \sqrt{m^2 + p^2}$. In the non-relativistic limit $c \rightarrow \infty$ we get $E \approx m$ and therefore recover the formula we derived for a Galilei boost from the Lorentz boost conserved quantity. The conserved quantity for a particle theory is then $M_i = (tp_i - x_iE)$. The generator of boosts is (see Eq. 3.248 with $K_i = M_{0i}$) $K_i = i(x_0\partial_i - x_i\partial_0)$. Comparing the two equations, with of course $x_0 = t$, yields $M_i = (tp_i - x_iE) \leftrightarrow K_i = i(t\partial_i - x_i\partial_0)$ The identification is now, with the identifications we made earlier, straightforward. Location $\hat{x}_i \rightarrow x_i$.

[2] If you don't know anything about quantum mechanics, it may seem strange to you why this little equation is so important, but maybe you have heard of the Heisenberg uncertainty principle. In Section 8.3 we will take a closer look at the formalism of quantum mechanics. Then we will be able to see that this equation tells us that we aren't able to measure the momentum and the location of a particle with arbitrary precision. Our physical quantities are interpreted as measurement operators and this equation tells us that a measurement of location followed by a measurement of momentum is not the same as a measurement of momentum followed by a measurement of location.

[3] If this is a new idea to you, take note that we could rewrite **every** vector equation as an equation that involves only numbers. For example Newton's second law: $\vec{F} = m\ddot{\vec{x}}$, could be written as $\vec{F}\vec{C} = m\ddot{\vec{x}}\vec{C}$, which is certainly true for any vector $\vec{C}$. Nevertheless, if the equation is true for any $\vec{C}$, writing it all the time makes little sense.

[4] Recall that this was the part of the conserved quantity that resulted from the invariance under mixing of the field components. Hence the finite-dimensional representation.

As explained in the text below Eq. 4.30 the relation between $S_{\mu\nu}$ and the generator of rotations S_i is $S_i = \frac{1}{2}\epsilon_{ijk}S_{jk}$.

For example, when we consider a spin $\frac{1}{2}$ field, we have to use the two-dimensional representation which we derived in Section 3.7.5:

$$\hat{S}_i = i\frac{\sigma_i}{2}\,, \tag{5.4}$$

where once more σ_i denotes the Pauli matrices. We will return to this very interesting and very strange type of angular momentum in Section 8.5.5, after we learned how to work with the operators that we derive in this chapter. It is important to keep in mind that only the **sum** of spin and orbital angular momentum is conserved.

5.2   The Operators of Quantum Field Theory

The central objects in a field theory are, of course, fields. A field is a **function of the location in space and time**[5] $\Phi = \Phi(x)$. Later we want to describe interactions at points in spacetime and therefore work with the densities of our dynamical variables $\pi = \pi(x)$ and not the total quantities that we get by integrating the densities over all space $\Pi = \int dx^3 \pi(x) \neq \Pi(x)$.

   We discovered in the last chapter for invariance under displacements of the field itself $\Phi \to \Phi - i\epsilon$ a new conserved quantity, called conjugate momentum Π. Analogous to the identifications we made in the last section, we identify the conjugate momentum density with the corresponding generator (Eq. 4.60)

conj. mom. density $\pi(x) \to$ gen. of displ. of the field itself $: -i\dfrac{\partial}{\partial \Phi(x)}$.

We could now go on, like we did for a particle theory, and identify momentum, angular momentum and energy with the corresponding generators. However, we will see later that quantum field theory works a little differently and the identification we make here will prove to be sufficient.

   For the same reasons discussed in the last section, we need something our operators act on. Thus again, we work again with an abstract Ψ, that we will specify in a later chapter. We are then again able to derive an incredibly important equation, this time of quantum field theory[6]

$$[\Phi(x), \pi(y)]\Psi = \left[\Phi(x), -i\frac{\partial}{\partial \Phi(y)}\right]\Psi$$

[5] Here $x = x_0, x_1, x_2, x_3$ includes time $x_0 = t$.

[6] For the last step we use the analogue to $\frac{\partial x_i}{\partial x_j} = \delta_{ij}$ for the delta distribution $\frac{\partial f(x_i)}{\partial f(x_j)} = \delta(x_i - x_j)$, which can be shown in a rigorous way. For some more information have a look at Appendix D.2.

$$\underbrace{=}_{\text{product rule}} -i\Phi(x)\cancel{\frac{\partial \Psi}{\partial \Phi(y)}} + i\Phi(x)\left(\cancel{\frac{\partial \Psi}{\partial \Phi(y)}}\right) + i\left(\frac{\partial \Phi(x)}{\partial \Phi(y)}\right)\Psi = i\delta(x-y)\Psi \,.$$

$$(5.5)$$

Again, the equations hold for arbitrary Ψ and we can therefore write

$$[\Phi(x), \pi(y)] = i\delta(x-y) \,. \qquad (5.6)$$

Analogously we have for more than one field component

$$[\Phi_i(x), \pi_j(y)] = i\delta(x-y)\delta_{ij}. \qquad (5.7)$$

As we will see later, almost everything in quantum field theory follows from this little equation.

6

Free Theory

In this chapter we will derive the basic equations for a physical theory of **free** (=non-interacting) fields[1] from symmetry. We will

- derive the Klein-Gordon equation using the $(0,0)$ representation of the Lorentz group,

- derive the Dirac equation using the $(\frac{1}{2},0) \oplus (0,\frac{1}{2})$ representation of the Lorentz group,

- derive the Proca equations from the vector $(\frac{1}{2},\frac{1}{2})$ representation of the Lorentz group which in the massless limit become the famous Maxwell equations.

[1] Although we specify here for fields we will see in a later chapter how the equations we derive here can be used to describe particles, too.

6.1 Lorentz Covariance and Invariance

In the following sections, we will derive the fundamental equations of motion of the standard model of particle physics, which is the best physical theory that we have. We want that these equations are the same in all inertial frames, because otherwise would have a different equation for each possible frame of reference. This would be useless because there is no preferred frame of reference in special relativity. The technical term for an equation that looks the same in all inertial frames of reference is **Lorentz covariant equation**. An object is Lorentz covariant if it transforms under a given representation of the Lorentz group. For example, a vector A_μ, transforms according to the $(\frac{1}{2},\frac{1}{2})$ representation and is therefore Lorentz covariant. This means $A_\mu \to A'_\mu$, but not something completely different. On the other hand, for example, a term of the form $A_1 + A_3$ is not Lorentz covariant, because it does not transform according to a representation of the Lorentz group. This does not mean that we do not know how it transforms. The transformation properties can be easily derived from the transformation laws for A_μ, but nevertheless this term looks

completely different in different inertial frames. In a boosted frame, it may look like $A_2 + A_4$. An equation that involves only Lorentz covariant objects is called a Lorentz covariant equation. For example,

$$A_\mu + 7B_\mu + C_\nu A^\nu D_\mu = 0$$

is a Lorentz covariant equation, because in another coordinate system it reads

$$\Lambda^\nu_\mu A_\nu + 7\Lambda^\nu_\mu B_\nu + \Lambda^\rho_\nu C_\rho \Lambda^\nu_\eta A^\eta \Lambda^\sigma_\mu D_\sigma = A'_\mu + 7B'_\mu + C'_\nu A'^\nu D'_\mu = 0.$$

We see that it looks the same. An equation containing only some components of such objects is, in general, not Lorentz covariant and therefore looks completely different in each inertial frame.

To make sure we only end up with Lorentz **covariant** equations, we require the action S to be Lorentz **invariant**. This means it should only contain terms that stay exactly the same when changing the frame of reference. In other words: The action is only allowed to contain terms that do not change under Lorentz transformations. We get the equations of motion from the action[2] S. Now if S depends on the frame of reference, so would the terms in the equations that follow from it and therefore these equations can not be Lorentz covariant.

As already discussed in the last chapter, we can use the more restrictive requirement that the Lagrangian should be invariant, because if the Lagrangian is invariant the action is, too.

6.2 Klein-Gordon Equation

We now start with the simplest possible case: scalars, which transform according to the $(0,0)$ representation of the Lorentz group. To specify the equation of motion for scalars we need to find the corresponding Lagrangian. A general Lagrangian that is compatible with our restrictions[3] is

$$\mathscr{L} = A\Phi^0 + B\Phi + C\Phi^2 + D\partial_\mu\Phi + E\partial_\mu\Phi\partial^\mu\Phi + F\Phi\partial_\mu\Phi. \qquad (6.1)$$

Firstly, take note that we are considering the Lagrangian density $\mathscr{L}$ and not L itself and we get our physical theory from the action

$$S = \int dx \mathscr{L}, \qquad (6.2)$$

where dx is to be understood as the integral over space and time. Therefore a term like $\Phi\partial_\mu\partial^\mu\Phi$ would be redundant, because it is equivalent to the term $\partial_\mu\Phi\partial^\mu\Phi$, as we can see if we integrate by parts[4].

[2] Recall that we minimize the action and the result of this minimization procedure is the Euler-Lagrange equation, which yields the equation of motion for our system.

[3] Discussed in Section 4.2: We only consider terms of order $0, 1$ and 2 in Φ. The term with the lowest possible derivative will become clear in a moment.

[4] The boundary terms, as usual, vanish, because fields vanish at infinity. Recall that this follows, because field configurations that do not vanish at infinity correspond to infinite field energy, which is non-physical (Section 2.3).

In addition, Lorentz **invariance** restricts the Lagrangian to be a scalar. Therefore, all odd powers in ∂_μ, like in $\partial_\mu \Phi$ are forbidden. What about the constants i.e. a and c etc. having a Lorentz index? This would mean that a, c are four-vectors, specifying a direction in spacetime and therefore violating the assumption of isotropy of space. We can neglect the constant term, i.e. $A = 0$, because we get our physical theory from the Euler-Lagrange equation and a constant in the Lagrangian has no influence on the equation of motion[5]. In addition, we ignore the term linear in Φ, i.e. $B = 0$, because it leads, using the Euler-Lagrange equation, to a constant in our equations of motion[6]. At least in a free theory, the absolute value of the field has no relevance and thus we can always perform a redefinition $\Phi \to \Phi' = \Phi + \text{const.}$ such that the constant term vanishes from the equation of motion[7]. What remains is

$$\mathscr{L} = C\Phi^2 + E\partial_\mu\Phi\partial^\mu\Phi. \qquad (6.3)$$

As the heading of this chapter indicates we want to develop a free theory, which means there is just one Φ and no terms of the form $\Phi_1\Phi_2$. Terms like this will be investigated in the next chapter, when we develop a theory describing interactions.

There is one last thing to note: we are left with only two constants C and E. By using variational calculus we are able to combine these into just one constant, because an overall constant in the Lagrangian has no influence on the physics[8]. Nevertheless, it is conventional to include an overall factor $\frac{1}{2}$ into the Lagrangian and call the remaining constant[9] $-m^2$. Therefore, we are finally left with

$$\mathscr{L} = \frac{1}{2}(\partial_\mu\Phi\partial^\mu\Phi - m^2\Phi^2). \qquad (6.4)$$

If we now use the variational calculus machinery, which means putting this Lagrangian into the Euler-Lagrange equation (Eq. 4.10), we get the equation of motion

$$0 = \frac{\partial\mathscr{L}}{\partial\Phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi)}\right)$$

$$\to 0 = \frac{\partial}{\partial\Phi}\left(\frac{1}{2}(\partial_\mu\Phi\partial^\mu\Phi - m^2\Phi^2)\right) - \partial_\mu\left(\frac{\partial}{\partial(\partial_\mu\Phi)}\left(\frac{1}{2}(\partial_\mu\Phi\partial^\mu\Phi - m^2\Phi^2)\right)\right)$$

$$\to 0 = (\partial_\mu\partial^\mu + m^2)\Phi. \qquad (6.5)$$

This is the famous **Klein-Gordon equation**, which is the correct equation of motion to describe free spin 0 fields and particles.

---

[5] See Eq. 4.10: $\frac{\partial\mathscr{L}}{\partial\Phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi)}\right) = 0$ and therefore $\mathscr{L} \to \mathscr{L} + A$ with some constant A does not change anything: $\frac{\partial(\mathscr{L}+A)}{\partial\Phi} - \partial_\mu\left(\frac{\partial(\mathscr{L}+A)}{\partial(\partial_\mu\Phi)}\right) = \frac{\partial\mathscr{L}}{\partial\Phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi)}\right)$

[6] $\mathscr{L} \to \mathscr{L} + B\Phi$ yields $\frac{\partial(\mathscr{L}+B\Phi)}{\partial\Phi} - \partial_\mu\left(\frac{\partial(\mathscr{L}+B\Phi)}{\partial(\partial_\mu\Phi)}\right) = \frac{\partial\mathscr{L}}{\partial\Phi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi)}\right) + B = 0$, which is just an additional constant in the equation of motion.

[7] However take note that if the field interacts with other fields, this is no longer true, because such a field shift then also affects the terms that describe the coupling between the fields. This will be discussed in detail in Section. 7.3 in the context of spontaneous symmetry breaking.

[8] $\mathscr{L} \to C\mathscr{L}$ yields $\frac{\partial(C\mathscr{L})}{\partial\Phi} - \partial_\mu\left(\frac{\partial(C\mathscr{L})}{\partial(\partial_\mu\Phi)}\right) = 0$
$\updownarrow$
$\frac{\partial(\mathscr{L})}{\partial\Phi} - \partial_\mu\left(\frac{\partial(\mathscr{L})}{\partial(\partial_\mu\Phi)}\right) = 0$

[9] The suggestive name of this constant will become clear later, because we will see that it coincides with the mass of particles described by this Lagrangian.

6.2.1 Complex Klein-Gordon Field

For spin 0 fields, we are able to construct a Lorentz invariant Lagrangian without using the complex conjugate of the scalar field. This will not be the case for spin $\frac{1}{2}$ fields Ψ and this curious fact will have very interesting consequences[10]. Nevertheless, nothing prevents us from investigating the equally Lorentz invariant Lagrangian

$$\mathscr{L} = \partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi$$

as many textbooks do. This is simply the same as investigating **two** scalar fields of equal mass at the same time

$$\mathscr{L} = \frac{1}{2} \partial_\mu \phi_1 \partial^\mu \phi_1 - \frac{1}{2} m^2 \phi_1^2 + \frac{1}{2} \partial_\mu \phi_2 \partial^\mu \phi_2 - \frac{1}{2} m^2 \phi_2^2,$$

because we have

$$\phi \equiv \frac{1}{\sqrt{2}} \left(\phi_1 + i \phi_2 \right).$$

Again Lorentz symmetry dictates the form of the Lagrangian. Elementary scalar (=spin 0) particles are very rare. In fact, only one is experimentally verified: **the Higgs boson**. However this Lagrangian can also be used to describe composite systems like mesons. We will not investigate this Lagrangian any further and most textbooks use it only for "training purposes".

6.3 Dirac Equation

> Another story told of Dirac is that when he first met Richard Feynman, he said after a long silence "I have an equation. Do you have one too?"
>
> - **Anthony Zee**[11]

In this section we want to find the equation of motion for free spin $\frac{1}{2}$ fields/particles. We will use the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation of the Lorentz group, because a theory that respects symmetry under parity transformations must include the $(\frac{1}{2}, 0)$ and $(0, \frac{1}{2})$ representations at the same time[12]. The objects transforming under this representation are called Dirac spinors and combine right-chiral and left-chiral Weyl spinors into one object[13]:

$$\Psi \equiv \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_a \\ \xi^{\dot{a}} \end{pmatrix}. \tag{6.6}$$

Now, we need to search for Lorentz invariant objects constructed from Dirac spinors, which we can then put into the Lagrangian. The first step is to search for invariants constructed from our left-chiral and right-chiral Weyl spinors.

[10] To spoil the surprise: there is an antiparticle for each spin $\frac{1}{2}$ particle. Using complex fields is the same as considering two fields at the same time as explained in the text below. Therefore we are forced by Lorentz invariance to use **two** (closely connected) fields at the same time, which are commonly interpreted as particle and antiparticle fields.

[11] Anthony Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 1st edition, 3 2003. ISBN 9780691010199

[12] We showed in Section 3.7.9 that a parity transformation transforms the $(\frac{1}{2}, 0)$ representation, into the $(0, \frac{1}{2})$ representation.

[13] This was discussion in Section 3.7.9.

We will use the Van-der-Waerden notation, which was introduced in Section 3.7.7. Two possibly invariants are[14]

$$I_1 := \chi_{\dot{a}}^T \xi^{\dot{a}} = (\chi_{\dot{a}}^\star)^T \xi^{\dot{a}} = (\chi_a)^\dagger \xi^{\dot{a}} = (\chi_L)^\dagger \xi_R \tag{6.7}$$

and

$$I_2 := (\xi^a)^T \chi_a = ((\xi^{\dot{a}})^\star)^T \chi_a = (\xi_R)^\dagger \chi_L, \tag{6.8}$$

because we always need to combine a lower dotted with an upper dotted and a lower undotted with an upper undotted index, in order to get a Lorentz invariant term. This was shown explicitly in Section 3.7.7. Here we see again that right-chiral and left-chiral spinors are needed in pairs.

Furthermore, we can construct two Lorentz-invariant combinations involving first order derivatives, as we will see in a moment. But first we need to understand how we can write the derivative of a spinor. We learned in Section 3.7.8 how we can construct four-vectors from spinors

$$v_{a\dot{b}} = v_\nu \sigma^\nu_{a\dot{b}},$$

where v_ν transforms like a four-vector. The differentiation operator is therefore in the spinor formalism

$$\partial_{a\dot{b}} = \partial_\nu \sigma^\nu_{a\dot{b}}. \tag{6.9}$$

It is conventional to define $\bar{\sigma}^0 = I_{2\times 2}$, $\bar{\sigma}^i = -\sigma^i$ and using this we can construct the following Lorentz invariant terms

$$I_3 := (\chi_{\dot{a}})^T \partial_\mu (\sigma^\mu)^{\dot{a}b} \chi_b = (\chi_L)^\dagger \partial_\mu \bar{\sigma}^\mu \chi_L \tag{6.10}$$

and

$$I_4 := (\xi^a)^T \partial_\mu (\sigma^\mu)_{a\dot{b}} \xi^{\dot{b}} = (\xi_R)^\dagger \partial_\mu \sigma^\mu \xi_R. \tag{6.11}$$

In addition to $(\sigma^\mu)_{a\dot{b}}$, we need here also $(\sigma^\mu)^{\dot{a}b}$. The first index must be dotted and the second index undotted to combine properly with the other spinor indices. We get $(\sigma^\mu)^{\dot{a}b}$ by using the spinor metric twice:

$$(\sigma^\mu)^{\dot{a}b} = ((\sigma^\mu)^T)^{b\dot{a}} = \epsilon^{bc}((\sigma^\mu)^T)_{c\dot{d}}(\epsilon^{\dot{a}\dot{d}})^T$$

$$= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} (\sigma^\mu)^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^T$$

$$= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} (\sigma^\mu)^T \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \bar{\sigma}_\mu. \tag{6.12}$$

[14] There are two other possibilities, which go by the name Majorana mass terms. We already know how we can move spinor indices up and down by using the spinor "metric" ϵ. We can write down a Lorentz invariant term of the form $\epsilon(\chi_L)^\dagger \chi_L$, because $\epsilon(\chi_L)^\dagger = \chi^a$ has an upper-undotted index. $\epsilon(\chi_L)^\dagger$ transforms like a right-chiral spinor, but by writing a term like this we have less degrees of freedom. ξ_R and χ_L both have two components, and therefore we have four degrees of freedom in a term like $(\xi_R)^\dagger \chi_L$. In the term $\epsilon(\chi_L)^\dagger \chi_L = \chi^a \chi_a$, the object that transforms like a right-chiral spinor is not independent of χ_L and we therefore only have two degrees of freedom here. There is a lot more one can say about Majorana spinors and it is currently under (experimental) investigation which type of term is the correct one for neutrinos. One more thing is worth noting: a Majorana spinor is a "real" Dirac spinor. I put real into quotation marks, because usually real means $\Psi^\star \overset{!}{=} \Psi$. For spinors this condition is not Lorentz invariant (because the Lorentz transformations are complex in this representation). If we impose the standard condition ($\Psi^\star \overset{!}{=} \Psi$) in one frame, it will, in general, not hold in another frame. Instead, it is possible to derive a Lorentz invariant "reality" condition for Dirac spinors: $\begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \Psi^\star \overset{!}{=} \Psi$, which is commonly interpreted as charge conjugation. This interpretation will be explained in Section 7.1.5. Therefore, a Majorana spinor describes a particle which is equivalent to its charge conjugated particle, commonly called anti-particle. Majorana particles are their own anti-particles and a Majorana spinor is a Dirac spinor with an extra-condition: $\Psi_M \equiv \begin{pmatrix} \chi_L \\ \epsilon \chi_L^\star \end{pmatrix}$ or $\Psi_M \equiv \begin{pmatrix} -\epsilon \xi_R^\star \\ \xi_R \end{pmatrix}$

For example, for σ_3, we have

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{=\sigma_3^T} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}}_{=\bar{\sigma}_3=-\sigma_3}.$$

Don't let yourself get confused why ∂_μ acts only on one spinor. We are going to use these invariants in Lagrangians, which we always evaluate inside of integrals[15] and therefore, we can always integrate by parts to get the other possibility. Therefore, our choice here is no restriction[16].

When we introduce the matrices[17]

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \bar{\sigma}_\mu & 0 \end{pmatrix} \rightarrow \gamma_\mu = \begin{pmatrix} 0 & \bar{\sigma}_\mu \\ \sigma_\mu & 0 \end{pmatrix}, \tag{6.13}$$

we can write the invariants we just found using the Dirac spinor formalism. Using the matrices γ_μ and Dirac spinors our invariants can be written as

$$\Psi^\dagger \gamma_0 \Psi \quad \text{and} \quad \Psi^\dagger \gamma_0 \gamma^\mu \partial_\mu \Psi, \tag{6.14}$$

because

$$\Psi^\dagger \gamma_0 \Psi = \left((\chi_L)^\dagger \quad (\xi_R)^\dagger \right) \begin{pmatrix} 0 & \bar{\sigma}_0 \\ \sigma_0 & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \underbrace{(\chi_L)^\dagger \bar{\sigma}_0 \xi_R}_{=I_1} + \underbrace{(\xi_R)^\dagger \sigma_0 \chi_L}_{=I_2}.$$

These are exactly the first two invariants we found earlier and[18]

$$\Psi^\dagger \gamma_0 \gamma^\mu \partial_\mu \Psi = \left((\chi_L)^\dagger \quad (\xi_R)^\dagger \right) \begin{pmatrix} 0 & \bar{\sigma}_0 \\ \sigma_0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \sigma^\mu \partial_\mu \\ \bar{\sigma}^\mu \partial_\mu & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}$$

$$= \underbrace{(\chi_L)^\dagger \bar{\sigma}_0 \bar{\sigma}^\mu \partial_\mu \chi_L}_{=I_3} + \underbrace{(\xi_R)^\dagger \sigma_0 \sigma^\mu \partial_\mu \xi_R}_{=I_4}$$

gives the other two invariants[19], as promised. To avoid writing γ_0 all the time it is conventional to introduce the notation

$$\bar{\Psi} := (\Psi)^\dagger \gamma_0. \tag{6.15}$$

Now we have everything we need to construct a Lorentz-invariant Lagrangian using Dirac spinors that is in agreement with the restrictions[20] discussed in Section 4.2:

$$\mathscr{L} = A\Psi^\dagger \gamma_0 \Psi + B\Psi^\dagger \gamma_0 \gamma^\mu \partial_\mu \Psi = A\bar{\Psi}\Psi + B\bar{\Psi}\gamma^\mu \partial_\mu \Psi.$$

[15] Remember that we get the equations of motion from the action, which is the integral over the Lagrangian.

[16] We will see this more clearly when we derive the corresponding equations of motion. We get the same equations regardless of where we put ∂_μ and we could put both possibilities into the Lagrangian. This would be longer, but doesn't gives us anything new.

[17] We get the matrix with lowered index by using the metric: $\gamma_\mu = \eta_{\mu\nu}\gamma^\nu = \eta_{\mu\nu}\begin{pmatrix} 0 & \sigma^\nu \\ \bar{\sigma}^\nu & 0 \end{pmatrix} = \begin{pmatrix} 0 & \eta_{\mu\nu}\sigma^\nu \\ \eta_{\mu\nu}\bar{\sigma}^\nu & 0 \end{pmatrix} = \begin{pmatrix} 0 & \bar{\sigma}_\mu \\ \sigma_\mu & 0 \end{pmatrix}$, because $\eta_{\mu\nu}\sigma^\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}\begin{pmatrix} \sigma^0 \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \end{pmatrix} = \begin{pmatrix} \sigma^0 \\ -\sigma^1 \\ -\sigma^2 \\ -\sigma^3 \end{pmatrix} = \bar{\sigma}_\mu$.

[18] Remember that σ_0 is just the unit matrix and we have $\bar{\sigma}_0 = \sigma_0$.

[19] Take note that $\sigma^\mu \partial_\mu = \partial_\mu \sigma^\mu$, because σ^μ are constant matrices.

[20] Recall that these were: maximum order two in Ψ and only the lowest, non-trivial order in ∂_μ, which is here order 1.

Putting in the constants ($A = -m$, $B = i$) gives us the final **Dirac-Lagrangian**

$$\mathscr{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma^\mu\partial_\mu\Psi = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi. \qquad (6.16)$$

Take note that what appears here in our Lagrangian are two distinct fields, because Ψ is complex[21]. This is a requirement, because otherwise we can't get something Lorentz invariant. More explicitly, we have

$$\Psi = \Psi_1 + i\Psi_2$$

with two real fields Ψ_1 and Ψ_2. Instead of working with two real fields it is conventional to work with two complex fields Ψ and $\bar{\Psi}$, as two distinct fields.

Now, if we put this Lagrangian into the Euler-Lagrange equation, which we recite here for convenience

$$\frac{\partial\mathscr{L}}{\partial\Psi} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Psi)}\right) = 0$$

we get

$$-m\bar{\Psi} - i\partial_\mu\bar{\Psi}\gamma^\mu = 0 \rightarrow (i\partial^\mu\bar{\Psi}\gamma_\mu + m\bar{\Psi}) = 0. \qquad (6.17)$$

With the Euler-Lagrange equation for the field $\bar{\Psi}$

$$\frac{\partial\mathscr{L}}{\partial\bar{\Psi}} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\bar{\Psi})}\right) = 0$$

we get the equation of motion for Ψ

$$(i\gamma_\mu\partial^\mu - m)\Psi = 0. \qquad (6.18)$$

This is the famous **Dirac equation**, which is the equation of motion for spin $\frac{1}{2}$ particles and fields. Take note that this is exactly what we get if we integrate the Lagrangian by parts

$$-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi \underbrace{=}_{\text{Integrate by parts. Just imagine here the action integral.}} -m\bar{\Psi}\Psi - (i\partial^\mu\bar{\Psi})\gamma_\mu\Psi$$

and then use the Euler-Lagrange equation,

$$\rightarrow -m\Psi + i\partial_\mu\gamma^\mu\Psi = 0.$$

So it really makes no difference and the way we wrote the Lagrangian is no restriction, despite its asymmetry[22].

6.4   Proca Equation

Now, we want to find the equation of motion for an object transforming according to the $(\frac{1}{2}, \frac{1}{2})$ representation of the Lorentz group. We

[21] Therefore the left-chiral and right-chiral spinors inside each Dirac spinor, are complex, too.

[22] You are free to use the longer Lagrangian that includes both possibilities, but the results are the same.

already saw that this representation **is** the vector representation and therefore this task is easy. We simply take an arbitrary vector field A_μ and construct all possible Lorentz-invariants from it that are in agreement with the restrictions from Section 4.2. We must combine an upper with a lower index, because we defined the scalar product of Minkowski space in Section 2.4 this way and scalars are what we need in the Lagrangian. The possible invariants are[23]

$$I_1 = \partial^\mu A^\nu \partial_\mu A_\nu \quad , \quad I_2 = \partial^\mu A^\nu \partial_\nu A_\mu$$
$$I_3 = A^\mu A_\mu \quad , \quad I_4 = \partial^\mu A_\mu$$

and the Lagrangian reads

$$\mathscr{L}_{\text{Proka}} = C_1 \partial^\mu A^\nu \partial_\mu A_\nu + C_2 \partial^\mu A^\nu \partial_\nu A_\mu + C_3 A^\mu A_\mu + C_4 \partial^\mu A_\mu. \quad (6.19)$$

We can neglect the term $\partial^\mu A_\mu$, because has no influence on the equations of motion, as can be seen by looking at the Euler-Lagrange equation[24]. Therefore order 1 in ∂_μ is trivial.

If we now want to compute the equations of motion using the Euler-Lagrange equation for each field component independently

$$\frac{\partial \mathscr{L}}{\partial A_\rho} = \partial_\sigma \left(\frac{\partial \mathscr{L}}{\partial (\partial_\sigma A_\rho)} \right),$$

we need to be very careful about the indices. Let us take a look at the right-hand side of the Euler-Lagrange equation and pick the term involving C_1:

$$\partial_\sigma \left(\frac{\partial}{\partial (\partial_\sigma A_\rho)} (C_1 \partial^\mu A^\nu \partial_\mu A_\nu) \right)$$

$$\underbrace{=}_{\text{product rule}} C_1 \partial_\sigma \left((\partial_\mu A_\nu) \frac{\partial (\partial^\mu A^\nu)}{\partial (\partial_\sigma A_\rho)} + (\partial^\mu A^\nu) \frac{\partial (\partial_\mu A_\nu)}{\partial (\partial_\sigma A_\rho)} \right)$$

$$\underbrace{=}_{\text{lowering indices with the metric}} C_1 \partial_\sigma \left((\partial_\mu A_\nu) g^{\mu\kappa} g^{\nu\lambda} \frac{\partial (\partial_\kappa A_\lambda)}{\partial (\partial_\sigma A_\rho)} + (\partial^\mu A^\nu) \frac{\partial (\partial_\mu A_\nu)}{\partial (\partial_\sigma A_\rho)} \right)$$

$$= C_1 \partial_\sigma \left((\partial_\mu A_\nu) g^{\mu\kappa} g^{\nu\lambda} \delta^\sigma_\kappa \delta^\rho_\lambda + (\partial^\mu A^\nu) \delta^\sigma_\mu \delta^\rho_\nu \right)$$

$$= C_1 \partial_\sigma (\partial^\sigma A^\rho + \partial^\sigma A^\rho)$$

$$= 2C_1 \partial_\sigma \partial^\sigma A^\rho . \quad (6.20)$$

Following similar steps we can compute

$$\partial_\sigma \left(\frac{\partial}{\partial (\partial_\sigma A_\rho)} (C_2 \partial^\mu A^\nu \partial_\nu A_\mu) \right) = 2C_2 \partial^\rho (\partial_\sigma A^\sigma).$$

Therefore the equation of motion, following from the Lagrangian in Eq. 6.19 reads

$$2C_3 A^\rho = 2C_1 \partial_\sigma \partial^\sigma A^\rho + 2C_2 \partial^\rho (\partial_\sigma A^\sigma),$$

which gives us, when we put in the conventional constants[25]

$$\to m^2 A^\rho = \partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma). \tag{6.21}$$

This is called the **Proca equation**, which is the equation of motion for massive spin 1 particles and fields. For massless ($m = 0$) spin 1 particles, e.g. photons, the equation reads

$$\to 0 = \partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma). \tag{6.22}$$

This is the **inhomogeneous Maxwell equation** in absence of electric currents. To unclutter the notation it is conventional to define the electromagnetic tensor

$$F^{\sigma\rho} := \partial^\sigma A^\rho - \partial^\rho A^\sigma. \tag{6.23}$$

Then the inhomogeneous Maxwell-equations read

$$\partial_\sigma F^{\sigma\rho} = 0 \tag{6.24}$$

and we can rewrite the Lagrangian for massless spin 1 fields

$$\mathscr{L}_{\text{Maxwell}} = \frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu)$$

as

$$\begin{aligned}
\mathscr{L}_{\text{Maxwell}} &= \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \\
&= \frac{1}{4}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) \\
&= \frac{1}{4}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\nu A^\mu \partial_\mu A_\nu + \partial^\nu A^\mu \partial_\nu A_\mu) \\
&\underbrace{=}_{\text{renaming dummy indices}} \frac{1}{4}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu A_\mu + \partial^\mu A^\nu \partial_\mu A_\nu) \\
&= \frac{1}{4}(2\partial^\mu A^\nu \partial_\mu A_\nu - 2\partial^\mu A^\nu \partial_\nu A_\mu) \\
&= \frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu) \quad \checkmark. \tag{6.25}
\end{aligned}$$

$\mathscr{L}_{\text{Maxwell}} = \frac{1}{4} F^{\mu\nu} F_{\mu\nu}$ is the conventional way to write the Lagrangian. Equivalently, the Lagrangian for a massive spin 1 field can be written as

$$\mathscr{L}_{\text{Proca}} = \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + m^2 A_\mu A^\mu = \frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu) + m^2 A_\mu A^\mu. \tag{6.26}$$

In this chapter we derived the equations of motion that describe free fields and particles. We want to understand what we can do with these equations in order to get predictions for experiments. However

first, we need to derive some more equations, because experiments always work through interactions. For example, we are only able to detect an electron if we use another particle, like a photon. Therefore, in the next chapter we will derive Lagrangians that describe the interaction between different fields and particles.

7

Interaction Theory

Summary

In this chapter we will derive how different fields and particles interact with each other. This will enable us, for example, to describe how electrons, interact with photons[1].

We will be guided to the correct form of the Lagrangians by internal symmetries, which are in this context often called gauge symmetries[2]. The starting point will be *local* $U(1)$ symmetry[3] and we end up with the Lagrangian

$$ \mathscr{L} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu \partial^\mu \Psi + A_\mu \bar{\Psi}\gamma^\mu \Psi + \frac{1}{2}\left(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu\right). $$

This is the Lagrangian of **quantum electrodynamics**. This Lagrangian describes the interaction between charged, massive spin $\frac{1}{2}$ fields and a massless spin 1 field (the photon field). The Lagrangian is only then locally $U(1)$ invariant, if we avoid "mass terms" of the form $mA_\mu A^\mu$ in the Lagrangian. This coincides with the experimental fact that photons, described by A^μ, are massless. Using Noether's theorem, we can derive a new conserved quantity from $U(1)$ symmetry, which is commonly interpreted as **electric charge**.

Then we move on to local $SU(2)$ symmetry. For this purpose a two component object

$$ \bar{\Psi} := \begin{pmatrix} \bar{\psi}_1 & \bar{\psi}_2 \end{pmatrix}, $$

called **doublet**, is introduced. Such a doublet contains two spin $\frac{1}{2}$ fields, for example, the electron and the electron neutrino field that are "rotated" by $SU(2)$ transformations into each other.

Using this doublet notation we are able to write down a locally $SU(2)$ invariant Lagrangian[4]

[1] From a different point of view: how the electron field (= a massive spin $\frac{1}{2}$ field) interacts with the photon field (= a massless spin 1 field).

[2] This name will be explained in a moment.

[3] This means we multiply the field at each point in spacetime with a different factor: $e^{i\alpha(x)}$, instead of using the same transformation everywhere: $e^{i\alpha}$. In other words: the transformation parameter $\alpha = \alpha(x)$ is now a function of x and has therefore a different value for different points in spacetime.

[4] The object $W_j^{\mu\nu}$ is for the three fields W_j^μ what $F_{\mu\nu}$ is for the $U(1)$ gauge field A^μ.

$$\mathscr{L} = i\bar{\Psi}\gamma_\mu\left(\partial^\mu - ig\mathcal{W}^\mu\right)\Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right),$$

which contains three spin 1 fields $\mathcal{W}^\mu \equiv (W^\mu)_i\frac{\sigma^i}{2}$, with $i = 1, 2, 3$. We need three fields to make the Lagrangian locally $SU(2)$ invariant, because we have three $SU(2)$ basis generators: $J_i = \frac{\sigma_i}{2}$. We will see that local $SU(2)$ symmetry can only be achieved without "mass terms" of the form $m\bar{\Psi}\Psi$, $m\mathcal{W}^\mu\mathcal{W}_\mu$, with some arbitrary mass **matrix** m, because Ψ is now a two-component object. So this time not only are the spin 1 fields $(W^\mu)_i$ required to be massless, but the spin $\frac{1}{2}$ fields, too. Also allowed are equal masses for the two spin $\frac{1}{2}$ fields, but from experiments we know that this is not the case: The electron mass is much bigger than the electron-neutrino mass. In addition, we know from experiments that the three spin 1 fields $(W^\mu)_i$ are not massless. This is commonly interpreted as the $SU(2)$ symmetry being **broken**.

This idea is the starting point for the **Higgs formalism**, which is introduced afterwards. This formalism enables us to get a locally $SU(2)$ invariant Lagrangian that includes mass terms. This works by adding the interaction with a spin 0 field, called Higgs field, into our considerations. The final locally $SU(2) \times U(1)$ invariant interaction Lagrangian describes the so called **electroweak interactions**. The notion electroweak interactions contains the electromagnetic interaction and a new type of interaction called the **weak interaction**. Through the Higgs formalism the $SU(2) \times U(1)$ is broken to a remnant $U(1)$ symmetry. The weak interaction is mediated by three massive spin 1 fields, called W^+, W^- and Z and the remnant $U(1)$ symmetry by one massless spin 1 field γ. Using Noether's theorem we will be able to derive from $SU(2)$ symmetry a new conserved quantity, called **isospin**, which is the charge of weak interactions analogous to electric charge for electromagnetic interactions.

Lastly, we will consider internal *local $SU(3)$* symmetry, which will lead us to a Lagrangian describing another new interaction, called the **strong interaction**. For this purpose, we will introduce triplet objects

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix},$$

that are transformed by $SU(3)$ transformations and which contain three spin $\frac{1}{2}$ fields. These three spin $\frac{1}{2}$ fields are interpreted as **quarks** carrying different **color**, which is the strong interaction analogue to the electrical charge of the electromagnetic interaction or isospin of the weak interaction. Again, mass terms are forbidden, but this time this coincides with the experimental fact that the 8 corresponding

bosons[5], called gluons, are massless. In addition, we know from experiments that the fields inside a $SU(3)$ triplet have the same mass, This is a good thing, because local $SU(3)$ symmetry forbids a term with arbitrary mass matrix m for terms like $m\bar{Q}Q$, but allows a term of the form $\bar{Q}\begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix} Q$, which means that the terms in the triplet have the same mass. Therefore local $SU(3)$ invariance provides no new obstacles regarding mass terms in the Lagrangian and we therefore say the $SU(3)$ symmetry is unbroken. From experiments we know that only quarks (spin $\frac{1}{2}$) and gluons (spin 1) carry color. The resulting Lagrangian

$$\mathscr{L} = -\frac{1}{4}\mathcal{G}_{\alpha\beta}\mathcal{G}^{\alpha\beta} + \overline{Q}(i(\partial_\mu - ig\mathcal{G}_\mu)\gamma^\mu - m)Q$$

will only be cited, because the derivation is completely analogous to what we did before.

To summarize the summary:

$U(1) \longrightarrow$ 1 gauge field $\longrightarrow$ **massless** photons $\longrightarrow$ electric charge

$SU(2) \longrightarrow$ 3 gauge fields $\longrightarrow$ **massive** W- and Z-bosons (Higgs needed) $\longrightarrow$ isospin

$SU(3) \longrightarrow$ 8 gauge fields $\longrightarrow$ **massless** gluons $\longrightarrow$ color charge

7.1 $U(1)$ **Interactions**

To derive the correct interaction terms in the Lagrangian, we are going to use internal symmetrie that are usually called gauge symmetries. The notion gauge symmetry, is used for historic reasons and doesn't make much sense for the type of symmetry we are considering here. Weyl tried to derive electromagnetism[6]

> "as a consequence of spacetime symmetry, specifically symmetry under local changes of length scale."

Naming this kind of symmetry **gauge symmetry** makes sense, because this means, for example, that we can change the platinum bar that defines a standard meter (and which was used to gauge objects that measure length in experiments), arbitrarily without changing physics. This attempt was unsuccessful, but some time later, Weyl found the correct symmetry to derive electromagnetism and the name was kept.

7.1.1 Internal Symmetry of Free Spin $\frac{1}{2}$ Fields

Let's have a look again at the Lagrangian that we derived for a free spin $\frac{1}{2}$ theory (Eq. 6.16)

$$\mathscr{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi = \bar{\Psi}(i\gamma_\mu\partial^\mu - m)\Psi. \qquad (7.1)$$

We derived it by demanding Lorentz symmetry, but if we take a sharp look we can discover another symmetry of this Lagrangian. The Lagrangian does not change if we transform the field Ψ as follows

$$\Psi \rightarrow \Psi' = e^{ia}\Psi$$

if we take into account that this implies that $\bar{\Psi}$ also gets transformed

$$\Rightarrow \bar{\Psi} \rightarrow \bar{\Psi}' = \Psi'^\dagger\gamma_0 = (e^{ia}\Psi)^\dagger\gamma_0 = \bar{\Psi}e^{-ia}. \qquad (7.2)$$

[7] Remember: $\bar{\Psi} = \Psi^\dagger\gamma_0$.

The minus sign comes from the complex conjugation[7] and a is an arbitrary real number. To see that the Lagrangian is invariant under this transformation, we transform the Lagrangian explicitly:

$$
\begin{aligned}
\mathscr{L}'_{\text{Dirac}} &= -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' \\
&= -m(\bar{\Psi}e^{-ia})(e^{ia}\Psi) + i(\bar{\Psi}e^{-ia})\gamma_\mu\partial^\mu(e^{ia}\Psi) \\
&= -m\bar{\Psi}\Psi\underbrace{e^{-ia}e^{ia}}_{=1} + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi\underbrace{e^{-ia}e^{ia}}_{=1} \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi = \mathscr{L}_{\text{Dirac}}, \qquad (7.3)
\end{aligned}
$$

[8] Speaking more technically: A complex number commutes with every matrix, like, for example, γ_μ.

where we used that e^{ia} is just a complex number, which we can move around freely[8]. Remembering that we learned in Chapter 3 that all unit complex numbers can be written as e^{ia} and form a group called $U(1)$, we can put what we just discovered into mathematical terms, by saying that the Lagrangian is $U(1)$ invariant. This symmetry is an internal symmetry, because it is clearly no spacetime transformation and therefore transforms the field internally. This internal symmetry does not look like a big thing. At a first glance it may seem like a nice, but rather useless, mathematical side note. However, quite surprisingly we will see in a moment that this observation is incredibly important!

Now let's have a deeper look at what we just discovered. We showed that we are free to multiply our field with an arbitrary unit complex number without changing anything. The symmetry transformation $\Psi \rightarrow \Psi' = e^{ia}\Psi$ is called a **global** transformation, because we multiply the field $\Psi = \Psi(x)$ at every point x with the same factor e^{ia}.

Now, why should this factor at one point in spacetime be correlated to the factor at another point in spacetime? The choice at

one point in spacetime shouldn't fix this immediately in the whole universe. This would be strange, because special relativity tells us that no information can spread faster than light, as was shown in Section 2.4. For a global symmetry the choice would be fixed immediately for any point in the whole universe.

Let's check if our Lagrangian is invariant if we transform each point in spacetime with a different factor $a = a(x)$. This is called a **local** transformation.

If we transform

$$\Psi \rightarrow \Psi' = e^{ia(x)}\Psi$$

$$\Rightarrow \bar{\Psi} \rightarrow \bar{\Psi}' = e^{-ia(x)}\bar{\Psi}, \tag{7.4}$$

where the factor $a = a(x)$ now depends on the position, we get the transformed Lagrangian[9]

$$\mathscr{L}'_{\text{Dirac}} = -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi'$$

$$= -m(\bar{\Psi}\underbrace{e^{-ia(x)})(e^{ia(x)}}_{=1}\Psi) + i(\bar{\Psi}e^{-ia(x)})\gamma_\mu\partial^\mu(e^{ia(x)}\Psi)$$

$$\underbrace{=}_{\text{Product rule}} -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu(\partial^\mu\Psi)\underbrace{e^{-ia(x)}e^{ia(x)}}_{=1} + i(e^{-ia(x)}\bar{\Psi})\gamma_\mu\Psi(\partial^\mu e^{ia(x)})$$

$$= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i^2(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi \neq \mathscr{L}_{\text{Dirac}} \tag{7.5}$$

Therefore, our Lagrangian is *not* invariant under local $U(1)$ symmetry, because the product rule produces an extra term. As discussed above, our Lagrangian should be locally invariant, but we just found out that it isn't. There is something we can do about it, but first we must investigate another symmetry.

7.1.2   Internal Symmetry of Free Spin 1 Fields

Next, let's take a look at the Lagrangian we derived for free spin 1 particles[10]

$$\mathscr{L}_{\text{Proca}} = \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu + m^2 A^\mu A_\mu. \tag{7.6}$$

We can discover a **global** internal symmetry here, too. If we transform

$$A_\mu \rightarrow A'_\mu = A_\mu + a_\mu \tag{7.7}$$

with some arbitrary constants a_μ[11], the Lagrangian reads[12]

$$\mathscr{L}'_{\text{Proca}} = (\partial^\mu A'^\nu\partial_\mu A'_\nu - \partial^\mu A'^\nu\partial_\nu A'_\mu) + m^2 A'^\mu A'_\mu$$

$$= \partial^\mu(A^\nu + a^\nu)\partial_\mu(A_\nu + a_\nu) - \partial^\mu(A^\nu + a^\nu)\partial_\nu(A_\mu + a_\mu)) + m^2(A_\mu + a_\mu)(A^\mu + a^\mu)$$

$$= \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu + m^2(A_\mu + a_\mu)(A^\mu + a^\mu). \tag{7.8}$$

[9] Maybe you wonder if the Lagrangian that includes both possible derivatives, which we neglected for brevity, is locally $U(1)$ invariant: $\mathscr{L} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i(\partial^\mu\bar{\Psi})\gamma_\mu\Psi$. This Lagrangian is indeed locally $U(1)$ invariant, as you can check, but take note that the addition of the second and the third term yields zero:

$$i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i(\partial^\mu\bar{\Psi})\gamma_\mu\Psi$$

$$\underbrace{=}_{\text{integration by parts}} i(\partial^\mu\bar{\Psi})\gamma_\mu\Psi - i(\partial^\mu\bar{\Psi})\gamma_\mu\Psi = 0.$$

The correct Lagrangian that includes both possible derivatives has a minus sign between those terms: $\mathscr{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - i(\partial^\mu\bar{\Psi})\gamma_\mu\Psi$ and is therefore **not** locally $U(1)$ invariant.

[10] See Eq. 6.26 and take note that we neglect, for brevity, a conventional factor $\frac{1}{2}$ here.

[11] Remember, as always, $\mu = 0, 1, 2, 3$.

[12] The a_μ are constants and therefore $\partial^\mu a^\nu = 0$ etc.

We conclude this transformation is a **global** symmetry transformation of this Lagrangian, if we restrict ourselves to **massless** fields, i.e. $m = 0$.

What about **local** symmetry here? We transform

$$A_\mu \to A'_\mu = A_\mu + a_\mu(x) \tag{7.9}$$

and the transformed **massless** Lagrangian reads

$$
\begin{aligned}
\mathscr{L}'_{\text{Maxwell}} &= (\partial^\mu A'^\nu \partial_\mu A'_\nu - \partial^\mu A'^\nu \partial_\nu A'_\mu) \\
&= \partial^\mu (A^\nu + a^\nu(x)) \partial_\mu (A_\nu + a_\nu(x)) - \partial^\mu (A^\nu + a^\nu(x)) \partial_\nu (A_\mu + a_\mu(x)) \\
&= \partial^\mu A^\nu \partial_\mu A_\nu + \partial^\mu a^\nu \partial_\mu A_\nu + \partial^\mu A^\nu \partial_\mu a_\nu(x) + \partial^\mu a^\nu(x) \partial_\mu a_\nu(x) \\
&\quad - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu a_\mu(x) - \partial^\mu a^\nu(x) \partial_\nu A_\mu - \partial^\mu a^\nu(x) \partial_\nu a_\mu(x),
\end{aligned}
\tag{7.10}
$$

which shows that $A_\mu \to A'_\mu = A_\mu + a_\mu(x)$ is *not* a local internal symmetry .

Nevertheless, we can find a local internal symmetry by considering the transformation $A_\mu \to A'_\mu = A_\mu + \partial_\mu a(x)$ instead. This means we add the derivative of some arbitrary function $\partial_\mu a(x)$ instead of an arbitrary function. This transformation has the following effect on the Lagrangian[13]

[13] The symmetry of partial derivatives $\partial_\nu \partial_\mu = \partial_\mu \partial_\nu$ is also known as "Schwarz's theorem".

$$
\begin{aligned}
\mathscr{L}'_{\text{Maxwell}} &= \partial^\mu A'^\nu \partial_\mu A'_\nu - \partial^\mu A'^\nu \partial_\nu A'_\mu \\
&= \partial^\mu (A^\nu + \partial^\nu a(x)) \partial_\mu (A_\nu + \partial_\nu a(x)) - \partial^\mu (A^\nu + \partial^\nu a(x)) \partial_\nu (A_\mu + \partial_\mu a(x)) \\
&= \partial^\mu A^\nu \partial_\mu A_\nu + \partial^\mu (\partial^\nu a(x)) \partial_\mu A_\nu + \partial^\mu A^\nu \partial_\mu (\partial_\nu a(x)) + \partial^\mu (\partial^\nu a(x)) \partial_\mu (\partial_\nu a(x)) \\
&\quad - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu (\partial_\mu a(x)) - \partial^\mu (\partial^\nu a(x)) \partial_\nu A_\mu - \partial^\mu (\partial^\nu a(x)) \partial_\nu (\partial_\mu a(x)) \\
&\underbrace{=} \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu = \mathscr{L}_{\text{Maxwell}} . \tag{7.11}
\end{aligned}
$$

$\partial_\nu \partial_\mu = \partial_\mu \partial_\nu$ and renaming dummy indices

We see this is indeed an internal local symmetry transformation. Again this may look like a technical side note. Okay, we found some internal, local symmetry; so what?

7.1.3   Putting the Puzzle Pieces Together

Let's summarize what we found out so far:

- We discovered the Lagrangian for free spin $\frac{1}{2}$ fields has an internal global symmetry $\Psi \to \Psi' = e^{ia} \Psi$. Formulated differently: the Lagrangian for free spin $\frac{1}{2}$ fields is invariant under global $U(1)$ transformations.

- We saw that this symmetry is not local (although it should be), because for $a = a(x)$, we get an extra term in the Lagrangian of the form (Eq. 7.5).

$$-(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi. \tag{7.12}$$

In other words: The Lagrangian isn't locally $U(1)$ invariant.

- In the last section we found an internal local symmetry for massless spin 1 fields

$$A_\mu \to A'_\mu = A_\mu + \partial_\mu a(x), \tag{7.13}$$

which is only a local symmetry if we add the **derivative** of an arbitrary function $\partial_\mu a(x)$, instead of an arbitrary function $a_\mu(x)$.

This really looks like two pieces of a puzzle we should put together. When we transform Ψ, $\bar{\Psi}$ and A_μ simultaneously, an additional term $A_\mu\bar{\Psi}\gamma^\mu\Psi$ in the Lagrangian becomes

$$A_\mu\bar{\Psi}\gamma^\mu\Psi \to (A_\mu + \partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi = A_\mu\bar{\Psi}\gamma^\mu\Psi + \partial_\mu a(x)\bar{\Psi}\gamma^\mu\Psi. \tag{7.14}$$

Compare the second term here to Eq. 7.12. The new term in the Lagrangian, coupling Ψ, $\bar{\Psi}$ and A_μ together, therefore cancels exactly the term which stopped the Lagrangian for free spin $\frac{1}{2}$ from being locally $U(1)$ invariant. In other words: **By adding this new term we can make the Lagrangian locally $U(1)$ invariant**.

Let's study this in more detail. First take note that it's conventional to factor out a constant g in the exponent of the local $U(1)$ transformation: $e^{iga(x)}$. Then the extra term becomes

$$-(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi \to -g(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi. \tag{7.15}$$

This extra factor g accounts for an arbitrary coupling constant[14], as we will see now. We then add to the Lagrangian for free spin $\frac{1}{2}$ fields the new term

$$gA_\mu\bar{\Psi}\gamma^\mu\Psi,$$

where we included γ^μ to make the term Lorentz invariant[15] and inserted the **coupling constant**[16] g. This yields the Lagrangian

$$\mathscr{L}_{\text{Dirac+Extra-Term}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi.$$

Transforming this Lagrangian according to the rules for local transformations of Ψ, $\bar{\Psi}$ and A_μ yields[17]

[14] A coupling constant always tells us how strong a given interaction is. Here we are talking about electromagnetic interactions and g determines its strength.

[15] Otherwise A_μ has an unmatched index μ and therefore wouldn't be Lorentz invariant.

[16] We can see here that g determines how strong Ψ, $\bar{\Psi}$ and A_μ couple together.

[17] The combined transformation of Ψ, $\bar{\Psi}$ and A_μ is called $U(1)$ **gauge transformation**.

$$\mathscr{L}'_{\text{Dirac+Extra-Term}} = -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' + gA'_\mu\bar{\Psi}'\gamma^\mu\Psi'$$

$$\underbrace{=}_{\text{See Eq. 7.5}} -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - g(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi + gA'_\mu\bar{\Psi}'\gamma^\mu\Psi'$$

$$= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - g(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi$$
$$+ g(A_\mu + \partial_\mu a(x))(e^{-iga(x)}\bar{\Psi})\gamma^\mu(e^{iga(x)}\Psi)$$

$$= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \cancel{g(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi}$$
$$+ gA_\mu\bar{\Psi}\gamma^\mu\Psi + g\underbrace{\cancel{(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi}}_{=(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi}$$

$$= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi = \mathscr{L}_{\text{Dirac+Extra-Term}}.$$
$$(7.16)$$

Therefore, by adding an extra term we get a locally $U(1)$ invariant Lagrangian. To describe a system consisting of massive spin $\frac{1}{2}$ and massless spin 1 fields we must add the Lagrangian for free massless spin 1 fields to the Lagrangian as well. This gives us the complete Lagrangian[18]

$$\mathscr{L}_{\text{Dirac+Extra-Term+Maxwell}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi$$
$$-\frac{1}{2}(\partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu). \quad (7.17)$$

To unclutter the notation it is conventional to introduce a new symbol

$$D^\mu \equiv i\partial^\mu - igA^\mu, \quad\quad\quad (7.18)$$

called **covariant derivative**. The Lagrangian then reads

$$\mathscr{L}_{\text{Dirac+Extra-Term+Maxwell}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\underbrace{(\partial^\mu - igA_\mu)}_{\equiv D^\mu}\Psi$$
$$-\frac{1}{2}(\partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu)$$
$$= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu D^\mu\Psi$$
$$-\frac{1}{2}(\partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu). \quad (7.19)$$

This is the correct Lagrangian for the quantum field theory of electrodynamics, commonly called **quantum electrodynamics**. We are able to derive this Lagrangian simply by making use of internal symmetries of the Lagrangians describing free spin $\frac{1}{2}$ fields and free spin 1 fields.

The next question we have to answer is: What equations of motion follow from this Lagrangian?

[18] We use here the conventional "normalization" with an additional factor $\frac{1}{2}$ in front of the last terms.

7.1.4   Inhomogeneous Maxwell Equations and Minimal Coupling

To spoil the surprise: This Lagrangian gives us the inhomogeneous Maxwell equations in the presence of currents.

The process is again straightforward: we simply put the Lagrangian (Eq. 7.17)

$$\mathscr{L}_{\text{Dirac+Extra-Term+Maxwell}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi$$
$$ - \frac{1}{2}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu).$$

into the Euler-Lagrange equation for each field

$$\frac{\partial \mathscr{L}}{\partial \Psi} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \Psi)} \right) = 0$$

$$\frac{\partial \mathscr{L}}{\partial \bar{\Psi}} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial(\partial_\mu \bar{\Psi})} \right) = 0$$

$$\frac{\partial \mathscr{L}}{\partial A_\rho} - \partial_\sigma \left(\frac{\partial \mathscr{L}}{\partial(\partial_\sigma A_\rho)} \right) = 0.$$

This yields

$$\bar{\Psi}(i\gamma_\mu\partial^\mu + m) + gA_\mu\bar{\Psi}\gamma^\mu = 0 \qquad (7.20)$$

$$(i\gamma_\mu\partial^\mu - m)\Psi + gA_\mu\gamma^\mu\Psi = 0 \qquad (7.21)$$

$$\partial_\nu(\partial^\nu A^\mu - \partial^\mu A^\nu) + g\bar{\Psi}\gamma^\mu\Psi = 0. \qquad (7.22)$$

The first two equations describe the behavior of spin $\frac{1}{2}$ particles/fields in an external electromagnetic field. In many books the derivation of these equations uses the notion **minimal coupling**, by which is meant that in the presence of an external field the derivative ∂^μ has to be changed into the covariant derivative

$$\partial^\mu \to D^\mu = \partial^\mu - igA^\mu \qquad (7.23)$$

to yield the correct equations. The word "minimal" is used, because only one gauge field A_μ, with four components $\mu = 0, 1, 2, 3$, is used.

Now that we have the equation that describes how Dirac spinors behave in the presence of an external electromagnetic field (Eq. 7.21), we can show something that we promised in Section 3.7.10. There we claimed that a transformation, which we called very suggestively charge conjugation, changes the electric charge of the object it describes. In other words, if Ψ describes something of charge $+e$, the charge conjugate spinor Ψ^C describes something of charge $-e$.

Electrical charge determines the coupling strength of spin $\frac{1}{2}$ particles/fields to an external spin 1 field and we therefore investigate now, which equation of motion holds for Ψ^C. Afterwards we will talk about the third equation, i.e. Eq. 7.22.

7.1.5   Charge Conjugation, Again

Before we can derive the corresponding equation, we need to find an explicit form of the charge conjugation operator for Dirac spinors. We derived in Section 3.7.10 the transformation (Eq. 3.244)

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \rightarrow \Psi^C = \begin{pmatrix} \xi_L \\ \chi_R \end{pmatrix}. \tag{7.24}$$

This transformation can now be described easily with the help of one of the γ_μ matrices. Using the definition of γ_2 in Eq. 6.13, we have

$$\Psi^C = i\gamma_2 \Psi^\star = i \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} \begin{pmatrix} \chi_L^\star \\ \xi_R^\star \end{pmatrix}, \tag{7.25}$$

because we can rewrite this, using that $i\sigma_2 = \epsilon$ is exactly the spinor metric

$$= \begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \begin{pmatrix} \chi_L^\star \\ \xi_R^\star \end{pmatrix} = \begin{pmatrix} \epsilon\xi_R^\star \\ -\epsilon\chi_L^\star \end{pmatrix}. \tag{7.26}$$

This is equivalent to

$$= \begin{pmatrix} \xi_L \\ \chi_R \end{pmatrix}, \tag{7.27}$$

as was shown in Section 3.7.7, specifically Eq. 3.202. Therefore we start with Eq. 7.21:

$$(i\gamma_\mu\partial^\mu - m)\Psi + gA_\mu\gamma^\mu\Psi = (\gamma_\mu(i\partial^\mu + gA^\mu) - m)\Psi = 0, \tag{7.28}$$

and complex conjugate this equation, as a first step towards an equation for Ψ^C:

$$\rightarrow (\gamma_\mu^\star(-i\partial^\mu + gA^\mu) - m)\Psi^\star = 0. \tag{7.29}$$

Now, we multiply this equation from the left-hand side with γ_2 and add a $1 = \gamma_2^{-1}\gamma_2$ in front of $\Psi^\star$:

$$\rightarrow \gamma_2(\gamma_\mu^\star(-i\partial^\mu + gA^\mu) - m)\underbrace{\gamma_2^{-1}\gamma_2}_{=1}\Psi^\star = 0 \tag{7.30}$$

$$\rightarrow \Big(\underbrace{\gamma_2\gamma_\mu^\star\gamma_2^{-1}}_{=-\gamma_\mu}(-i\partial^\mu + gA^\mu) - m\gamma_2\gamma_2^{-1}\Big)\gamma_2\Psi^\star = 0 \tag{7.31}$$

$$\underbrace{\rightarrow}_{\text{Multiplying the equation with } i} \left(-\gamma_\mu(-i\partial^\mu + gA^\mu) - m \right) \underbrace{i\gamma_2\Psi^\star}_{=\Psi^C \text{ see Eq. 7.25}} = 0 \qquad (7.32)$$

$$\rightarrow \left((\gamma_\mu(i\partial^\mu - gA^\mu) - m)\Psi^C = 0 \right). \qquad (7.33)$$

This is exactly the same equation of motion as for Ψ, but with opposite coupling strength $g \rightarrow -g$. This justifies the name charge conjugation[19].

Next, we turn to the third equation of motion derived in the last section (Eq. 7.22) which is called **inhomogeneous Maxwell equation in the presence of an electric current**. To make the notion "electric current" precise we need again Noether's theorem.

7.1.6   Noether's Theorem for Internal $U(1)$ Symmetry

In Section 4.5.5 we learned that Noether's theorem connects each internal symmetry with a conserved quantity. What conserved quantity follows from the $U(1)$ symmetry we just discovered? Noether's theorem for internal symmetries tells us that a transformation of the form

$$\Psi \rightarrow \Psi' = \Psi + \delta\Psi$$

leads to a Noether current

$$J^\mu = \frac{\partial \mathscr{L}}{\partial(\partial_\mu\Psi)}\delta\Psi$$

which fulfils a continuity equation

$$\partial_\mu J^\mu = 0. \qquad (7.34)$$

A global[20] $U(1)$ transformation is

$$\Psi \rightarrow \Psi' = e^{iga}\Psi = (1 + iga + \ldots)\Psi.$$

We stop the series expansion of the exponential function, as usual, after the first term, because $U(1)$ is a Lie group and arbitrary transformations can be built of infinitesimal ones. An infinitesimal transformation reads

$$\Psi \rightarrow \Psi' = \Psi + iga\Psi.$$

Therefore we have $\delta\Psi = iga\Psi$ and as we derived in Section 4.5.5 the corresponding Noether current is

$$\begin{aligned}
J^\mu &= \frac{\partial \mathscr{L}}{\partial(\partial_\mu\Psi)}\delta\Psi \\
&= \frac{\partial(-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi)}{\partial(\partial_\mu\Psi)}iga\Psi \\
&= -\bar{\Psi}\gamma^\mu ga\Psi = -ga\bar{\Psi}\gamma^\mu\Psi. \qquad (7.35)
\end{aligned}$$

[19] It is important to note that $\Psi^c \neq \bar{\Psi}$. $\Psi^c = i\gamma_2\Psi^\star$ and $\bar{\Psi} = \Psi^\dagger\gamma_0 = (\Psi^\star)^T\gamma_0$. Charge conjugation is the correct transformation that enables us to interpret things in terms of antiparticles, as we will discuss later in detail.

[20] Recall that the Lagrangian for free spin $\frac{1}{2}$ fields was only *globally* $U(1)$ invariant. The final Lagrangian of the last section was *locally* $U(1)$ invariant. Global symmetry is a special case of local symmetry with $a = $const. Therefore, if we have a locally $U(1)$ invariant Lagrangian, it is automatically globally $U(1)$ invariant, too. Considering **global** $U(1)$ symmetry here will give us a quantity that is conserved for free **and** interacting fields.

We can ignore[21] the arbitrary constant a, because the continuity equation holds for arbitrary a and therefore, we define

$$J^\mu \equiv -g\bar{\Psi}\gamma^\mu\Psi. \tag{7.36}$$

This is usually called the **electric four-current**. The zeroth component is the **electric charge density**, which gives us if we integrate it, a quantity that is conserved in time[22]

[22] This can be seen by following the same steps as in Eq. 4.39.

$$Q = \int d^3x \underbrace{\rho}_{\text{Charge density}} = \int d^3x J^0 = -g\int d^3x \bar{\Psi}\gamma^0\Psi. \tag{7.37}$$

In the quantum framework the objects Ψ will be related to probability amplitudes and this interpretation requires that $\int d^3x \bar{\Psi}\gamma^0\Psi = 1$, because the overall probability must be $100\% = 1$. Therefore, the conserved quantity is in fact the coupling strength g, which is for electromagnetism the electric charge. **Therefore, global $U(1)$ symmetry leads to the conservation of electric charge**.

If we now take a look again at Eq. 7.22, we can write it, using the definition in Eq. 7.36, as

$$\partial_\nu(\partial^\nu A^\mu - \partial^\mu A^\nu) + \underbrace{g\bar{\Psi}\gamma^\mu\Psi}_{=-J^\mu} = 0$$

$$\to \partial_\nu(\partial^\nu A^\mu - \partial^\mu A^\nu) = J^\mu. \tag{7.38}$$

Using the electromagnetic tensor as defined in Eq. 6.23 this equation reads

$$\partial_\nu F^{\nu\mu} = J^\mu. \tag{7.39}$$

These are the inhomogeneous Maxwell equations in the presence of an external electromagnetic current. These equations[23], together with the homogeneous Maxwell equations, which follow immediately from the definition of $F^{\nu\mu}$, are the basis for the classical theory of electrodynamics[24].

[23] Plural, because we have an equation for each component μ.

[24] We will see this explicitly in Chapter 11.

Next we take a quick look at interactions of **massive** spin 1 and spin 0 fields.

7.1.7 Interaction of Massive Spin 0 Fields

Take note that the Lagrangian we derived for spin 0 fields

$$\mathscr{L} = \frac{1}{2}(\partial_\mu\Phi\partial^\mu\Phi - m^2\Phi^2)$$

is not $U(1)$ invariant, as we can see by transforming $\Phi \to \Phi' = e^{ia}\Phi$. Nevertheless, the complex scalar theory

$$\mathscr{L} = \frac{1}{2}(\partial_\mu\Phi^\star\partial^\mu\Phi - m^2\Phi^\star\Phi). \tag{7.40}$$

has $U(1)$ symmetry, because then we have $\Phi \rightarrow \Phi' = e^{ia}\Phi$ and $\Phi^\star \rightarrow (\Phi^\star)' = e^{-ia}\Phi^\star$. Therefore it is possible to derive, analogous to what we did in Section 7.1 for spin $\frac{1}{2}$ fields, an interaction theory for this Lagrangian. The derivation is completely analogous[25] and one gets

[25] The correct Lagrangian can be computed by substituting $\partial^\mu \rightarrow D^\mu = \partial^\mu - igA^\mu$ as introduced in Eq. 7.23.

$$\mathscr{L} = \frac{1}{2}\left(\left((\partial_\mu + iqA_\mu)\Phi^\star\right)\left((\partial^\mu - iqA^\mu)\Phi\right) - m^2\Phi^\star\Phi\right). \qquad (7.41)$$

Using the Euler-Lagrange equation

$$\frac{\partial\mathscr{L}}{\partial\Phi^\star} - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\Phi^\star)}\right) = 0,$$

we find the corresponding equation of motion

$$(\partial_\mu - iqA_\mu)(\partial^\mu - iqA^\mu)\Phi - m^2\Phi = 0, \qquad (7.42)$$

which describes a charged spin 0 field coupled to a massless spin 1 field.

7.1.8 Interaction of Massive Spin 1 Fields

The interaction of a **massive** spin 1 field with a **massless** spin 1 field is dictated by symmetry, too. The Lagrangian for massless spin 1 fields is given by (Eq. 6.25)

$$\mathscr{L}_{\text{Maxwell}} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu}.$$

To distinguish between a massless and a massive spin 1 field, we name the massive field B^μ and define

$$G^{\mu\nu} := \partial^\mu B^\nu - \partial^\nu B^\mu.$$

The Lagrangian for this massive spin 1 field reads (Eq. 6.19)

$$\mathscr{L}_{\text{Proca}} = -\frac{1}{4}G^{\mu\nu}G_{\mu\nu} + m^2 B^\mu B_\mu.$$

Lorentz symmetry dictates the interaction term in the Lagrangian to be of the form

$$\mathscr{L}_{\text{Proca-interaction}} = CG_{\mu\nu}F^{\mu\nu},$$

with a coupling constant C that we need to measure in experiments. If you're interested you can derive yourself the corresponding equations of motion, by using the Euler-Lagrange equations.

7.2 $SU(2)$ Interactions

Motivated by the success with $U(1)$ symmetry we want to answer the question: Is $U(1)$ the only internal symmetry of our Lagrangians?

It turns out that we can find an internal symmetry for **two massless** spin $\frac{1}{2}$ fields. We get the Lagrangian for two spin $\frac{1}{2}$ fields by adding two copies of the Lagrangian that we derived in Section 6.3. The final Lagrangian can be found in Eq. 6.16 and we recite it here for convenience:

$$\mathscr{L}_{\text{Dirac}} = \bar{\psi}(i\gamma_\mu\partial^\mu - m)\psi.$$

Here we neglect mass terms, which means $m = 0$, because otherwise the Lagrangian isn't invariant as we will see in a moment. We will see later how we can include mass terms, without spoiling the symmetry. The addition yields

$$\mathscr{L}_{\text{D1+D2}} = i\bar{\psi}_1\gamma_\mu\partial^\mu\psi_1 + i\bar{\psi}_2\gamma_\mu\partial^\mu\psi_2\,. \tag{7.43}$$

This can be rewritten, if we define

$$\Psi := \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

$$\rightarrow \bar{\Psi} := \begin{pmatrix} \bar{\psi}_1 & \bar{\psi}_2 \end{pmatrix}\,,$$

where the newly defined object Ψ is called a **doublet**:

$$\mathscr{L}_{\text{D1+D2}} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi. \tag{7.44}$$

[26] To calculate the transformation behavior of $\bar{\Psi}$, we use $\sigma_i^\dagger = \sigma_i$, which we already noted in Eq. 3.213.

This Lagrangian is invariant under **global** $SU(2)$ transformations[26]

$$\Psi \rightarrow \Psi' = e^{ia_i\frac{\sigma_i}{2}}\Psi \tag{7.45}$$

$$\Rightarrow \bar{\Psi} \rightarrow \bar{\Psi}' = \bar{\Psi}e^{-ia_i\frac{\sigma_i}{2}}\,, \tag{7.46}$$

where a sum over the index "i" is implicitly assumed, a_i denotes arbitrary real constants and $\frac{\sigma_i}{2}$ are the usual generators of $SU(2)$, with the Pauli matrices σ_i.

To see the invariance we take a look at the transformed Lagrangian[27]

[27] We neglect mass terms here, because they are, in general, not invariant under $SU(2)$ transformations. The mass terms would be $-m_1\bar{\Psi}_1\Psi_1$ and $-m_2\bar{\Psi}_2\Psi_2$ and we could write them, using the two component definition for Ψ and by defining

$$m := \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$$

as

$$\mathscr{L}_{\text{D1+D2}} = -\bar{\Psi}m\Psi.$$

Such a term is not invariant under $SU(2)$ transformations, because, in general,

$$\mathscr{L}_{\text{D1+D2}} = -\bar{\Psi}'m\Psi' = \bar{\Psi}\underbrace{e^{-ia_i\frac{\sigma_i}{2}}me^{ia_i\frac{\sigma_i}{2}}}_{\neq m}\Psi.$$

For equal masses $m_1 = m_2$ it would be invariant, but we are going to see how we can include arbitrary mass terms without violating this symmetry. We know from experiments that the two fields in the doublet do not create particles of equal mass, i.e. $m_1 \neq m_2$. This will be discussed later in detail.

$$\begin{aligned} \mathscr{L}'_{\text{D1+D2}} &= i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' \\ &= i\bar{\Psi}e^{-ia_i\frac{\sigma_i}{2}}\gamma_\mu\partial^\mu e^{ia_i\frac{\sigma_i}{2}}\Psi \\ &= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi = \mathscr{L}_{\text{D1+D2}}, \qquad \checkmark \end{aligned} \tag{7.47}$$

where we got to the last line because our transformation $e^{ia_i\frac{\sigma_i}{2}}$ acts on our newly defined two-component object Ψ, whereas γ_μ acts on the objects in our doublet, i.e. the Dirac spinors. We can express this using indices

$$\left[\left(e^{-ia_i\frac{\sigma_i}{2}}\right)_{ab}\delta_{\alpha\beta}\right]\left[\delta_{bc}\gamma_\mu^{\beta\delta}\right]\left[\left(e^{ia_i\frac{\sigma_i}{2}}\right)_{cd}\delta_{\delta\epsilon}\right] = \left[\delta_{ad}\gamma_\mu^{\alpha\epsilon}\right]\,.$$

This symmetry should be a *local* symmetry, too. The $SU(2)$ transformations mix the two components of the doublet. Later we will give these two fields names like electron and electron-neutrino field. Our symmetry here tells us that it does not matter what we call electron and what electron-neutrino field, because by using $SU(2)$ transformations we can mix them as we like. If this is only a global symmetry, as soon as we fix one choice[28], which means we decide what we call electron and what electron-neutrino field, this choice would be fixed immediately for the complete universe. Therefore we investigate if this is a local symmetry. Again we find that it isn't, but as for local $U(1)$ symmetry, we will do everything we can to make the Lagrangian locally $SU(2)$ invariant.

The problem here is again the derivative, which produces an extra term. To unclutter the notation, we define $U(x) \equiv e^{-ia_i(x)\frac{\sigma_i}{2}}$:

$$\Psi \to \Psi' = U(x)\Psi \tag{7.48}$$

$$\Rightarrow \bar{\Psi} \to \bar{\Psi}' = \bar{\Psi}U^\dagger(x). \tag{7.49}$$

Our transformed Lagrangian then reads

$$\begin{aligned}
\mathscr{L}'_{\text{D1+D2}} &= i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' \\
&= i\bar{\Psi}U^\dagger(x)\gamma_\mu\partial^\mu(U(x)\Psi) \\
&= \underbrace{}_{\text{product rule}} i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i\bar{\Psi}\gamma_\mu U^\dagger(x)(\partial_\mu U(x))\Psi \neq \mathscr{L}_{\text{D1+D2}}. \tag{7.50}
\end{aligned}$$

We can see that the Lagrangian is not invariant because we get an additional term, $i\bar{\Psi}\gamma_\mu U^\dagger(x)(\partial_\mu U(x))\Psi$, after the transformation.

So how can we modify our Lagrangian $\mathscr{L}_{\text{D1+D2}}$ such that is invariant under local $SU(2)$ transformations? In principle the same method that we discovered in the last sections to ensure local $U(1)$ works again. However, some details are a bit more complicated. We already noted in Eq. 7.23 that the essence of how we need to change our Lagrangian can be summarized through the replacement of the ordinary derivative ∂^μ, with a "covariant derivative" $D^\mu = \partial^\mu - igA^\mu$. With this in mind, we replace the derivative ∂^μ in our Lagrangian $\mathscr{L}_{\text{D1+D2}}$ with a new object D^μ and then try to derive how D^μ looks like in order to ensure local $SU(2)$ invariance.

Concretely, we now write

$$\widetilde{\mathscr{L}}_{\text{D1+D2}} = i\bar{\Psi}\gamma_\mu D^\mu\Psi \tag{7.51}$$

and under a local $SU(2)$ transformation this Lagrangian becomes

$$\widetilde{\mathscr{L}'_{\mathrm{D1+D2}}} = i\bar{\Psi}'\gamma_\mu(D^\mu\Psi)'$$
$$= i\bar{\Psi}U^\dagger(x)\gamma_\mu(D^\mu\Psi)'$$
$$\overset{!}{=} \widetilde{\mathscr{L}_{\mathrm{D1+D2}}}. \tag{7.52}$$

We can see here that this new Lagrangian is invariant under the local $SU(2)$ transformation $U(x) = \mathrm{e}^{-ia_i(x)\frac{\sigma_i}{2}}$ if $(D^\mu\Psi)' = U(x)D^\mu\Psi$, because then[29]

[29] The reason that we can change the positions of $U^\dagger(x)$ and γ_μ, as already discussed in the text below Eq. 7.47, is that γ_μ is a matrix that acts on the components of the Dirac spinors, whereas the $SU(2)$ transformation mixes the two Dirac spinors ψ_1 and ψ_2 inside the doublet Ψ.

$$\widetilde{\mathscr{L}'_{\mathrm{D1+D2}}} = i\bar{\Psi}'\gamma_\mu(D^\mu\Psi)'$$
$$= i\bar{\Psi}U^\dagger(x)\gamma_\mu(D^\mu\Psi)'$$
$$= i\bar{\Psi}\gamma_\mu \underbrace{U^\dagger(x)U(x)}_{=1} D^\mu\Psi$$
$$= \widetilde{\mathscr{L}_{\mathrm{D1+D2}}} \quad \checkmark. \tag{7.53}$$

So our goal is to find an object D^μ that has exactly this transformation behavior $(D^\mu\Psi)' = U(x)D^\mu\Psi$. Take note that this property is the reason why we call D^μ the "covariant derivative"[30]. The covariant derivative of Ψ, which we denote by $D^\mu\Psi$, transforms exactly like Ψ, i.e. $\Psi' = U(x)\Psi$. Thus the form stays the same and we don't get additional terms, like we do if we use the ordinary derivative ∂^μ[31].

[30] The notion "covariance" was discussed in Section 2.6 and means roughly that the form of an equation or an object is not changed under a given transformation.

[31] This was demonstrated in Eq. 7.50.

From our experience with local $U(1)$ symmetry, we already know that the crucial trick is to make use of spin 1 fields. However, there is one crucial difference. We saw in Eq. 7.50 that again the reason for the non-invariance is that the derivative ∂_μ produces an extra term $i\bar{\Psi}\gamma_\mu U^\dagger(x)(\partial_\mu U(x))\Psi$. Our local $SU(2)$ transformations $U(x) \equiv \mathrm{e}^{-ia_i(x)\frac{\sigma_i}{2}}$ are a bit more complicated than local $U(1)$ transformations $U(x) \equiv \mathrm{e}^{-i\alpha(x)}$, because of the generators $\sigma_i/2$ in the exponent. For local $SU(2)$ transformations, we have

$$\partial_\mu U(x) = \partial_\mu \mathrm{e}^{-ia_i(x)\frac{\sigma_i}{2}} = -i(\partial_\mu a_i(x))\frac{\sigma_i}{2}\mathrm{e}^{-ia_i(x)\frac{\sigma_i}{2}}. \tag{7.54}$$

[32] We have here an implicit sum, in the sense of Einstein's summation convention, over the index i.

We can see here that we actually get three extra terms[32], one for each Pauli matrix σ_i. Thus in contrast to the $U(1)$ case, to make the Lagrangian locally $SU(2)$ invariant we don't need one additional spin 1 field, but three! In addition, we can see that the troublesome terms, that are produced through the derivative in the Lagrangian ∂_μ, are proportional to the generators $\frac{\sigma_i}{2}$.

To summarize: we need three spin 1 fields to cancel the terms that make our Lagrangian non-invariant under local $SU(2)$ transformations and must introduce these new fields in such a way that they are able to cancel terms that involve the generators $\frac{\sigma_i}{2}$.

We therefore try

$$D^\mu = \partial^\mu - ig\frac{\sigma^a}{2}W_a^\mu , \qquad (7.55)$$

where σ^a are the Pauli matrices and $a = 1, 2, 3$. With this ansatz, the requirement $(D^\mu\Psi)' \overset{!}{=} U(x)D^\mu\Psi$, can now be translated into a transformation law for the spin 1 fields W_a^μ:

$$(D^\mu\Psi)' \overset{!}{=} U(x)D^\mu\Psi$$

$$\underset{\text{Eq. 7.55}}{\rightarrow} ((\partial^\mu - ig\frac{\sigma^a}{2}W_a^\mu)\Psi)' \overset{!}{=} U(x)(\partial^\mu - ig\frac{\sigma^a}{2}W_a^\mu)\Psi$$

$$\rightarrow \partial^\mu\Psi' - ig\frac{\sigma^a}{2}W_a'^\mu\Psi' \overset{!}{=} U(x)(\partial^\mu - ig\frac{\sigma^a}{2}W_a^\mu)\Psi$$

$$\underset{\text{Eq. 7.48}}{\rightarrow} \partial^\mu(U(x)\Psi) - ig\frac{\sigma^a}{2}W_a'^\mu U(x)\Psi \overset{!}{=} U(x)(\partial^\mu - ig\frac{\sigma^a}{2}W_a^\mu)\Psi$$

$$\underset{\text{product rule}}{\rightarrow} (\partial^\mu U(x))\Psi + \underline{U(x)(\partial^\mu\Psi)} - ig\frac{\sigma^a}{2}W_a'^\mu U(x)\Psi \overset{!}{=} \underline{U(x)\partial^\mu\Psi} - igU(x)\frac{\sigma^a}{2}W_a^\mu\Psi$$

$$\rightarrow (\partial^\mu U(x))\Psi - ig\frac{\sigma^a}{2}W_a'^\mu U(x)\Psi \overset{!}{=} -igU(x)\frac{\sigma^a}{2}W_a^\mu\Psi .$$

$$(7.56)$$

The Lagrangian should be invariant for arbitrary Ψ and therefore, we write the last line in Eq. 7.56 without it

$$(\partial^\mu U(x)) - ig\frac{\sigma^a}{2}W_a'^\mu U(x) \overset{!}{=} -igU(x)\frac{\sigma^a}{2}W_a^\mu . \qquad (7.57)$$

We can now calculate the correct transformation behavior of the spin 1 fields $W_a'^\mu$ by "solving" this equation for $W_a'^\mu$. To achieve this we multiply Eq. 7.57 from the right with $U^{-1}(x)$, which yields

$$(\partial^\mu U(x))U^{-1}(x) - ig\frac{\sigma^a}{2}W_a'^\mu \underbrace{U(x)U^{-1}(x)}_{=1} \overset{!}{=} -igU(x)\frac{\sigma^a}{2}W_a^\mu U(x)^{-1}$$

$$\rightarrow \frac{\sigma^a}{2}W_a'^\mu \overset{!}{=} U(x)\frac{\sigma^a}{2}W_a^\mu U^{-1}(x) - \underbrace{\frac{1}{ig}}_{=\frac{i}{i^2 g}=-\frac{i}{g}} (\partial^\mu U(x))U^{-1}(x)$$

$$\rightarrow \frac{\sigma^a}{2}W_a'^\mu \overset{!}{=} U(x)\frac{\sigma^a}{2}W_a^\mu U^{-1}(x) + \frac{i}{g}(\partial^\mu U(x))U^{-1}(x) \qquad (7.58)$$

This is how our gauge fields $W_a'^\mu$, that we introduced in Eq. 7.55 as part of the covariant derivative $D^\mu = \partial^\mu - ig\frac{\sigma^a}{2}W_a^\mu$, need to transform such that $(D^\mu\Psi)' = U(x)D^\mu\Psi$. We saw in Eq. 7.53 that this transformation behavior is needed to get a locally $SU(2)$ invariant Lagrangian.

There is one last thing, we need to take care of. We introduced three new spin 1 fields W_a^μ and we saw that they need to have very specific transformation properties when we want a locally $SU(2)$ invariant Lagrangian. However, the "naive" Lagrangian for these spin 1 fields (Eq. 6.25)

$$
\begin{aligned}
\mathscr{L}_{\text{3xMaxwell}} &= \frac{1}{4}(W_{\mu\nu})_1(W^{\mu\nu})_1 + \frac{1}{4}(W_{\mu\nu})_2(W^{\mu\nu})_2 + \frac{1}{4}(W_{\mu\nu})_3(W^{\mu\nu})_3 \\
&= \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i
\end{aligned} \tag{7.59}
$$

with

$$
(W_{\mu\nu})_i = \partial_\mu(W_\nu)_i - \partial_\nu(W_\mu)_i
$$

[33] This can be seen by using the explicit transformation behavior that we derived in Eq. 7.58.

is not invariant under such transformations[33].

We saw above that the demand for local $U(1)$ symmetry was powerful and yielded the correct Lagrangian that describes electromagnetic interactions. Thus instead of discarding the transformation behavior that successfully makes the spin $\frac{1}{2}$ part of of the Lagrangian invariant under local $SU(2)$ transformations, because it is not a symmetry of the free spin 1 field Lagrangian, we try to find a better Lagrangian for these spin 1 fields which has the desired symmetry. In other words, our final task is to derive a Lagrangian that describes how these new spin 1 fields behave when they are on their own that is invariant under the transformation in Eq. 7.58.

To find this Lagrangian, we need to note several things:

1. The new spin 1 fields $(W_\mu)_i$ always appeared in the previous Lagrangians[34] in combination with the generators $\frac{\sigma^i}{2}$. It is thus useful to introduce a new object $\mathcal{W}_\mu \equiv (W_\mu)_i \frac{\sigma^i}{2}$. Next, we might try to use this new object $\mathcal{W}_\mu$ instead of $(W_\mu)_i$ in the Lagrangian Eq. 7.59[35]. The important difference is that $\mathcal{W}_\mu$ is a matrix, because the generators are matrices. Therefore, when we define our field strength tensor in terms of the new object $\mathcal{W}_\mu$:

[34] See, e.g. Eq. 7.58.

[35] It will become clear in a moment, why this is useful, although at a first glance it seems to make things much more complicated.

$$
\mathcal{W}_{\mu\nu} = \partial_\mu \mathcal{W}_\nu - \partial_\nu \mathcal{W}_\mu , \tag{7.60}
$$

it becomes a matrix, too.

2. We want a Lagrangian that is invariant under local $SU(2)$ transformations, and therefore need to combine our fields in such a way that their transformation behavior cancels exactly. Now that we have new objects $(\mathcal{W}_\mu, \mathcal{W}_{\mu\nu})$ that are matrices, we can try to make use of the following nice property of the trace of matrices[36]

[36] This property is known as "cyclic property" of the trace. In words it means that the trace stays the same when we perform cyclic permutations among the matrices A, B, C, D that appear here. We always take the last element that appears in the product and put it at the beginning. However, take note that arbitrary permutations do not lead to the same trace. For example, $\text{Tr}(ACBD) \neq \text{Tr}(ABCD)$.

$$
\text{Tr}(ABCD) = \text{Tr}(DABC) = \text{Tr}(CDAB) = \text{Tr}(BCDA) . \tag{7.61}
$$

If we somehow manage that our field strength tensor transforms as $\mathcal{W}_{\mu\nu} \rightarrow U(x)\mathcal{W}_{\mu\nu}U^{-1}(x)$, the term $\text{Tr}(\mathcal{W}_{\mu\nu}\mathcal{W}_{\mu\nu})$ would be invariant, because[37]

$$\text{Tr}(\mathcal{W}_{\mu\nu}\mathcal{W}_{\mu\nu}) \rightarrow \text{Tr}\left(U(x)\mathcal{W}_{\mu\nu}U^{-1}(x)U(x)\mathcal{W}_{\mu\nu}U^{-1}(x)\right)$$
$$\underset{\text{Eq. 7.61}}{=} \text{Tr}\left(\underbrace{U^{-1}(x)U(x)}_{=1}\mathcal{W}_{\mu\nu}\underbrace{U^{-1}(x)U(x)}_{=1}\mathcal{W}_{\mu\nu}\right)$$
$$= \text{Tr}(\mathcal{W}_{\mu\nu}\mathcal{W}_{\mu\nu}) \quad \checkmark . \tag{7.62}$$

3. Unfortunately, the naive field strength tensor that we wrote down in Eq. 7.60 does not transform so nicely. However, we can construct a different field strength tensor that has exactly the needed transformation behavior. Above, we derived the covariant derivative D^μ that transforms exactly in the way we need it to get a locally $SU(2)$ invariant Lagrangian for the spin $\frac{1}{2}$ fields: $(D^\mu\Psi)' = U(x)D^\mu\Psi$. The spin $\frac{1}{2}$ doublet Ψ transforms as[38] $\Psi \rightarrow U(x)\Psi$. Therefore, we can conclude that the covariant derivative transforms as $D^\mu \rightarrow U(x)D^\mu U^{-1}(x)$, because

$$(D^\mu\Psi)' = D'^\mu\Psi'$$
$$= U(x)D^\mu\underbrace{U^{-1}(x)U(x)}_{=1}\Psi$$
$$= U(x)D^\mu\Psi \quad \checkmark \tag{7.63}$$

Now the final crucial trick to get a locally $SU(2)$ invariant Lagrangian for the spin 1 fields is the observation that the object[39]

$$\mathcal{W}_{\mu\nu} \equiv \frac{i}{g}[D^\mu, D^\nu] = \frac{i}{g}\left(D^\mu D^\nu - D^\nu D^\mu\right) \tag{7.64}$$

has exactly the correct transformation behavior[40] $\mathcal{W}_{\mu\nu} \rightarrow U(x)\mathcal{W}_{\mu\nu}U^{-1}(x)$:

$$\frac{g}{i}\mathcal{W}_{\mu\nu} = D^\mu D^\nu - D^\nu D^\mu$$
$$\rightarrow U(x)D^\mu\underbrace{U^{-1}(x)U(x)}_{=1}D^\nu U^{-1} - U(x)D^\nu\underbrace{U^{-1}(x)U(x)}_{=1}D^\mu U^{-1}$$
$$= U(x)D^\mu D^\nu U^{-1} - U(x)D^\nu D^\mu U^{-1}$$
$$= U(x)\left(D^\mu D^\nu - D^\nu D^\mu\right)U^{-1}$$
$$= U(x)\mathcal{W}_{\mu\nu}U^{-1} \quad \checkmark \tag{7.65}$$

4. When we now calculate explicitly how our field strength tensor looks like in terms of the fields $\mathcal{W}_\mu$, we can see what we missed in our naive definition in Eq. 7.60[41]

$$
\begin{aligned}
\mathcal{W}_{\mu\nu} f(x) &= \frac{i}{g}\left(D^{\mu}D^{\nu} - D^{\nu}D^{\mu}\right)f(x) \\
&= \left(\frac{i}{g}(\partial^{\mu} - ig\mathcal{W}^{\mu})(\partial^{\nu} - ig\mathcal{W}^{\nu}) - \frac{i}{g}(\partial^{\nu} - ig\mathcal{W}^{\nu})(\partial^{\mu} - ig\mathcal{W}^{\mu})\right)f(x) \\
&= \left(\frac{i}{g}\cancel{\partial^{\mu}\partial^{\nu}} + \partial^{\mu}\mathcal{W}^{\nu} + \mathcal{W}^{\mu}\partial^{\nu} - ig\mathcal{W}^{\mu}\mathcal{W}^{\nu} \right. \\
&\quad \left. - (\frac{i}{g}\cancel{\partial^{\nu}\partial^{\mu}} + \partial^{\nu}\mathcal{W}^{\mu} + \mathcal{W}^{\nu}\partial^{\mu} - ig\mathcal{W}^{\nu}\mathcal{W}^{\mu})\right)f(x) \\
&\underbrace{=}_{\text{product rule}} \left(\partial^{\mu}\mathcal{W}^{\nu} + \cancel{\mathcal{W}^{\nu}\partial^{\mu}} + \cancel{\mathcal{W}^{\mu}\partial^{\nu}} - ig\mathcal{W}^{\mu}\mathcal{W}^{\nu}\right. \\
&\quad \left. - (\partial^{\nu}\mathcal{W}^{\mu} + \cancel{\mathcal{W}^{\mu}\partial^{\nu}} + \cancel{\mathcal{W}^{\nu}\partial^{\mu}} - ig\mathcal{W}^{\nu}\mathcal{W}^{\mu})\right)f(x) \\
&= \left(\partial^{\mu}\mathcal{W}^{\nu} - \partial^{\nu}\mathcal{W}^{\mu} - ig\mathcal{W}^{\mu}\mathcal{W}^{\nu} + ig\mathcal{W}^{\nu}\mathcal{W}^{\mu})\right)f(x) \\
&= \left(\partial^{\mu}\mathcal{W}^{\nu} - \partial^{\nu}\mathcal{W}^{\mu} - ig[\mathcal{W}^{\mu}, \mathcal{W}^{\nu}]\right)f(x)\,. \tag{7.66}
\end{aligned}
$$

The last term here is what we missed previously[42]. It is important for local $SU(2)$ symmetry, because the $\mathcal{W}^{\mu}$ are now matrices and therefore $\mathcal{W}^{\mu}\mathcal{W}^{\nu} \neq \mathcal{W}^{\nu}\mathcal{W}^{\mu}$. In contrast, in order to ensure local $U(1)$ symmetry there was no need for such a term. There was just one spin 1 field, hence no matrix and therefore this last term vanishes.

This was quite a long journey, but now we have everything we need to write down a Lagrangian that is invariant under local $SU(2)$ transformations:

$$
\mathscr{L}_{\text{locally SU(2) invariant}} = i\bar{\Psi}\gamma_{\mu}D^{\mu}\Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right), \tag{7.67}
$$

where

$$
\begin{aligned}
D^{\mu} &\equiv \partial^{\mu} - ig\mathcal{W}^{\mu} \\
\mathcal{W}^{\mu\nu} &\equiv \partial^{\mu}\mathcal{W}^{\nu} - \partial^{\nu}\mathcal{W}^{\mu} - ig[\mathcal{W}^{\mu}, \mathcal{W}^{\nu}] \\
\mathcal{W}^{\mu} &\equiv (W^{\mu})_{i}\frac{\sigma^{i}}{2}\,. \tag{7.68}
\end{aligned}
$$

7.3   Mass Terms and "Unification" of $SU(2)$ and $U(1)$

In the last section we couldn't add mass terms like $m_1\bar{\Psi}\Psi$ and $m_2(W^{\mu})_i(W_{\mu})_i$ to the Lagrangian without destroying the $SU(2)$ symmetry. From experiments we know that the corresponding particles[43] have mass and this is conventionally interpreted as the $SU(2)$ symmetry being

broken. This means the symmetry exists at high energy and sponta-
neously breaks at lower energies.

So far we derived a locally $U(1)$ invariant Lagrangian and in this
context it's conventional to name the corresponding spin 1 field B_μ:

$$\mathscr{L}_{\text{locally U(1) invariant}} = -m\bar{\psi}\psi + i\bar{\psi}\gamma_\mu(\partial^\mu - igB_\mu)\psi - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \quad (7.69)$$

with

$$B^{\mu\nu} := \partial^\mu B^\nu - \partial^\nu B^\mu$$

The spin 1 field B^μ is often called $U(1)$ **gauge field**, because it makes
the Lagrangian $U(1)$ invariant.

The locally $SU(2)$ invariant Lagrangian is (Eq. 7.67)

$$\mathscr{L}_{\text{locally SU(2) invariant}} = i\bar{\Psi}\gamma_\mu D^\mu\Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right). \quad (7.70)$$

As above, the three spin 1 fields $(W_\nu)_i$ are called the $SU(2)$ gauge
fields, because they make the Lagrangian locally $SU(2)$ invariant.

We can combine them into one locally $U(1)$ **and** locally $SU(2)$
invariant Lagrangian[44]

$$\mathscr{L}_{\text{SU(2) and U(1)}} = i\bar{\Psi}\gamma_\mu(\partial^\mu - igB^\mu - ig'\mathcal{W}^\mu)\Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right) - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.$$
$$(7.71)$$

Now, how can we add mass terms to this Lagrangian without
spoiling the $SU(2)$ symmetry? The only ingredient we haven't used
so far is a spin 0 field; so let's see. The globally $U(1)$ invariant La-
grangian for a complex spin 0 field is given by (Eq. 7.40)

$$\mathscr{L}_{\text{spin0}} = \frac{1}{2}\left(\partial_\mu\phi^\dagger\partial^\mu\phi - m^2\phi^\dagger\phi\right). \quad (7.72)$$

We can add to this Lagrangian the next higher power in ϕ without
violating any symmetry constraints[45]. Thus we write, renaming the
constants to their conventional names

$$\mathscr{L}_{\text{spin0+extraTerm}} = \partial_\mu\phi^\dagger\partial^\mu\phi + \rho^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2. \quad (7.73)$$

We already know from Eq. 7.41 how we can add a coupling term
between this spin 0 field and a $U(1)$ gauge field B_μ, which makes the
Lagrangian locally $U(1)$ invariant:

$$\mathscr{L}_{\text{spin0+extraTerm+spin1Coupling}} = \left((\partial_\mu + igB_\mu)\phi^\dagger\right)\left((\partial^\mu - igB^\mu)\phi\right)$$
$$+ \rho^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \quad (7.74)$$

with the symmetries[46]

[44] We use here again the notation $\mathcal{W}^\mu = \frac{\sigma_j}{2}W_j^\mu$.

[45] Recall that only higher order deriva-
tives were really forbidden in order to
get a sensible theory. Higher powers
of ϕ describe the self-interaction of the
field ϕ and were omitted in order to get
a "free" theory.

[46] See Eq. 7.13 and recall that an overall
constant in the Lagrangian has no
influence.

$$B_\mu \rightarrow B'_\mu = B_\mu + \partial_\mu a(x) \tag{7.75}$$

$$\phi(x) \rightarrow \phi'(x) = e^{ia(x)} \phi(x). \tag{7.76}$$

[47] $\Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$

In the same way how we derived in the last chapter the locally $SU(2)$ invariant Lagrangian for spin $\frac{1}{2}$ fields, we can write a locally $SU(2)$ invariant Lagrangian for doublets of spin 0 fields[47] as

$$\mathscr{L}_{\text{SU(2) and U(1)}} = \left((\partial_\mu + ig'\mathcal{W}^\mu + igB_\mu)\Phi^\dagger \right) \left((\partial^\mu - ig'\mathcal{W}^\mu - igB^\mu)\Phi \right)$$
$$+ \underbrace{\rho^2\Phi^\dagger\Phi - \lambda(\Phi^\dagger\Phi)^2}_{\equiv -V(\Phi)} \tag{7.77}$$

[48] We use again the abbreviation $U(x) = e^{ib_i(x)\frac{\sigma_i}{2}}$ for our local $SU(2)$ transformations.

with the doublet $\Phi := \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ and the symmetries[48] (Eq. 7.58)

$$\mathcal{W}^\mu \rightarrow \mathcal{W}^\mu = U(x)\mathcal{W}^\mu U^{-1}(x) + \frac{i}{g}(\partial^\mu U(x))U^{-1}(x) \tag{7.78}$$

and

$$\Phi \rightarrow \Phi' = U(x)\Phi, \quad \overline{\Phi} \rightarrow \overline{\Phi}' = \overline{\Phi}U^\dagger(x). \tag{7.79}$$

We start with this locally $SU(2)$ invariant Lagrangian and investigate in the following how this Lagrangian gives us mass terms for the fields $\mathcal{W}^\mu = W_i^\mu \frac{\sigma_i}{2}$ and B^μ. Adding mass terms "by hand" to the Lagrangian does not work, because these terms spoil the symmetry and this leads to an insensible theory[49].

[49] The reason is quite complicated and will not be discussed in this book. In technical terms: We need a locally $SU(2)$ symmetric Lagrangian to get a renormalizable theory. You are encouraged to read about this in the books mentioned at the end of this chapter.

The term we defined above

$$V(\Phi) = -\rho^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2$$
$$= -\rho^2\phi_1^\dagger\phi_1 + \lambda(\phi_1^\dagger\phi_1)^2 - \rho^2\phi_2^\dagger\phi_2 + \lambda(\phi_2^\dagger\phi_2)^2$$
$$= V_1(\phi_1) + V_2(\phi_2) \tag{7.80}$$

is often called Higgs potential. A two-dimensional plot for different values of ρ, with $\lambda > 0$ can be seen in Fig. 7.1.
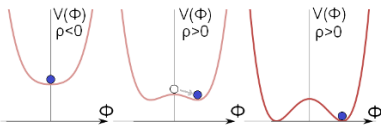
Fig. 7.1: Two-dimensional illustration of the Higgs potential for different values of ρ, which is believed to have changed as the universe cooled down as a result of the expansion of the universe. Figure adapted from "Spontaneous symmetry breaking" by FT2 (Wikimedia Commons) released under a CC BY-SA 3.0 licence: http://creativecommons.org/licenses/by-sa/3.0/deed.en. URL: http://commons.wikimedia.org/wiki/File:Spontaneous_symmetry_breaking_(explanatory_diagram).png, Accessed: 8.12.2014

The idea is that at very high temperatures, e.g. in the early universe, the potential looks like in the image to the left. The minimum, in this context called the **vacuum expectation value**, is without ambiguity at $\phi = 0$. With sinking temperature the parameters λ and ρ change, and with them the shape of the potential. After the temperature dropped below some critical value the potential no longer has its minimum at $\phi = 0$, as indicated in the pictures to the right. Now there is not only one location with the minimum value, but many.

In fact, the potential has an infinite number of possible minima. The minima of the potential can be computed in the usual way

$$V(\phi) = -\rho^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)(\phi^\dagger\phi) \tag{7.81}$$

$$\frac{\partial V(\phi)}{\partial \phi} = -2\rho^2 \phi^\dagger + 2\lambda |\phi|^2 \phi^\dagger \overset{!}{=} 0 \tag{7.82}$$

$$\rightarrow \phi^\dagger (-2\rho^2 + 4\lambda |\phi|^2) \overset{!}{=} 0 \tag{7.83}$$

$$\rightarrow |\phi|^2 \overset{!}{=} \frac{\rho^2}{2\lambda} \tag{7.84}$$

$$\rightarrow |\phi| \overset{!}{=} \sqrt{\frac{\rho^2}{2\lambda}} \tag{7.85}$$

$$\phi_{\min} = \sqrt{\frac{\rho^2}{2\lambda}} e^{i\varphi}. \tag{7.86}$$

This is a minimum for every value of φ and we therefore have an infinite number of minima. All these minima lie on a circle with radius $\sqrt{\frac{\rho^2}{2\lambda}}$. This can be seen in the three dimensional plot of the Higgs potential in Fig. 7.2. Like a marble that rolls down from the top of a sombrero, spontaneously one new vacuum value is chosen out of the infinite possibilities.

From Eq. 7.80 we see that for the doublet, both components have this choice to make. We therefore have for the doublet the minimum

$$\Phi_{\min} = \begin{pmatrix} \phi_{1\min} \\ \phi_{2\min} \end{pmatrix}. \tag{7.87}$$

An economical **choice**[50] for the minimum is

$$\Phi_{\min} = \begin{pmatrix} 0 \\ \sqrt{\frac{\rho^2}{2\lambda}} \end{pmatrix} \equiv \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \tag{7.88}$$

where the factor $\frac{1}{2}$ is just a convention to make computations easier and we define for brevity $v \equiv \sqrt{\frac{\rho^2}{\lambda}}$. We will learn later that in quantum field theory computations are always done as a series expansion around the minimum, because no exact solutions are available. In order to get sensible results, we must therefore shift the field Φ to the new minimum. We therefore consider the field

$$\Phi = \begin{pmatrix} \phi_{1r} + i\phi_{1c} \\ \frac{v}{\sqrt{2}} + \phi_{2r} + i\phi_{2c} \end{pmatrix}. \tag{7.89}$$

This can be rewritten as[51]

$$\Phi = e^{i\theta_i \frac{\sigma_i}{2}} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}, \tag{7.90}$$
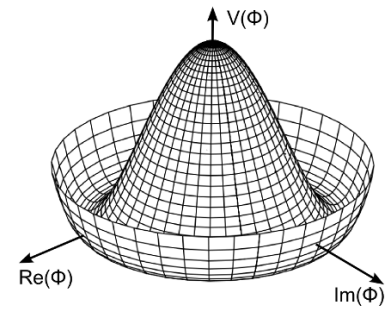
Fig. 7.2: 3-dimensional plot of the Higgs potential. Figure adapted from "Mexican hat potential polar" by Rupert Millard (Wikimedia Commons) released under a public domain licence. URL: `http://commons.wikimedia.org/wiki/File:Mexican_hat_potential_polar.svg` , Accessed: 7.5.2014

[50] Recall that symmetry breaking means that **one** minimum is chosen out of the infinite possibilities.

[51] This form is very useful as we will see in a moment.

because if we consider the series expansion of the exponential function and the explicit form of the Pauli matrices σ_i (Eq. 3.80), we can see that in first order

$$
\begin{aligned}
e^{i\theta_i \frac{\sigma_i}{2}} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} &\approx (1 + i\frac{1}{2}\theta_i\sigma_i) \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \\
&= (1 + i\frac{1}{2}\theta_1\sigma_1 + i\frac{1}{2}\theta_2\sigma_2 + i\frac{1}{2}\theta_3\sigma_3) \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \\
&= \begin{pmatrix} 1 + i\frac{1}{2}\theta_3 & i\frac{1}{2}\theta_1 + \frac{1}{2}\theta_2 \\ i\frac{1}{2}\theta_1 - \frac{1}{2}\theta_2 & 1 - i\frac{1}{2}\theta_3 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \\
&= \begin{pmatrix} (i\frac{1}{2}\theta_1 + \frac{1}{2}\theta_2)\frac{v+h}{\sqrt{2}} \\ (1 - i\frac{1}{2}\theta_3)\frac{v+h}{\sqrt{2}} \end{pmatrix} \\
\text{redefinitions} \rightarrow &\equiv \begin{pmatrix} \phi_{1r} + i\phi_{1c} \\ \frac{v}{\sqrt{2}} + \phi_{2r} + i\phi_{2c} \end{pmatrix}.
\end{aligned}
$$

$$(7.91)$$

Writing the complex spin 0 doublet in this form is useful, because we can now use the local $SU(2)$ (gauge) symmetry to make computations simpler. In order to get physical results **one** gauge must be chosen and we prefer to work with a gauge that makes life the easiest[52].

A general local $SU(2)$ transformation is

$$
\Phi \rightarrow \Phi' = e^{ib_i(x)\frac{\sigma_i}{2}}\Phi, \tag{7.92}
$$

which enables us to eliminate the exponential factor in Eq. 7.90, by choosing appropriate $b_i(x)$. The complex scalar doublet is then, in this **unitary gauge**

$$
\Phi_{un} = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}. \tag{7.93}
$$

Another possible way to understand this is that of the original four components that appeared in our complex scalar doublet, three are equivalent to our $SU(2)$ gauge freedom[53]. Therefore, these three fields aren't physical[54] and can't be measured in experiments. What remains is **one** physical field h, which is called the **Higgs field**.

Next, we want to take a look at the implications of this symmetry breaking on the Lagrangian. We recite here the Lagrangian in question for convenience, which was derived in Eq. 7.74 and include an additional factor $\frac{1}{2}$ in front of the field B_μ to unclutter the notation in the calculations that follow[55]

[52] For different computations, different gauges can be useful. Here we will work with what is called the unitary gauge, that is particularly useful to understand the physical particle content of a theory.

[53] Take note that this is only possible, because we have a local $SU(2)$ theory, because our fields $\theta = \theta(x)$, of course depend on the location in spacetime. For a global symmetry, these components can't be gauged away, and are commonly interpreted as massless bosons, called Goldstone bosons.

[54] The local $SU(2)$ symmetry is nothing that can be measured in experiments. This is merely a symmetry of our equations and the gauge freedom disappears from everything that is measurable in experiments. Otherwise there would be no possible way to make predictions from our theory, because we would have an infinite number of equivalently possible predictions (that are connected by $SU(2)$ transformations). Nevertheless, this symmetry is far from being useless, because it guides us to the correct form of the Lagrangian.

[55] Such a factor changes nothing, because it corresponds simply to a redefinition of the coupling constant $\frac{1}{2}\tilde{g} \equiv g$. Thus, strictly speaking, we should use $\tilde{g}$ here instead of g. However, the name of the constant doesn't matter and we continue to call the coupling constant simply g.

$$\mathscr{L} = \left((\partial_\mu + ig'\frac{\sigma_i}{2}(W_\mu)_i + i\frac{1}{2}gB_\mu)\Phi^\dagger \right) \left((\partial^\mu - ig'\frac{\sigma_i}{2}(W^\mu)_i - i\frac{1}{2}gB^\mu)\Phi \right)$$
$$- V(\Phi). \tag{7.94}$$

We now substitute the field Φ with the shifted field in the unitary gauge, which was defined in Eq. 7.93. Of particular interest for us will be the newly appearing terms that include the constant vacuum value v. The other terms describe the self-interaction of the Higgs-field and the interaction of the Higgs field with the other fields, which we will not examine any further. If we put in the minimum value $\Phi \to \Phi_{\text{min}} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}$, which means we ignore h, we get

$$\left((\partial_\mu + i\frac{\sigma_i}{2}g'(W_\mu)_i + i\frac{1}{2}gB_\mu)\Phi^\dagger_{\text{min}} \right) \left((\partial^\mu - ig'\frac{\sigma_i}{2}(W^\mu)_i - i\frac{1}{2}gB^\mu)\Phi_{\text{min}} \right)$$

$$= \left| \left((\partial^\mu + ig'\frac{\sigma_i}{2}(W^\mu)_i + i\frac{1}{2}gB^\mu)\Phi_{\text{min}} \right) \right|^2$$

$$= \left| \left((\partial^\mu + ig'\frac{\sigma_i}{2}(W^\mu)_i + i\frac{1}{2}gB^\mu)\sqrt{\frac{1}{2}}\begin{pmatrix} 0 \\ v \end{pmatrix} \right) \right|^2$$

$$= \frac{v^2}{8} \left| \left((g'\sigma_i(W^\mu)_i + gB^\mu)\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \right|^2.$$

Now using that we have behind B^μ an implicit 2×2 identity matrix and the explicit form of the Pauli matrices[56] σ_i yields

[56] $\sigma_i W_i = \begin{pmatrix} W_3 & W_1 - iW_2 \\ W_1 + iW_2 & -W_3 \end{pmatrix}$

$$= \frac{v^2}{8} \left| \begin{pmatrix} g'W_3^\mu + gB^\mu & g'W_1^\mu - ig'W_2^\mu \\ g'W_1^\mu + ig'W_2^\mu & -g'W_3^\mu + gB^\mu \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \right|^2$$

$$= \frac{v^2}{8} \left| \left(\begin{pmatrix} g'W_1^\mu - ig'W_2^\mu \\ -g'W_3^\mu + gB^\mu \end{pmatrix} \right) \right|^2$$

$$= \frac{v^2}{8} \left((g')^2 \left((W_1^\mu)^2 + (W_2^\mu)^2 \right) + (g'W_3^\mu - gB^\mu)^2 \right) \tag{7.95}$$

Next we define two new spin 1 fields from the old ones we have been using so far

$$W_+^\mu \equiv \frac{1}{\sqrt{2}}(W_1^\mu - iW_2^\mu) \tag{7.96}$$

$$W_-^\mu \equiv \frac{1}{\sqrt{2}}(W_1^\mu + iW_2^\mu), \tag{7.97}$$

where W_+^μ is the complex conjugate of W_-^μ. The first term in Eq. 7.95 is then

$$(W_1^\mu)^2 + (W_2^\mu)^2 = 2(W^+)_\mu(W^-)^\mu \tag{7.98}$$

and thus we have, including the constants,

$$\underbrace{\left(\frac{g'v}{2}\right)}_{\equiv m_W}^2 (W^+)_\mu (W^-)^\mu \tag{7.99}$$

which then looks like a typical "mass" term.

The second term in Eq. 7.95 can be written in matrix form

$$(g'W_3^\mu - gB^\mu)^2 = \left(W_3^\mu, B_\mu\right) \underbrace{\begin{pmatrix} g'^2 & -gg' \\ -gg' & g^2 \end{pmatrix}}_{\equiv G} \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix}. \tag{7.100}$$

In order to be able to interpret this as mass-terms, we need to diagonalize[57] the matrix G. The standard linear-algebra way to do this needs the eigenvalues λ_1, λ_2 and normalized[58] eigenvectors $\vec{v}_1, \vec{v}_2$ of the matrix G, which are

$$\lambda_1 = 0 \rightarrow \vec{v}_1 = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g \\ g' \end{pmatrix}$$

$$\lambda_2 = (g^2 + g'^2) \rightarrow \vec{v}_2 = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g' \\ -g \end{pmatrix}.$$

The matrix G is then diagonalized by the matrix M build from the eigenvectors as its columns, i.e. $G_{diag} = M^{-1}GM$, with

$$M = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \tag{7.101}$$

and

$$G_{diag} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & (g^2 + g'^2) \end{pmatrix}. \tag{7.102}$$

The matrix M is orthogonal ($M^T = M^{-1}$), because we work with normalized eigenvectors:

$$M^T M = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix}$$

$$= \frac{1}{(g^2 + g'^2)} \begin{pmatrix} g^2 + g'^2 & gg' - gg' \\ gg' - gg' & g^2 + g'^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7.103}$$

We therefore add two unit matrices $1 = M^T M$, into Eq. 7.100:

$$\left(W_3^\mu, B_\mu\right) \underbrace{MM^T}_{=1} G \underbrace{MM^T}_{=1} \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix} = \left(W_3^\mu, B_\mu\right) M \underbrace{M^T G M}_{= G_{diag}} M^T \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix}.$$

$$\tag{7.104}$$

[57] We will see in a moment that a diagonalized matrix gives us terms that look exactly like the other mass terms. This enables us to interpret the corresponding fields as physical fields that can be observed in experiments. We could work with fields W_3^μ and B_μ, but the physical interpretation would be much harder.

[58] Which means length 1, i.e. $\vec{v} \cdot \vec{v} = 1$.

The remaining task is then to evaluate $M^T \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix}$, in order to get the definition of two new fields, which have easily interpretable mass terms in the Lagrangian:

$$
M^T \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix} = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix}
$$

$$
= \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} (gW_3^\mu + g'B^\mu) \\ (g'W_3^\mu - gB^\mu) \end{pmatrix} \equiv \begin{pmatrix} A^\mu \\ Z_\mu \end{pmatrix}. \qquad (7.105)
$$

We can therefore write the second term as

$$
\begin{pmatrix} A^\mu & Z_\mu \end{pmatrix} G_{diag} \begin{pmatrix} A^\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} A^\mu & Z_\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & (g^2 + g'^2) \end{pmatrix} \begin{pmatrix} A^\mu \\ Z_\mu \end{pmatrix}
$$

$$
= (g^2 + g'^2)(Z^\mu)^2 + 0 \cdot (A^\mu)^2. \qquad (7.106)
$$

To summarize: We started with a Lagrangian, without mass terms for the spin 1 fields W_i^μ and B^μ (Eq. 7.71)

$$
\mathscr{L}_{\text{SU(2) and U(1)}} = i\bar{\Psi}\gamma_\mu(\partial^\mu - igB^\mu - ig'\mathcal{W}^\mu)\Psi - \frac{1}{4}Tr(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}) - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}.
$$
$$
(7.107)
$$

Then we included interactions with a doublet of spin 0 fields (Eq. 7.77) and after the process of spontaneous symmetry breaking, we have new terms in the Lagrangian that are interpreted as mass terms[59]

$$
\underbrace{\frac{1}{4}v^2 g'^2 (W^+)_\mu (W^-)^\mu}_{=M_W^2} + \underbrace{\frac{1}{8}v^2(g^2 + g'^2)\, Z_\mu^2}_{=\frac{1}{2}M_Z^2} + \underbrace{\frac{1}{8}v^2 0 \cdot A_\mu^2}_{\text{photon mass} = 0}. \qquad (7.108)
$$

We can see that one of the spin 1 fields A_μ remains massless after spontaneous symmetry breaking. This is the photon field of electromagnetism and all experiments up to now verify that the photon is massless [60]. An important observation is that the field responsible for Z-bosons Z_μ and the field responsible for photons A_μ are orthogonal linear combinations of the fields B_μ and W_μ^3. Therefore we can see that both have a common origin!

The same formalism can be used to get mass terms for spin $\frac{1}{2}$ fields without spoiling the local $SU(2)$ symmetry, but before we discuss this, we need to talk about one very curious fact of nature: parity violation.

[59] For the first term here, we combine Eq. 7.95 with Eq. 7.98.

[60] Take note that I omitted some very important notions in this section: Hypercharge and the Weinberg angle. The Weinberg angle θ_W is simply defined as $\cos(\theta_W) = \frac{g}{\sqrt{g^2+g'^2}}$ or $\sin(\theta_W) = \frac{g'}{\sqrt{g^2+g'^2}}$. This can be used to simplify some of the definitions mentioned in this section. Hypercharge is a bit more complicated to explain and those who want to dig deeper are referred to the standard texts about quantum field theory, some of which are recommended at the end of Chapter 9.

7.4   Parity Violation

One of the biggest discoveries in the history of science was that nature is not invariant under parity transformations. In layman's terms this means that some experiments behave differently than their mirrored analogue. The experiment that discovered the violation of parity symmetry was the Wu experiment. A full description of this experiment, although fascinating, strays from our current subject, so let's just discuss the final result.

The Wu experiment discovered that the particles mediating the weak force (the W^+, W^-, Z bosons) only couple to left-chiral particles. In other words: **only left-chiral particles interact via the weak force**[61]. All particles produced in weak interactions are left-chiral. Neutrinos interact exclusively via the weak force and therefore it is possible that there are no right-chiral neutrinos[62]. All other particles can be produced via other interactions and therefore can be right-chiral, too.

Up to this point, we used left-chiral and right-chiral as labels for objects transforming according to different representations of the Lorentz group. Although this seems like something very abstract, we can measure the chirality of particles, because there is a correlation to a more intuitive concept called helicity[63].

In fact, most of the time particles do not have a specified **chirality**, which means they aren't definitely left-chiral or right-chiral and the corresponding Dirac spinor Ψ has both components. Parity violation was **no** prediction of the theory and a total surprise for every physicist. Until the present day, no one knows why nature behaves so strangely. Nevertheless, it's easy to accommodate this discovery into our framework. We only need *something* that makes sure we always deal with left-chiral spinors when we describe weak interactions.

Recall that the symbols χ and ξ denote Weyl spinors (two component objects), ψ Dirac spinors (four component objects, consisting of two Weyl spinors)

$$\psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \tag{7.109}$$

and Ψ doublets of Dirac spinors

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \tag{7.110}$$

Then the "something" we need is the projection operator P_L:

$$P_L \psi = P_L \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} \equiv \psi_L. \tag{7.111}$$

[61] We will discuss at the end of this section why this means that parity is violated

[62] We will see in a moment that massive, left-chiral particles always get a right-chiral component during propagation. It is known from experiments that neutrinos have mass and therefore we would expect a right-chiral neutrino component. Nevertheless, this right-chiral component does not participate in any known interaction.

[63] We will not discuss this here, because the details make no difference for the purpose of the text. The message to take away is that it can be done. A very nice discussion of these matters can be found in Alessandro Bettini. *Introduction to Elementary Particle Physics*. Cambridge University Press, 2nd edition, 4 2014. ISBN 9781107050402

Such a projection operator can be constructed using the matrix[64]

$$\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad (7.112)$$

The matrix γ_5 is called the **chirality operator**, because states of pure chirality $\begin{pmatrix} \chi_L \\ 0 \end{pmatrix}$ or $\begin{pmatrix} 0 \\ \xi_R \end{pmatrix}$ are eigenstates of γ_5 with eigenvalue -1 and $+1$ respectively.

The projection operator P_L can then be written[65]

$$P_L = \frac{1 - \gamma_5}{2} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \qquad (7.113)$$

and we can define analogously

$$P_R = \frac{1 + \gamma_5}{2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \qquad (7.114)$$

Now, in order to accommodate for the fact that only left-chiral particles interact via the weak force, we must simply include P_L into all terms of the Lagrangian that describe the interaction of $W_\mu^\pm$ and Z_μ with different fields. The corresponding terms were derived in Section 7.2, and the final result was Eq. 7.67, which we recite here for convenience:

$$\mathscr{L}_{\text{locally SU(2) invariant}} = i\bar{\Psi}\gamma_\mu D^\mu \Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right)$$
$$= i\bar{\Psi}\gamma_\mu(\partial^\mu - ig\mathcal{W}^\mu)\Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right). \qquad (7.115)$$

The relevant term that describes the interaction is $g\bar{\Psi}\gamma_\mu\mathcal{W}^\mu\Psi$ and we simply add P_L:

$$\rightarrow \mathscr{L} = i\bar{\Psi}\gamma_\mu(\partial^\mu - ig\mathcal{W}^\mu P_L)\Psi - \frac{1}{4}Tr\left(\mathcal{W}_{\mu\nu}\mathcal{W}^{\mu\nu}\right). \qquad (7.116)$$

Here P_L acts on a doublet and is therefore

$$P_L\Psi = \begin{pmatrix} P_L & 0 \\ 0 & P_L \end{pmatrix}\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} (\psi_1)_L \\ (\psi_2)_L \end{pmatrix}. \qquad (7.117)$$

One P_L is enough to project the left-chiral component out of both doublets $\bar{\Psi}$ and Ψ. To see this we need three identities:

- $(P_L)^2 = P_L$, which is obvious from the explicit matrix form and because every projection operator must have this property[66]. Projecting twice must be the same as projecting one time.

---

[64] Recall the definition of the γ_μ matrices in Eq. 6.13 and don't let yourself get confused about the missing γ_4 matrix. There is an alternative convention that uses γ_4 instead of γ_0 and to avoid interference between those conventions, the matrix here is commonly called γ_5.

[65] Maybe you wonder why we define P_L as so complicated and do not start with the explicit matrix form right away. We do this, because it's possible to work in a different basis where the matrices γ_μ look completely different. (For more information about this have a look at Section 8.10). Take note that in the Lagrangian the Dirac spinors appear always in combination with the matrices γ_μ. We can always add a $1 = U^{-1}U$, with some arbitrary invertible matrix U, between them. For example, $\partial_\mu\bar{\Psi}\gamma_\mu\Psi = \partial_\mu\bar{\Psi}\underbrace{U^{-1}U}_{=1}\gamma_\mu\underbrace{U^{-1}U}_{=1}\Psi = \partial_\mu\underbrace{\bar{\Psi}U^{-1}}_{=\bar{\Psi}'}\underbrace{U\gamma_\mu U^{-1}}_{\gamma_\mu'}\underbrace{U\Psi}_{\Psi'}$. Physics is of course completely independent of such transformations, but we can use this to simplify computations. The basis we prefer to work with in this text is called Weyl Basis. In other bases the two components of a Dirac spinor are mixtures of χ_L and ξ_R. Nevertheless, the projection operator defined as $P_L = \frac{1-\gamma_5}{2}$, always projects out the left-chiral component, because $P_L^{\text{Weyl}}\Psi^{\text{Weyl}} = \Psi_L^{\text{Weyl}} \Rightarrow P_L'\Psi' = \frac{1-\gamma_5'}{2}\underbrace{\Psi'}_{U\Psi^{\text{Weyl}}} = \frac{1-Ui\gamma_0U^{-1}U\gamma_1U^{-1}U\gamma_2U^{-1}U\gamma_3U^{-1}}{2}U\Psi^{\text{Weyl}} = \frac{U-Ui\gamma_0\gamma_1\gamma_2\gamma_3}{2}\Psi^{\text{Weyl}} = U\left(\frac{1-\gamma_5}{2}\right)\Psi^{\text{Weyl}} = U\Psi_L^{\text{Weyl}} = \Psi_L'$ ✓

[66] Another defining condition of any projection operator is $P_L P_R = P_R P_L = 0$, which is here fulfilled as you can check by using the explicit form of P_L, P_R and γ_5.

- $\{\gamma_5, \gamma_\mu\} = \gamma_5 \gamma_\mu + \gamma_\mu \gamma_5 = 0$, which you can check by brute force computation[67].

[67] Or using another identity $\{\gamma_\mu, \gamma_\nu\} = \gamma_\mu \gamma_\nu + \gamma_\nu \gamma_\mu = \frac{1}{2}\eta_{\mu\nu}$, where $\eta_{\mu\nu}$ is the Minkowski metric and the definition $\gamma_5 = i\gamma_0 \gamma_1 \gamma_2 \gamma_3$.

- $(P_L)^\dagger = P_L$, because γ_5 is real, as can be seen from the explicit matrix form: $\gamma_5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

The second identity simply tells us that $\gamma_5 \gamma_\mu = -\gamma_\mu \gamma_5$, i.e. that we can switch the position of γ_5 and any γ_μ matrix, as long as we include a minus sign. This tells us

$$\gamma_\mu P_L = \gamma_\mu \frac{1-\gamma_5}{2} = \frac{1+\gamma_5}{2}\gamma_\mu = P_R \gamma_\mu. \tag{7.118}$$

We can now rewrite the relevant term of Eq. 7.116:

$$
\begin{aligned}
g\bar{\Psi}\gamma_\mu \mathcal{W}^\mu P_L \Psi &= g\bar{\Psi}\gamma_\mu \mathcal{W}^\mu \underbrace{P_L^2}_{=P_L} \Psi \\
&= g \underbrace{\bar{\Psi}}_{\Psi^\dagger \gamma_0} \underbrace{\gamma_\mu P_L}_{P_R \gamma_\mu} \mathcal{W}^\mu \underbrace{P_L \Psi}_{\Psi_L} \\
&= g\Psi^\dagger \underbrace{\gamma_0 P_R}_{P_L \gamma_0} \gamma_\mu \mathcal{W}^\mu \Psi_L \\
&\underbrace{=}_{} (P_L \Psi)^\dagger \gamma_0 \gamma_\mu \mathcal{W}^\mu \Psi_L \\
&\text{Using } P_L^\dagger = P_L \text{ and } (AB)^\dagger = ((AB)^T)^\star = (B^T A^T)^\star = B^\dagger A^\dagger \\
&= g\big(\underbrace{P_L \Psi}_{=\Psi_L}\big)^\dagger \gamma_0 \gamma_\mu \mathcal{W}^\mu \Psi_L \\
&= g\bar{\Psi}_L \gamma_\mu \mathcal{W}^\mu \Psi_L \quad \checkmark \tag{7.119}
\end{aligned}
$$

We can see here that one projection operator is really enough to guarantee that both Ψ and $\bar{\Psi}$ are left-chiral.

Now we know how we can describe mathematically that only left-chiral fields interact via the weak force, but why does this mean that parity is violated? To understand this we need to parity transform this term, because if it isn't invariant, the physical system in question is different from its mirror image[68]. Here we need the parity operator for spinors[69] P_spinor and vectors[70] P_vector. The transformation yields

[68] This means an experiment, whose outcome depends on this term of the Lagrangian, will find a different outcome if everything in the experiment is arranged mirrored.

[69] The parity operator for spinors was derived in Section 3.7.9. Using the γ_μ matrices, we can write the parity operator derived there as $P = \gamma_0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

[70] The parity operator for vectors is simply $P_\text{vector} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$ as already mentioned in Eq. 3.140.

$$
\begin{aligned}
\underbrace{\bar{\Psi}}_{=\Psi^\dagger \gamma_0} \gamma_\mu \mathcal{W}^\mu P_L \Psi &\to (P_\text{spinor}\Psi)^\dagger \gamma_0 \gamma_\mu (P_\text{vector}\mathcal{W}^\mu) P_L (P_\text{spinor}\Psi) \\
&= (\Psi)^\dagger \gamma_0 \gamma_0 \gamma_\mu (P_\text{vector}\mathcal{W}^\mu) P_L \gamma_0 \Psi \\
&\underbrace{=}_{} (\Psi)^\dagger \gamma_0 \gamma_0 \gamma_\mu \gamma_0 (P_\text{vector}\mathcal{W}^\mu) P_R \Psi. \tag{7.120}
\end{aligned}
$$
$$\text{using } \{\gamma_5, \gamma_0\} = 0 \text{ and } P_L = \tfrac{1-\gamma_5}{2}$$

Then we can use $\gamma_0\gamma_0\gamma_0 = \gamma_0$ and $\gamma_0\gamma_i\gamma_0 = -\gamma_i$, as you can check by looking at the explicit form of the matrices. Furthermore, we have $P_{\text{vector}}W^0 = W^0$ and $P_{\text{vector}}W^i = -W^i$, which follows from the explicit form of P_{vector}. We conclude these two minus signs cancel each other and the parity transformed term of the Lagrangian reads:

$$(P_{\text{spinor}}\Psi)^\dagger \gamma_0 \gamma_\mu (P_{\text{vector}}\mathcal{W}^\mu) P_L (P_{\text{spinor}}\Psi) = \bar{\Psi}\gamma_\mu \mathcal{W}^\mu P_R \Psi \neq \bar{\Psi}\gamma_\mu \mathcal{W}^\mu P_L \Psi$$
$$(7.121)$$

Therefore this term is *not* invariant and **parity is violated**.

Parity violation has another important implication. Recall that we always write things below each other between two big brackets if they can transform into each other[71]. For example, we use four-vectors, because their components can transform into each other through rotations or boosts. In this section we learned that only left-chiral particles interact via the weak force and the correct term in the Lagrangian is $\bar{\Psi}\gamma_\mu \mathcal{W}^\mu P_L \Psi = \bar{\Psi}_L \gamma_\mu \mathcal{W}^\mu \Psi_L$. In physical terms this term means that the components of the left-chiral doublets, which means the two spin $\frac{1}{2}$ fields $(\psi_1)_L$ and $(\psi_2)_L$ can transform into each other through weak-interactions. Right-chiral fields do not interact via the weak force and therefore $(\psi_1)_R$, $(\psi_2)_R$ aren't transformed into each other. Therefore writing them below each other between two big brackets makes no sense. In mathematical terms this means that right-chiral fields form $SU(2)$ **singlets**, i.e. are objects transforming according to the 1 dimensional representation of $SU(2)$, which do not change at all[72]. So let's summarize:

- Left-chiral fields are written as $SU(2)$ **doublets**: $\Psi_L = \begin{pmatrix} (\psi_1)_L \\ (\psi_2)_L \end{pmatrix}$, because they interact via the weak force and therefore can transform into each other. They transform under the two-dimensional representation of $SU(2)$:

$$\Psi_L \to \Psi_L' = e^{i\vec{a}\frac{\vec{\sigma}}{2}}\Psi_L.$$
$$(7.122)$$

- Right-chiral fields are described by $SU(2)$ **singlets**: $(\psi_1)_R$, $(\psi_2)_R$, because they do not interact via the weak force and therefore can't transform into each other. Therefore they transform under the one-dimensional representation of $SU(2)$:

$$(\psi_1)_R \to (\psi_1)_R' = e^0(\psi_1)_R = (\psi_1)_R$$
$$(\psi_2)_R \to (\psi_2)_R' = e^0(\psi_2)_R = (\psi_2)_R.$$
$$(7.123)$$

Now we move on and try to understand how mass terms for spin $\frac{1}{2}$ particles can be added in the Lagrangian without spoiling any crucial symmetry.

7.5   Lepton Mass Terms

At the beginning of Section 7.2, we discovered that we can't include arbitrary mass terms $\bar{\Psi}m\Psi$ without spoiling the $SU(2)$ symmetry. Now we will see that parity violation makes this problem even bigger. After discussing the problem, we will see that again the Higgs mechanism is a solution.

In the last section we talked a bit about the chirality of the coupling term: $\bar{\Psi}\gamma_\mu\sigma_j W_j^\mu P_L\Psi$. What about the chirality of a mass term? Take a look again at the invariants without derivatives for spinors, which we derived in Eq. 6.7 and Eq. 6.8:

[73] The Dirac spinors ψ_L and ψ_R are defined using the chiral-projection operators introduced in the last section: $\psi_L = P_L\psi$ and $\psi_R = P_R\psi$. And we have as always $\bar{\psi} = \psi^\dagger\gamma_0$.

$$I_1 := (\chi_{\hat{a}})^T\xi^{\hat{a}} = (\chi_L)^\dagger\xi_R \qquad \text{and} \quad I_2 := (\xi^a)^T\chi_a = (\xi_R)^\dagger\chi_L \quad (7.124)$$

We can write these invariants using Dirac spinors as[73]

$$\bar{\psi}\psi = \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L$$

$$= \begin{pmatrix} \chi_L^\dagger & 0 \end{pmatrix}\begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ \xi_R \end{pmatrix} + \begin{pmatrix} 0 & \xi_R^\dagger \end{pmatrix}\begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix}\begin{pmatrix} \chi_L \\ 0 \end{pmatrix}$$

$$= \chi_L^\dagger\xi_R + \xi_R^\dagger\chi_L \quad \checkmark \qquad\qquad\qquad (7.125)$$

We can see that Lorentz invariant mass terms always combine left-chiral with right-chiral fields. This is a problem, because left-chiral and right-chiral fields transform differently under $SU(2)$ transformations, as explained at the end of the last section. The left-chiral fields are doublets, whereas the right-chiral fields are singlets. The multiplication of a doublet and singlet is not $SU(2)$ invariant. For example

$$\underbrace{\bar{\Psi}_L}_{\text{doublet}}\underbrace{\psi_R}_{\text{singlet}} \rightarrow \bar{\Psi}'_L\psi'_R = \bar{\Psi}_L e^{-ib_i\frac{\sigma_i}{2}}\psi_R \neq \bar{\Psi}_L\psi_R \qquad (7.126)$$

From the experience with mass terms for spin 1 fields, we know what to do: Instead of considering terms as above, we add $SU(2)$ invariant coupling terms with 0 fields to the Lagrangian. Then, by choosing the vacuum value for the spin 0 field we break the symmetry and generate mass terms.

A $SU(2)$, $U(1)$ **and** Lorentz invariant term, coupling a spin 0 doublet and our spin $\frac{1}{2}$ fields together, is given by

$$\bar{\Psi}_L\Phi\psi_R. \qquad\qquad\qquad (7.127)$$

To see the invariance we transform this term with a $SU(2)$ trans-formation[74]

[74] Remember $\Phi_L \rightarrow \Phi'_L = e^{ib_i(x)\sigma_i}\Phi_L$ and $\sigma_i^\dagger = \sigma_i$

$$\bar{\Psi}_L \Phi \psi_R \rightarrow \bar{\Psi}'_L \Phi' \psi_R = \bar{\Psi}_L e^{-ib_i(x)\sigma_i} e^{ib_i(x)\sigma_i} \Phi \psi_R = \bar{\Psi}_L \Phi \psi_R \quad \checkmark$$

and equally for a $U(1)$ transformation:

$$\bar{\Psi}_L \Phi \psi_R \rightarrow \bar{\Psi}'_L \Phi \psi'_R = \bar{\Psi}_L e^{-ia(x)} \Phi e^{ia(x)} \psi_R = \bar{\Psi}_L \Phi \psi_R \quad \checkmark$$

The spin 0 field does not transform at all under Lorentz transformations[75] and therefore the term is Lorentz invariant, because we have the same Lorentz invariant terms as in Eq. 7.125.

This kind of term is called **Yukawa term** and we add it, with the equally allowed Hermitian conjugate to the Lagrangian[76] $-\lambda_2$

$$\mathscr{L} = -\lambda_2 (\bar{\Psi}_L \Phi \psi_{2R} + \bar{\psi}_{2R} \Phi \bar{\Psi}_L). \tag{7.128}$$

The coupling constant λ_2 is called a **Yukawa coupling**. This extra term does not only describe the interaction between the fermions and the Higgs field, but also leads to mass terms for the spin $\frac{1}{2}$ fields after the symmetry breaking. We put the expansion around the vacuum expectation value (Eq. 7.88)

$$\Phi = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix}$$

into the Lagrangian, which yields

$$\mathscr{L} = -\frac{\lambda_2}{\sqrt{2}} \left(\left(\bar{\Psi}_{1L}, \bar{\Psi}_{2L} \right) \begin{pmatrix} 0 \\ v+h \end{pmatrix} \psi_{2R} + \bar{\psi}_{2R} \left(0, v+h \right) \begin{pmatrix} \Psi_{1L} \\ \Psi_{2L} \end{pmatrix} \right)$$

$$= -\frac{\lambda_2(v+h)}{\sqrt{2}} \left(\bar{\Psi}_{2L} \psi_{2R} + \bar{\psi}_{2R} \Psi_{2L} \right). \tag{7.129}$$

Equation 7.125 tells us this is equivalent to

$$= -\frac{\lambda_2(v+h)}{\sqrt{2}} \bar{\psi}_2 \psi_2 \tag{7.130}$$

$$= \underbrace{-\frac{\lambda_2 v}{\sqrt{2}} (\bar{\psi}_2 \psi_2)}_{\text{Fermion mass term}} - \underbrace{\frac{\lambda_f h}{\sqrt{2}} (\bar{\psi}_2 \psi_2)}_{\text{Fermion-Higgs interaction}}. \tag{7.131}$$

We see that through the Higgs mechanism we get indeed the required mass terms. Again, we used symmetry constraints to add a term to the Lagrangian and this yields after spontaneous symmetry breaking mass terms for the spin $\frac{1}{2}$ fields. Take note that we only generated mass terms for the second field inside the doublet ψ_2. What about mass terms for the first field ψ_1?

To get mass terms for the first field ψ_1 we need to consider coupling terms to the charge-conjugated[77] Higgs field $\tilde{\Phi} = \epsilon \Phi^\star$, because

[75] By definition a spin 0 field transforms according to the $(0,0)$ representation of the Lorentz group. In this representation all Lorentz transformations are trivially the identity transformation. This was derived in Section 3.7.4.

[76] The strange name $-\lambda_2$ and why only add ψ_{2R} here will become clear in a moment, because terms including ψ_{1R} and $-\lambda_1$ will be discussed afterwards.

[77] Charge conjugation is explained in Section 3.7.10.

$$\Phi = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \rightarrow \tilde{\Phi} = \epsilon \Phi^{\star} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix}. \quad (7.132)$$

Following the same steps as above with the charge conjugated Higgs field leads to mass terms for ψ_1:

$$\mathscr{L} = -\lambda_f(\bar{\Psi}_L \tilde{\Phi} \psi_{1R} + \bar{\psi}_{1R} \tilde{\bar{\Phi}} \Psi_L)$$

$$= -\frac{\lambda_1}{\sqrt{2}} \left(\left(\bar{\Psi}_{1L}, \bar{\Psi}_{2L} \right) \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix} \Psi_{1R} + \bar{\Psi}_{1R} \begin{pmatrix} \frac{v+h}{\sqrt{2}} & 0 \end{pmatrix} \begin{pmatrix} \Psi_{1L} \\ \Psi_{2L} \end{pmatrix} \right)$$

$$= -\frac{\lambda_1(v+h)}{\sqrt{2}} \left(\bar{\Psi}_{1L} \Psi_{1R} + \bar{\Psi}_{1R} \Psi_{1L} \right). \quad (7.133)$$

To understand the rather abstract spin $\frac{1}{2}$ doublets better, we rewrite them more suggestively[78]

$$\Psi = \begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad (7.134)$$

and equivalently for the other leptons μ, ν_μ and τ, ν_τ. This form of the doublets is suggested by experiments, because an electron e is always transformed by weak interactions into another electron e, with possibly different momentum, or an electron-neutrino ν_e plus other particles. In weak interactions e and ν_e (equivalently μ and ν_μ or τ and ν_τ) always appear in pairs. This can be understood by looking at the coupling term $\bar{\Psi}\gamma_\mu \mathcal{W}^\mu P_L \Psi = \bar{\Psi}\gamma_\mu (W^\mu)_i \frac{\sigma^i}{2} P_L \Psi$. As discussed in the last section this can be rewritten using the explicit matrix form of the Pauli matrices[79] σ_i, which then gives us terms coupling the components of the doublets together:

$$\bar{\Psi}\gamma_\mu \sigma_j W_j^\mu P_L \Psi = \begin{pmatrix} \bar{\nu}_e & \bar{e} \end{pmatrix} \gamma_\mu \begin{pmatrix} W_3^\mu & \sqrt{2}W_+ \\ \sqrt{2}W_- & -W_3^\mu \end{pmatrix} P_L \begin{pmatrix} \nu_e \\ e \end{pmatrix}$$

$$\underbrace{=}_{\text{using Eq. 7.116}} \begin{pmatrix} (\bar{\nu}_e)_L & (\bar{e})_L \end{pmatrix} \gamma_\mu \begin{pmatrix} W_3^\mu & \sqrt{2}W_+ \\ \sqrt{2}W_- & -W_3^\mu \end{pmatrix} \begin{pmatrix} (\nu_e)_L \\ (e)_L \end{pmatrix}$$

$$= (\bar{\nu}_e)_L \gamma_\mu W_3^\mu (\nu_e)_L + (\bar{\nu}_e)_L \gamma_\mu \sqrt{2}W_+(e)_L$$

$$+ (\bar{e})_L \gamma_\mu \sqrt{2}W_-(\nu_e)_L - (\bar{e})_L \gamma_\mu W_3^\mu (e)_L. \quad (7.135)$$

If we want to consider all lepton generations at once, i.e. e, μ and τ, we need to write down three terms like this into the Lagrangian:

$$\bar{\Psi}_e \gamma_\mu \sigma_j W_j^\mu P_L \Psi_e + \bar{\Psi}_\mu \gamma_\mu \sigma_j W_j^\mu P_L \Psi_\mu + \bar{\Psi}_\tau \gamma_\mu \sigma_j W_j^\mu P_L \Psi_\tau, \quad (7.136)$$

which can be written more compactly by introducing $\Psi_l = \begin{pmatrix} \nu_l \\ l \end{pmatrix}$,

where $l = e, \mu, \tau$:

$$\bar{\Psi}_l \gamma_\mu \sigma_j W_j^\mu P_L \Psi_l.$$

Using the notation $l = \begin{pmatrix} l_L \\ l_R \end{pmatrix}$ the coupling term between the spin 0

and the electrically charged spin $\frac{1}{2}$ fields after spontaneous symmetry breaking reads

$$\underbrace{-\frac{\lambda_l v}{\sqrt{2}} (\bar{l}l)}_{\text{Fermion mass } m_L} \qquad \underbrace{- \frac{\lambda_l h}{\sqrt{2}} (\bar{l}l)}_{\text{Fermion-Higgs interaction strength } c_L}.$$

The terms for the corresponding neutrinos follow analogously[80].

This Lagrangian enables us to predict something about the Higgs field h that can be tested in experiments. For a given lepton l, the mass is given by

$$m_l = \frac{\lambda_l v}{\sqrt{2}} \rightarrow \lambda_l = \frac{m_l \sqrt{2}}{v} \tag{7.137}$$

and the coupling strength of this lepton to the Higgs is given by

$$c_l = \frac{\lambda_l h}{\sqrt{2}} \underbrace{=}_{\text{Eq. 7.137}} \frac{m_l h}{v}. \tag{7.138}$$

The last equation means that the coupling strength of the Higgs to a lepton is proportional to the mass of the lepton. The heavier the lepton the stronger the coupling. The same is true for all particles and the derivation is completely analogous.

There are other spin $\frac{1}{2}$ particles, called **quarks**, that interact via the weak force. In addition, quarks interact via a third force, called the strong force and this will be the topic of Section 7.8, but first we want to talk about mass terms for quarks. Luckily, these can be incorporated analogous to the lepton mass terms.

7.6   Quark Mass Terms

We learned in the last section that an $SU(2)$ doublet contains the particles that are transformed into each other via the weak force. For quarks[81] these are the up- and down quark:

$$q = \begin{pmatrix} u \\ d \end{pmatrix} \tag{7.139}$$

[80] As already mentioned above, we can only write down such mass terms if there are right-chiral neutrinos. However, so far right-chiral neutrinos were never observed in experiments, although we know that neutrinos have mass. This is one of the open problems of the standard model.

[81] If you've never heard of quarks before, have a look at Section 1.3.

and equally for the strange and charm or top and bottom quarks.

Again, we must incorporate the experimental fact that only left-chiral particles interact via the weak force. Therefore, we have left-chiral doublets and right-chiral singlets:

$$\underbrace{q_L}_{\text{doublet}} = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \qquad \rightarrow \qquad e^{ia_i\frac{\sigma_i}{2}}q_L \qquad (7.140)$$

$$\underbrace{u_R}_{\text{singlet}} \qquad \rightarrow \qquad u_R$$

$$\underbrace{d_R}_{\text{singlet}} \qquad \rightarrow \qquad d_R. \qquad (7.141)$$

Again, right-chiral particles do not interact via the weak force and therefore they aren't transformed into anything and form a $SU(2)$ singlet (=one component object).

The problem is the same as for leptons: To get something Lorentz invariant, we need to combine left-chiral with right-chiral spinors. Such a combination is not $SU(2)$ invariant and we use again the Higgs mechanism. This means, instead of terms like

$$\bar{q}_L u_R + \bar{q}_L d_R + \bar{u}_R q_L + \bar{d}_R q_L, \qquad (7.142)$$

which aren't $SU(2)$ invariant, we consider the coupling of the quarks to a spin 0 field doublet Φ:

$$\lambda_u \bar{q}_L \tilde{\Phi} u_R + \lambda_d \bar{q}_L \Phi d_R + \lambda_u \bar{u}_R \tilde{\Phi} q_L + \lambda_d \bar{d}_R \Phi q_L, \qquad (7.143)$$

with coupling constants λ_u, λ_d and the charge conjugated Higgs doublet[82] which is needed in order to get mass terms for the up quarks[83].

Then, everything is analogous to the lepton case: We put the expansion of the Higgs field around its minimum[84] into the Lagrangian, which gives us mass terms plus quark-Higgs coupling terms.

7.7 Isospin

Now it's time we talk about the conserved quantity that follows from $SU(2)$ symmetry. The free Lagrangians are only globally invariant and we need interaction terms to make them locally symmetric. Recall that global symmetry is a special case of local symmetry. Therefore we have global symmetry in every locally invariant Lagrangian and the corresponding conserved quantity is conserved for

[82] This is defined in Eq. 7.132:
$$\tilde{\Phi} = \epsilon \Phi^\star \underbrace{=}_{} \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix}.$$

[83] We defined the doublets as $\begin{pmatrix} u \\ d \end{pmatrix}$. Multiplication of this doublet with $\Phi = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}$ always results in terms proportional to d.

[84] $\Phi = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}$ and equivalently for the charge-conjugated Higgs field.

both cases. The result will be that global $SU(2)$ invariance gives us through Noether's theorem, a new conserved quantity called **isospin**. This is similar to electric charge, which is the conserved quantity that follows from global $U(1)$ invariance.

Noether's theorem for internal symmetry (Section 4.5.5, especially Eq. 4.57) tells us that

$$\partial_0 \int d^3x \underbrace{\frac{\partial \mathscr{L}}{\partial(\partial_0 \Psi)} \delta \Psi}_{=Q} = 0 \qquad (7.144)$$

The Lagrangian is invariant under transformations of the form

$$\Psi \to e^{ia_i \frac{\sigma_i}{2}} \Psi = (1 + ia_i \frac{\sigma_i}{2} + \ldots) \Psi \qquad (7.145)$$

Therefore our infinitesimal variation is $\delta \Psi = ia_i \frac{\sigma_i}{2} \Psi$, with arbitrary a_i. This tells us we get one conserved quantity for each generator, because the Lagrangian is invariant regardless of if two of the three a_i are zero and one isn't. For example, $a_2 = a_3 = 0$ and $a_1 \neq 0$ or $a_1 = a_2 = 0$ and $a_3 \neq 0$. Of course we get another conserved quantity for $a_1 \neq 0$, $a_2 \neq 0$ and $a_3 \neq 0$, which is just the sum of the conserved quantities we get from the individual generators. In other words: we get three independently conserved quantities, one for each generator of $SU(2)$.

The globally invariant, free Lagrangian (Eq. 7.44) is

$$\mathscr{L}_{\text{D1+D2}} = i \bar{\Psi} \gamma_\mu \partial^\mu \Psi.$$

The corresponding conserved quantities Q_i, for example for the electron-neutrino doublet, are[85]

[85] See Eq. 7.144 and as always defined without the arbitrary constants a_i.

$$Q_i = i \bar{\Psi} \gamma_0 \frac{\sigma_i}{2} \Psi$$
$$= \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \underbrace{\gamma_0 \gamma_0}_{=1} \frac{\sigma_i}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}. \qquad (7.146)$$

Recall that only σ_3 is diagonal. This means we are only able to assign a definite value to the two components of the doublet (v_e, e) for the conserved quantity $i = 3$. For the other generators, σ_1 and σ_2, our two components v_e and e aren't eigenstates. We are of course free to choose a different basis, where for example σ_2 is diagonal. Then we can simply redefine what we call v_e and e and get the same result. The thing to take away is that although we have three conserved quantities, one for each generator, we can only use one at a time to label our particles/states.

For $i = 3$ we have

$$Q_3 = \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \frac{\sigma_3}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} v_e \\ e \end{pmatrix}$$

$$= \frac{1}{2} v_e^\dagger v_e - \frac{1}{2} e^\dagger e \qquad (7.147)$$

This means we can assign $Q_3(v_e) = \frac{1}{2}$ and $Q_3(e) = -\frac{1}{2}$ as new particle labels. In contrast for $i = 1$, we have

$$Q_1 = \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \frac{\sigma_1}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} v_e \\ e \end{pmatrix}$$

$$= \frac{1}{2} v_e^\dagger e + \frac{1}{2} e^\dagger v_e \qquad (7.148)$$

and we can't assign any particle labels here, because the matrix σ_1 isn't diagonal.

7.7.1   Labelling States

Recall that in Section 3.5, we introduced the notion of Cartan generators, which is the set of diagonal generators of a given group. In the last section we learned that these generators become especially useful if we want to give new labels to our particles inside a doublet[86] object. A typical $SU(2)$ doublet is of the form

$$\begin{pmatrix} v_e \\ e \end{pmatrix}. \qquad (7.149)$$

We can't diagonalize two or more elements of the Lie algebra $\mathfrak{su}(2)$ at the same time and thus the "Cartan subalgebra" consists of only only one element[87]. It is conventional to choose $J_3 = \frac{1}{2}\sigma_3$ as diagonal. The corresponding eigenvalues are $+\frac{1}{2}$ and $-\frac{1}{2}$. A left-chiral neutrino $\begin{pmatrix} v_e \\ 0 \end{pmatrix}$ is an eigenstate of this generator, with eigenvalue $+\frac{1}{2}$ and a left-chiral electron $\begin{pmatrix} 0 \\ e \end{pmatrix}$ an eigenstate of this generator, with eigenvalue $-\frac{1}{2}$. These are new particle labels, which are called the isospin of the neutrino and the electron.

[86] In a later section we will learn that the same can be done for triplets and the conserved quantities following from $SU(3)$ invariance.

[87] This can be seen by noting that there are no commuting generators in $\mathfrak{su}(2)$. Only commuting generators can be diagonalized at the same time.

Following the same line of thoughts we can assign an isospin value to the right-chiral singlets. These transform according to the one-dimensional representation of $SU(2)$, and the generators are in this representation simply zero[88]: $J_i = 0$. Therefore, in this one-dimensional representation, the singlets are eigenstates of the Cartan generator J_3 with eigenvalue zero. The right-chiral singlets, like e_R carry isospin zero. This coincides with the remarks above that right-chiral fields do not interact via the weak force. Just as electrically uncharged objects do not interact via electromagnetic interactions, fields without isospin do not take part in weak interactions.

Finally, we can assign isospin values to the three gauge fields $W_+^\mu, W_-^\mu, W_3^\mu$. The three gauge fields form a $SU(2)$ triplet

$$W^\mu = \begin{pmatrix} W_+^\mu \\ W_-^\mu \\ W_3^\mu \end{pmatrix}, \tag{7.150}$$

which transforms according to the three dimensional representation of $SU(2)$. In this representation the Cartan generator J_3 has eigenvalues[89] $+1, -1, 0$ and therefore we assign $Q_3(W_+^\mu) = 1$, $Q_3(W_-^\mu) = -1$, $Q_3(W_3^\mu) = 0$. This is the isospin of the W_+ and the W_- bosons.

[89] This can be seen directly from the explicit matrix form of J_3 in Eq. 3.129:
$$J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Take note that the triplet $\begin{pmatrix} W_1^\mu & W_2^\mu & W_3^\mu \end{pmatrix}$ simply belongs to a different basis, where J_3 isn't diagonal. This can be seen as another reason for our introduction of $W_\pm^\mu$.

If this is unclear, have a look at how we introduced the three gauge fields W_i^μ. They were included into the Lagrangian in combination with the generators $\sigma_i W_i^\mu$. This can be seen as a basis expansion of some objects W^μ in terms of the basis σ_i:

$$\mathcal{W}^\mu = \sigma_i W_i^\mu = \sigma_1 W_1^\mu + \sigma_2 W_2^\mu + \sigma_3 W_3^\mu,$$

analogous to how we can write a vector in terms of basis vectors:

$$\vec{v} = v_1 \vec{e_1} + v_2 \vec{e_2} + v_3 \vec{e_3}.$$

The generators σ_i live in the Lie algebra[90] of $SU(2)$ and consequently our object $\mathcal{W}^\mu$ lives there, too. Therefore, if we want to know how $\mathcal{W}^\mu$ transforms, we need to know the representation of $SU(2)$ on this vector space, i.e. on its own Lie algebra. In other words: We need to know how the group elements of $SU(2)$ act on their own Lie algebra elements, i.e. its generators. This may seem like a strange idea at first, but actually is a quite natural idea. Recall how we defined a representation: a representation is a map[91] from the group to the space of linear operators over a vector space. So far we only looked

[90] The Lie algebra is a vector space!

[91] To be precise: A homomorphism, which is a map that satisfies some special conditions.

[93] $\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$

[94] $W^\mu = \begin{pmatrix} W_1^\mu \\ W_2^\mu \\ W_3^\mu \end{pmatrix}$

[95] As already noted in Section 2.4, Capital Roman letters $A, B, \ldots$ are always summed from 1 to 8.

[96] See Eq. 3.79 and the following text plus equations, where the basis was given by the 2×2 Pauli matrices.

[97] It can be shown that in general for $SU(N)$ the Lie-Algebra is $N^2 - 1$ dimensional.

at "external" vector spaces like Minkowski space. The only intrinsic vector space that comes with a group is its Lie algebra[92] ! Therefore it isn't that strange to ask what a group representation on this vector space looks like. This very important representation is called the **adjoint representation**.

Gauge fields (like W_+, W_-, W_3) are said to live in the adjoint representation of the corresponding group. For $SU(2)$ the Lie algebra is three dimensional, because we have three basis generators and therefore the adjoint representation is three dimensional. Exactly how we are able to write the components of a vector between two brackets[93], we can write the component of W^μ between two brackets[94], which is what we call a triplet . The generators in the adjoint representation are connected to the three dimensional generator we derived earlier through a basis transformation.

In the following section we move on to the "next higher" internal symmetry group $SU(3)$. Demanding local $SU(3)$ invariance of the Lagrangian gives us the correct Lagrangian which describes strong interactions.

7.8 $SU(3)$ **Interactions**

For **three** fermion fields we can find a locally $SU(3)$ invariant Lagrangian in exactly the same way we did in the last chapter for two fields and $SU(2)$. This symmetry is not broken and the corresponding spin 1 fields, called gluon fields, are massless. $SU(3)$ is the group of all unitary 3×3 matrices with unit determinant, i.e.

$$U^\dagger U = UU^\dagger = 1 \qquad \det U = 1. \tag{7.151}$$

As usual for Lie groups we can write this as an exponential function[95]

$$U = e^{iT_A \theta_A}. \tag{7.152}$$

The defining equations (Eq. 7.151) of the group require, as for[96] $SU(2)$, the generators to be Hermitian and traceless

$$T_A^\dagger = T_A \tag{7.153}$$

$$\text{tr}(T_A) = 0. \tag{7.154}$$

A basis for those traceless, Hermitian generators is, at least in *one* representation, given by eight[97] 3×3 matrices, called **Gell-Mann matrices**:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \tag{7.155}$$

The generators of the group are connected to these Gell-Mann matrices, just as the Pauli matrices were connected to the generators[98] of the $SU(2)$ group via $T_A = \frac{1}{2}\lambda_A$. The Lie algebra for this group is given by

$$[T_A, T_B] = if^{ABC} T^C, \tag{7.156}$$

where we adopted the standard convention that capital letters like A, B, C can take on every value from 1 to 8. f^{ABC} are called the structure constants of $SU(3)$, which for $SU(2)$ were given by the Levi-Civita symbol ϵ_{ijk}. They can be computed by brute-force computation, which yields[99]

$$f^{123} = 1 \tag{7.157}$$

$$f^{147} = -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} = \frac{1}{2} \tag{7.158}$$

$$f^{458} = f^{678} = \frac{\sqrt{3}}{2}, \tag{7.159}$$

where all others can be computed from the fact that the structure constants f^{ABC} are antisymmetric under permutation of any two indices. For example

$$f^{ABC} = -f^{BAC} = -f^{CBA}. \tag{7.160}$$

All other possibilities, which cannot be computed by permutation, vanish.

Analogous to what we did for $SU(2)$ in Section 7.2, we introduce a triplet of spin $\frac{1}{2}$ fields

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \tag{7.161}$$

and exactly as for $SU(2)$ we get new labels for the objects inside this triplet. We will discuss these new labels in the next section.

[98] $J_i = \frac{\sigma_i}{2}$, see Eq. 3.81 and the explanations there.

[99] This is not very enlightening, but we list it here for completeness.

To make the Lagrangian

$$\mathcal{L} = i\overline{Q}\partial_\mu\gamma^\mu Q - \overline{Q}mQ \tag{7.162}$$

locally $SU(3)$ invariant, one again adds coupling terms between the spin $\frac{1}{2}$ fields and new spin 1 fields. The derivation is analogous as for $SU(2)$, but the computations are quite cumbersome, so we just quote the final Lagrangian[100]

$$\mathcal{L} = -\frac{1}{4}\mathcal{G}_{\alpha\beta}\mathcal{G}^{\alpha\beta} + \overline{Q}(iD_\mu\gamma^\mu - m)Q, \tag{7.163}$$

and the field strength tensor $\mathcal{G}_{\alpha\beta}$ for the spin 1 gluon fields $\mathcal{G}_\alpha \equiv T^C G^C_\alpha$ is defined as

$$\mathcal{G}_{\alpha\beta} = \partial_\alpha\mathcal{G}_\beta - \partial_\beta\mathcal{G}_\alpha - g[\mathcal{G}_\alpha, \mathcal{G}_\beta]. \tag{7.164}$$

Here T^C denotes the generators of $SU(3)$ that were defined at the beginning of this section. Furthermore, D_α is defined as

$$D_\alpha = \partial_\alpha - igT^C G^C_\alpha = \partial_\alpha - ig\mathcal{G}_\alpha. \tag{7.165}$$

As you can check every term here is completely analogous to the $SU(2)$ case, except we now have different generators with different commutation properties.

7.8.1 Color

From global $SU(3)$ symmetry we get through Noether's theorem new conserved quantities. This is analogous to what we discussed for $SU(2)$ in Section 7.7. Following the same lines of thought as for $SU(2)$ tells us that we have 8 conserved quantities, one for each generator. Again, we can only use the conserved quantities that belong to the diagonal generators as particle labels. $SU(3)$ has two Cartan[101] generators $\frac{1}{2}\lambda_3$ and $\frac{1}{2}\lambda_8$. Therefore, every particle that interacts via the strong force carries two additional labels, corresponding to the eigenvalues of the Cartan generators.

The eigenvalues[102] of $\frac{1}{2}\lambda_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ are $+\frac{1}{2}, -\frac{1}{2}, 0$. For

$\lambda_8 = \frac{1}{2\sqrt{3}}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$ the eigenvalues[103] are $\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{-1}{\sqrt{3}}$. There-

fore if we arrange the strong interacting fermions into triplets (in the basis spanned by the eigenvectors of the Cartan generators), we can assign them the following labels, with some arbitrary spinor ψ:

$$\left(+\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \text{ for } \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}\psi,$$

[100] Remember: the sum over capital letters $(A, B, C, ...)$ runs from 1 to 8

[101] Recall, Cartan generators = diagonal generators.

[102] The eigenvectors are of course $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

[103] Corresponding to the same eigenvectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

where one usually defines **red**:= $\left(\frac{1}{2}, \frac{1}{2\sqrt{3}}\right)$. This means something of

the form $\begin{pmatrix} \Psi \\ 0 \\ 0 \end{pmatrix}$ is called red[104]. Analogously

$$\left(-\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \text{ for } \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}\psi$$

with **blue**:= $\left(\frac{-1}{2}, \frac{1}{2\sqrt{3}}\right)$ and for the third possibility we define
green:= $\left(0, \frac{-1}{\sqrt{3}}\right)$.

Completely analogous to what we did for $SU(2)$ we assign the color-charge zero to all $SU(3)$ singlets, which are then particles that do not interact via the strong force. Formulated differently: they are colorless. In addition, one can use the (8-dimensional[105]) adjoint representation of $SU(3)$ to assign color to the gauge fields G^μ_A, i.e. the gluons, completely analogous to how we assigned isospin to the W-Bosons in Section 7.7.

7.8.2   Quark Description

Recall that spin $\frac{1}{2}$ particles[106], which interact via the strong force are called quarks. If we want to talk about quarks we have to consider quite a lot of things:

- Quarks are $SU(3)$ triplets, denoted by Q. Inside a triplet we have

  the same quark, say an up-quark, in different colors: $U = \begin{pmatrix} u_r \\ u_b \\ u_g \end{pmatrix}$.

  The triplets always appear in pairs $\bar{Q}Q$ in order to get something $SU(3)$ invariant, exactly as we always need doublet pairs in order to get something $SU(2)$ invariant. Instead of writing $\bar{Q}Q$, we can use an index notation: $\bar{Q}Q = \bar{q}_c q_c$, where the index c stands for color $c = r, g, b$.

- In addition, left-chiral quarks are $SU(2)$ doublets, because they interact via the weak force, too. Each object in this doublet[107]

  (=each quark) is a triplet: $q = \begin{pmatrix} u_c \\ d_c \end{pmatrix}$. This can become very confus-

  ing, very fast and therefore the color index c is suppressed unless strong interactions are considered.

- As if this weren't enough, we need to remember that each quark is described by a Dirac spinor, which consits of two Weyl spinors

[104] Although we use here the familiar sounding labels "red", blue" and "green", there is absolutely no connection with the usual meaning of these words. In the context of $SU(3)$ interactions, we use these words simply as convenient labels. For a nice discussion on why we use color labels, see Chapter 1 in Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1st edition, 1 1984. ISBN 9780471887416.

[105] The adjoint representation of $SU(3)$ is 8 dimensional, because we have 8 generators.

[106] Spin $\frac{1}{2}$ are created by spin $\frac{1}{2}$ fields, as we will learn in Chapter 6.

[107] Each quark doublet consists of two different quarks, for example an up- and a down-quark or a top- and a bottom-quark.

$u_c = \begin{pmatrix} (\chi_u^L)_c \\ (\xi_u^R)_c \end{pmatrix}$. The upper Weyl-Spinor describes a left-chiral and the lower component the same quark with right-chirality.

Having talked about this, let's return to $SU(3)$ interactions. Happily, there is no experimental need for mass terms for the gauge bosons in the Lagrangian, because all experiments indicate that the gauge bosons of $SU(3)$, called gluons, are massless. Therefore $SU(3)$ is not broken.

Furthermore, the $SU(3)$ symmetry poses no new problems regarding mass terms for the fermions in the triplet, because a term of the form

$$\bar{Q}mQ \tag{7.166}$$

is $SU(3)$ invariant, as long as all particles in the triplet have equal mass. This means m is proportional to the unit matrix[108]. The objects inside a triplet describe the same quark in different colors, which indeed have equal mass. For example, for an up-quark the triplet is

$$U = \begin{pmatrix} u_r \\ u_b \\ u_g \end{pmatrix}, \tag{7.167}$$

where u_r denotes a red, u_b a blue and u_g green up-quark, which all have the same mass.

The other spin $\frac{1}{2}$ particles, like electrons or neutrinos, do not carry color and therefore do not couple to gluons. The interactions following from local $SU(3)$ invariance are called **strong interactions**, because the coupling constant is much bigger than for electromagnetic ($U(1)$) or weak ($SU(2)$) interactions.

[108] $m = m \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, instead of

$m = m \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}$

7.9   The Interplay Between Fermions and Bosons

This section summarizes what we discovered in this chapter and puts it in a more physical context. We will learn later that spin $\frac{1}{2}$ fields create and destroy spin $\frac{1}{2}$ particles. Analogously, spin 1 fields create and destroy spin 1 particles. In this chapter we derived the Lagrangians that describe how different fields and therefore particles interact with each other.

As already mentioned in Section 1.3 we call spin $\frac{1}{2}$ particles fermions and spin 1 particles bosons. The standard interpretation is that fermions make up matter and bosons mediate the forces between matter. We can now understand how this comes about.

We started the chapter with Lagrangians describing free fields, which we derived in Chapter 6. Then we discovered internal symmetries for the Lagrangian describing one, two or three free spin $\frac{1}{2}$ fields. These internal symmetries are only global symmetries, which is quite unconvincing because of special relativity. More natural would be local symmetries.

We then discovered that we could make the Lagrangians locally invariant by introducing additional coupling terms. These coupling terms describe the interaction of our spin $\frac{1}{2}$ fields with new spin 1 fields. For historic reasons the internal symmetries here are called gauge symmetries and we therefore call these new spin 1 fields, gauge fields. Through Noether's theorem we get for each internal symmetry new conserved quantities. These are interpreted as charges, analogous to electric charge that follows for $U(1)$ symmetry.

- To get a locally $U(1)$ invariant Lagrangian, we need one gauge field A^μ. The final Lagrangian describes correctly electromagnetic interactions. $U(1)$ symmetry tells us that electric charge is conserved.

- To get a locally $SU(2)$ invariant Lagrangian, we need three gauge fields $W_1^\mu, W_2^\mu, W_3^\mu$. The final Lagrangian describes correctly weak interactions. $SU(2)$ symmetry tells us that isospin is conserved.

- To get a locally $SU(3)$ invariant Lagrangian, we need eight such fields $G_1^\mu, G_2^\mu, \ldots$. The final Lagrangian describes correctly strong interactions. $SU(3)$ symmetry tells us that color is conserved.

Different gauge bosons (spin 1 particles) are responsible for a different kind of force. The electromagnetic force is mediated by photons, which is created by the $U(1)$ gauge field A_μ. The weak force is mediated by W^+, W^- and Z bosons and the strong force by 8 different gluons, which are created by the corresponding $SU(2)$ and $SU(3)$ gauge fields.

In addition, we discovered that $SU(2)$ symmetry forbids mass terms in the Lagrangian. From experiments we know this is incorrect. The solution that enables us to include mass terms without spoiling any symmetry is the Higgs mechanism. It works by including additional terms, describing coupling of our spin 1 and spin $\frac{1}{2}$ fields to a new spin 0 field, called the Higgs field. By breaking $SU(2)$ symmetry spontaneously and expanding the Higgs field around a new, no longer symmetric minimum, we get the required mass terms in the Lagrangian.

Part IV

Applications

8

Quantum Mechanics

Summary

In this chapter we will talk about quantum mechanics. The foundation for everything here are the identifications we discussed in Chapter 5. As a first result, we derive the **relativistic energy-momentum relation**.

After a discussion of how the quantum formalism works, we take the non-relativistic limit of the Klein-Gordon equation, because this is the equation of motion for the simplest type of particles: scalars. This results in the famous **Schrödinger equation**. The solution of this equation is interpreted as a probability amplitude and two simple examples are analysed using this **wave-mechanic** approach.

Afterwards, the **Dirac notation** is introduced, which is useful to understand the general structure of quantum mechanics. In this notation, the initial state of a system is denoted by an abstract state vector $|i\rangle$, called **ket**. The probability amplitude for measuring this initial state in a specific final state can then be computed formally by multiplying it with a **bra**, denoted $\langle f|$. The combination of a bra with a ket results in a complex number that is interpreted as probability amplitude A for the process $i \rightarrow f$. The probability for this process is $|A|^2$. Then we talk about **projection operators**. We will see how they can be used, together with the **completeness relation**, to expand an arbitrary state in the eigenstate basis of an arbitrary operator. The previously used wave-mechanics can then be seen as a special case, where we expand the states in the location basis. In the Dirac notation the Schrödinger equation is used to compute the time-evolution of states. To make the connection explicit, we use the Dirac notation to reanalyze an example which we already solved using wave-mechanics.

8.1 Particle Theory Identifications

The equations we derived so far[1] can be used in particle- and field-theories. In this chapter we want to investigate how they can be used in a particle-theory. Our dynamic variables are then the location, the angular momentum, the energy and the momentum of the particle in question. As explained in Chapter 5, we identify these with the generators of the corresponding symmetry[2]:

- momentum $\hat{p}_i = -i\partial_i$,

- location $\hat{x}_i = x_i$,

- energy $\hat{E} = i\partial_0$,

- angular momentum $\hat{L}_i = i\frac{1}{2}\epsilon_{ijk}(x^j\partial^k - x^k\partial^j)$.

Before we discuss how these operators are used in quantum mechanics, we use them to derive one of the most important equations of modern physics.

8.2 Relativistic Energy-Momentum Relation

In Section 6.2 we derived the equation of motion for a free spin 0 particle, the Klein-Gordon equation:

$$(\partial_\mu\partial^\mu + m^2)\Phi = 0.$$

With the identifications reiterated above this equation reads[3]

$$\begin{aligned}(\partial_\mu\partial^\mu + m^2)\Phi &= (\partial_0\partial_0 - \partial_i\partial_i + m^2)\Phi \\ &= \left(\left(\frac{1}{i}E\right)\left(\frac{1}{i}E\right) - \left(-\frac{1}{i}p_i\right)\left(-\frac{1}{i}p_i\right) + m^2\right)\Phi \\ &= (-E^2 + \vec{p}^2 + m^2)\Phi = 0.\end{aligned}$$

(8.1)

Therefore

$$\rightarrow E^2 = \vec{p}^2 + m^2 \quad \text{or using four-vectors} \quad p_\mu p^\mu = m^2, \qquad (8.2)$$

which is the famous **energy-momentum relation** of special-relativity. For a particle at rest ($\vec{p} = 0$) this gives us Einstein's famous equation

$$E^2 = m^2 \leftrightarrow E = mc^2$$

where we restored c^2 for clarity. We can now understand why we gave the scalar value of the first Casimir operator of the Poincaré group $p_\mu p^\mu$ the suggestive name m^2 (Eq. 3.266). The combination $p_\mu p^\mu$ is indeed the squared mass of the particle in question, which can be measured in experiments, for example, by measuring the energy and momentum of the particle: $m = \sqrt{E^2 - \vec{p}^2}$. For the same reason, we understand now why the constant in the Lagrangian we derived in Section 6.2 is also called m^2.

8.3   The Quantum Formalism

Now that our physical quantities are given by operators, we need something they can act on. First take note that we have for each operator a set of eigenfunctions, which is completely analogous to the eigenvectors of matrices. Matrices are finite-dimensional and therefore we get finite dimensional eigenvectors. Here our operators act on an infinite-dimensional vector space and we therefore have eigenfunctions. For example, the equation we must solve to get the eigenfunctions of the momentum operator is

$$\underbrace{-i\partial_i}_{\text{operator}}\Psi = \underbrace{p_i}_{\text{eigenvalue}}\overbrace{\Psi}^{\text{eigenfunction}},\tag{8.3}$$

with some number p_i. A solution is

$$\Psi = \underbrace{C}_{=const}e^{ip_ix_i} \quad \text{because}$$

$$\rightarrow -i\partial_i Ce^{ip_ix_i} = p_iCe^{ip_ix_i} \quad \checkmark \tag{8.4}$$

but take note that this a solution for arbitrary p_i. Therefore we have found an infinite number of eigenfunctions for the momentum operator $\hat{p}_i = -i\partial_i$. Equivalently, we can search for energy eigenfunctions

$$i\partial_0\Phi = E\Phi \tag{8.5}$$

or angular momentum eigenfunctions[4]. Analogous to eigenvectors for matrices, these eigenfunctions are bases for the corresponding abstract vector space[5]. This means we can expand an arbitrary function Ψ in terms of eigenfunctions. For example, we can expand Ψ in terms of momentum eigenfunctions[6] (for brevity in one dimension)

$$\Psi = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dp\Psi_p e^{-ipx}, \tag{8.6}$$

where Ψ_p are the coefficients in this expansion analogous to v_1, v_2, v_3 in $\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3$.

For some systems we have boundary conditions such that we have a discrete instead of a continuous basis. Then we can expand an arbitrary state, for example, in terms of discrete set of energy eigenfunctions Φ_{E_n}:

$$\Psi = \sum_n c_n\Phi_{E_n}. \tag{8.7}$$

An important observation is that, in general, a set of eigenfunctions for one operator is not a set of eigenfunctions for another operator.

[4] The discussion for angular momentum eigenfunctions is a bit more complicated, because the operator is more complicated than the others. We can't find a set of eigenfunctions for all three components at the same time, because $[\hat{L}_i, \hat{L}_j] \neq 0$. This will be discussed in a moment. The final result of a lengthy discussion is that the corresponding eigenfunctions of the third component of the angular momentum operator $\hat{L}_3$ (and of the squared angular momentum operator $\hat{L}^2$, which commutes with all components $[\hat{L}^2, \hat{L}_j] = 0$) are the famous **spherical harmonics**. These form an orthonormal basis.

[5] Recall that for matrices the eigenvectors are a basis for the corresponding vector space.

[6] Take note that this is exactly the Fourier transform of a function Ψ, which is introduced in Appendix D.1 and the factor $\frac{1}{\sqrt{2\pi}}$ is a matter of convention.

Only for operators that commute $[A, B] = AB - BA = 0$, we can find a simultaneous set of eigenfunctions for both operators. We don't prove this here, but to see the connection between a common set of eigenfunctions and the commutator property, let's assume we have two commuting operators $[A, B] = 0 \rightarrow AB = BA$. For the set of eigenfunctions Ψ_i that belong to A, we have $A\Psi_i = a_i\Psi_i$, where a_i denotes the corresponding eigenvalues. Now, we can calculate[7]

$$A(B\Psi_i) \underbrace{=}_{AB=BA} BA\Psi_i \underbrace{=}_{A\Psi_i=a_i\Psi_i} Ba_i\Psi_i \overbrace{=}^{a_i \text{ is just a number}} a_i(B\Psi_i). \tag{8.8}$$

This shows that $(B\Psi_i)$ is also an eigenfunction of A for all Ψ_i, with exactly the same eigenvalues a_i. Therefore the only difference between $(B\Psi_i)$ and Ψ_i can be a constant: $B\Psi_i = b_i\Psi_i$. We conclude that the A eigenfunctions Ψ_i are also eigenfunctions of B if $AB = BA$ holds.

In general our operators act on something we call[8] Ψ, which denotes the **state** of the physical system in question. We get this Ψ, by solving the corresponding equation of motion.

Usually, such a solution will have more than one term if we expand it in some basis. For example, consider a state that can be written in terms of two energy eigenstates[9] $\Psi = c_1\Phi_{E_1} + c_2\Phi_{E_2}$. Acting with the energy operator on this state yields

$$\hat{E}\Psi = \hat{E}(c_1\Phi_{E_1} + c_2\Phi_{E_2}) = c_1E_1\Phi_{E_1} + c_2E_2\Phi_{E_2} \neq E(c_1\Phi_{E_1} + c_2\Phi_{E_2}). \tag{8.9}$$

A **superposition** of states with different energy is therefore, in general, no eigenstate of the energy operator, because for an eigenstate we have by definition $\hat{E}\Psi = E\Psi$ for some number E. But what is then the energy of the system described by Ψ? What does it mean that a state is a superposition of two energy eigenstates? How can we interpret all this in physical terms?

A first hint towards an interpretation is the $U(1)$ symmetry of our Lagrangians, which shows us that the solution of an equation of motion Ψ cannot be directly physically relevant[10]. Secondly, we observe that a solution to any equation of motion we derived so far is a function of[11] $\vec{x}$ and t, i.e. $\Psi = \Psi(\vec{x}, t)$.

The standard interpretation is that the absolute value squared $|\Psi(\vec{x}, t)|^2$ of the **wave function** $\Psi(\vec{x}, t)$ gives the **probability density** of its location. Observe that the $U(1)$ symmetry has no influence on this quantity $|\Psi|^2 = \Psi^\dagger\Psi \rightarrow (\Psi')^\dagger(\Psi)' = \Psi^\dagger e^{-i\alpha}e^{i\alpha}\Psi = \Psi^\dagger\Psi$. In

[7] The brackets () are used here solely for illustrative purposes.

[8] The usage of Ψ is conventional in quantum mechanics and we use it here, although we used it so far exclusively for spinors, to describe spin 0 particles, too.

[9] This means all other coefficients in the expansion $\Psi = \sum_n c_n\Phi_{E_n}$ are zero.

[10] If we assume that Ψ describes our particle directly in some way, what would the $U(1)$ transformed solution $\Psi' = e^{i\alpha}\Psi$, which is equally allowed, describe?

[11] This will be made explicit in the next sections.

other words: $\Psi(x,t)$ is the probability amplitude that a measurement of the location gives a value in the interval $[x, x + dx]$. Consequently we have, if we integrate over all space

$$\int dx \Psi^\star(x,t)\Psi(x,t) \overset{!}{=} 1, \tag{8.10}$$

which is called the normalization prescription, because the probability for finding the particle anywhere in space must be $100\% = 1$.

If we want to make predictions about any other physical quantity, we must expand the wave-function in terms of the corresponding basis. For example, we can expand it in terms of energy eigenfunctions: $\Psi = c_1 \Phi_{E_1} + c_2 \Phi_{E_2} + \ldots$. Then, the standard interpretation of quantum mechanics is: the probability for measuring a given energy value E_1 for the system described by Ψ is given by the absolute value squared of the overlap between Ψ and Φ_{E_1}

$$P(E_1) = \left| \int dx \Phi_{E_1}^\star(x,t)\Psi(x,t) \right|^2.$$

In the example above this means

$$P(E_1) = \left| \int dx \Phi_{E_1}^\star(x,t)\Psi(x,t) \right|^2 = \left| \int dx \Phi_{E_1}^\star(x,t) \left(c_1 \Phi_{E_1} + c_2 \Phi_{E_2} \right) \right|^2$$

$$= \left| c_1 \underbrace{\int dx \Phi_{E_1}^\star(x,t)\Phi_{E_1}}_{=1 \text{ as explained above}} + c_2 \underbrace{\int dx \Phi_{E_1}^\star(x,t)\Phi_{E_2}}_{=0 \text{ because eigenstates are orthogonal}} \right|^2$$

$$= |c_1|^2. \tag{8.11}$$

Analogously, when we expand some other $\Psi(\vec{x},t)$ in terms of momentum eigenfunctions

$$\Psi(\vec{x},t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dp \, \tilde{\Psi}(\vec{p},t) e^{-i\vec{p}\vec{x}},$$

we have $\tilde{\Psi}(\vec{p},t)$ as the probability amplitude for finding the system with momentum in the interval $[p, p + dp]$.

This interpretation can be used to make probabilistic predictions about the system, for example using the statistical expectation value, which is the topic of the next section. Afterwards we will derive the equation of motion for non-relativistic quantum mechanics and look at two examples.

8.3.1 Expectation Value

In statistics the expectation value is defined in analogy to the weighted average. For example, if tossing a dice ten times results in 2, 4, 1, 3, 3,

6, 3, 1, 4, 5, the average value is

$$\langle x \rangle = (2 + 4 + 1 + 3 + 3 + 6 + 3 + 1 + 4 + 5) \cdot \frac{1}{10} = 3.2.$$

An alternative way of computing this is to collect equal results and weighing them by their empirical probability:

$$\langle x \rangle = \frac{2}{10} \cdot 1 + \frac{1}{10} \cdot 2 + \frac{3}{10} \cdot 3 + \frac{2}{10} \cdot 4 + \frac{1}{10} \cdot 5 + \frac{1}{10} \cdot 6 = 3.2.$$

We can write this in general as

$$\langle x \rangle = \sum_i \rho_i x_i \tag{8.12}$$

where ρ_i denotes the probability. Equally for a continuous distribution we have

$$\langle x \rangle = \int dx \rho(x) x \tag{8.13}$$

In quantum mechanics the **expectation value** for a physical quantity $\hat{O}$ is defined analogously

$$\langle \hat{O} \rangle = \int d^3 x \Psi^\star \hat{O} \Psi. \tag{8.14}$$

In general, we must expand Ψ in terms of eigenfunctions of $\hat{O}$, for example momentum eigenfunctions. Then, acting with the operator $\hat{O}$ on these states yields the corresponding eigenvalues and we get a weighted sum.

For example, the expectation value for the location of some particle is

$$\langle \hat{x} \rangle = \int d^3 x \Psi^\star \hat{x} \Psi = \int d^3 x \Psi^\star x \Psi = \int d^3 x \, x \underbrace{\Psi^\star \Psi}_{\text{probability density of its location}}. \tag{8.15}$$

In the next section, we consider the non-relativistic limit of the Klein-Gordon equation and will derive this way the equation of motion of non-relativistic quantum mechanics[12].

[12] It is also possible to consider the non-relativistic limit of the other equations that we have derived so far. For example, the Dirac equation becomes in the non-relativistic limit the Pauli equation. However, these additional equations are less important for the goals of this book.

8.4   The Schrödinger Equation

The Klein-Gordon equation is solved by plane waves

$$\Phi = e^{\pm i p_\mu x^\mu} \equiv e^{\pm i p \cdot x},$$

where $p_\mu = (E, \vec{p})^T$ is the conserved four-momentum of the particle. We check

$$\begin{aligned}
0 &= (\partial_\mu \partial^\mu + m^2) \Phi \\
&= (\partial_\mu \partial^\mu + m^2) e^{\pm i p_\mu x^\mu} \\
&= (i^2 p_\mu p^\mu + m^2) e^{\pm i p_\mu x^\mu} = 0 \\
&= (-m^2 + m^2) e^{\pm i p_\mu x^\mu} = 0 \qquad \checkmark
\end{aligned} \tag{8.16}$$

We now write the solution with a minus sign a little differently

$$\Phi = e^{-ip_\mu x^\mu} = \Phi = e^{i(-Et + \vec{p}\cdot\vec{x})}$$

and then we see that the time dependence is given by e^{-iEt}, i.e. $\Phi \propto e^{-iEt}$. From Eq. 8.2 we know the energy is

$$E = \sqrt{\vec{p}^2 + m^2} = \sqrt{m^2(\frac{\vec{p}^2}{m^2} + 1)} = m\sqrt{\frac{\vec{p}^2}{m^2} + 1}.$$

In the non-relativistic limit $|\vec{p}| \ll m$, i.e. for an object that moves much slower than the speed of light, which implies that its momentum is much smaller than its mass, we can approximate the energy using the Taylor-series as

$$\begin{aligned}
E &= m\sqrt{\frac{\vec{p}^2}{m^2} + 1} \\
&= m(1 + \frac{1}{2}\frac{\vec{p}^2}{m^2} + \ldots) \\
&\approx \underbrace{m}_{\text{rest-mass}} + \underbrace{\frac{\vec{p}^2}{2m}}_{\text{kinetic energy}}.
\end{aligned} \tag{8.17}$$

We can therefore write

$$\Phi = e^{i(-Et + \vec{p}\cdot\vec{x})} \approx e^{-imt}\underbrace{e^{i\vec{p}\cdot\vec{x} - i(\vec{p}^2/2m)t}}_{\equiv \phi(\vec{x},t)} = e^{-imt}\phi(\vec{x},t). \tag{8.18}$$

From $|\vec{p}| \ll m$ it follows that the rest-mass is much bigger than the kinetic energy and therefore the remaining time dependence in $\phi(\vec{x},t)$ oscillates more slowly than e^{-imt}. If we put our ansatz into the Klein-Gordon equation we get

$$(\partial_\mu\partial^\mu + m^2)e^{-imt}\phi(\vec{x},t) = (\partial_t\partial^t - \partial_i\partial^i + m^2)e^{-imt}\phi(\vec{x},t) = 0.$$

Then we use $\partial_t e^{-imt}(\ldots) = e^{-imt}(-im + \partial_t)(\ldots)$, which is just the product rule, twice. This yields

$$e^{-imt}((-im + \partial_t)^2 - \partial_i\partial^i + m^2)\phi(\vec{x},t) = 0,$$

which we can divide by e^{-imt}, because this never becomes zero. Therefore

$$\begin{aligned}
&\rightarrow ((-im + \partial_t)^2 - \nabla^2 + m^2)\phi(\vec{x},t) = 0 \\
&\rightarrow (-m^2 - 2im\partial_t + (\partial_t)^2 - \nabla^2 + m^2)\phi(\vec{x},t) = 0.
\end{aligned}$$

Comparing now the third term

$$\begin{aligned}
(\partial_t)^2\phi(\vec{x},t) &= (\partial_t)^2\exp\left[i\vec{p}\cdot\vec{x} - i(\vec{p}^2/2m)t\right] \\
&= (\vec{p}^2/2m)^2\exp\left[i\vec{p}\cdot\vec{x} - i(\vec{p}^2/2m)t\right] \propto \frac{p^4}{m^2} \tag{8.19}
\end{aligned}$$

with the second term

$$
\begin{aligned}
im\partial_t\phi(\vec{x},t) &= im\partial_t \exp\left[i\vec{p}\cdot\vec{x} - i(\vec{p}^2/2m)t\right] \\
&= m(\vec{p}^2/2m)\exp\left[i\vec{p}\cdot\vec{x} - i(\vec{p}^2/2m)t\right] \propto p^2 \qquad (8.20)
\end{aligned}
$$

shows us that, because of $|\vec{p}| \ll m$, we can neglect the third term in this limit. Therefore

$$
(-2im\partial_t - \nabla^2)\phi(\vec{x},t) = 0
$$

$$
\underset{\text{dividing by (-2m)}}{\longrightarrow} (i\partial_t + \frac{1}{2m}\nabla^2)\phi(\vec{x},t) = 0
$$

$$
\rightarrow (i\partial_t + \frac{\nabla^2}{2m})\phi(\vec{x},t) = 0, \qquad (8.21)
$$

which is the famous **Schrödinger equation**. When we now make the identifications we recited at the beginning of this chapter, the equation reads

$$
\rightarrow (E - \underbrace{\frac{\vec{p}^2}{2m}}_{=\text{kinetic energy}})\phi(\vec{x},t) = 0
$$

$$
\rightarrow E = \frac{\vec{p}^2}{2m}, \qquad (8.22)
$$

which is the usual **non-relativistic energy-momentum relation**. From this point-of-view, it is easy to see how we can include an external potential, because movement in an external potential simply adds a term describing the potential energy to the energy equation:

$$
\rightarrow E = \frac{\vec{p}^2}{2m} + V
$$

A famous example is the potential of a harmonic oscillator $V = -kx^2$. It is conventional to rewrite the Schrödinger equation, using the Hamiltonian operator $\hat{H}$, which collects all contributing energy operators, for example, the operator for the kinetic energy $\frac{\nabla^2}{2m}$ and the operator for the potential energy $\hat{V}$. Then we have

$$
i\partial_t\phi(\vec{x},t) = \underbrace{\left(\frac{\nabla^2}{2m} + \hat{V}\right)}_{\equiv\hat{H}} \phi(\vec{x},t) \rightarrow i\partial_t\phi(\vec{x},t) = \hat{H}\phi(\vec{x},t). \qquad (8.23)
$$

This is the standard way to write the Schrödinger equation.

8.4.1 Schrödinger Equation with an External Field

We can also follow the same procedure as in the last section and derive the non-relativistic limit of the **interacting** Klein-Gordon equation (Eq. 7.42), i.e. the equation that describes the interaction between a massive spin 0 field ϕ and a massless spin 1 field A_μ (the photon field). The resulting equation is

$$\left(i\partial_t - \frac{1}{2m} \left(\nabla - iq\vec{A} \right)^2 - qA_0 \right) \phi(\vec{x}, t) = 0. \qquad (8.24)$$

8.5   From Wave Equations to Particle Motion

Now, let's look at two examples of how the standard interpretation of quantum mechanics works in practice.

8.5.1   Example: Free Particle

A solution for the free (without external potential) Schrödinger equation (Eq. 8.21) is given by

$$\Psi = \mathrm{e}^{-i(Et - \vec{p}\vec{x})} \qquad (8.25)$$

because

$$i\partial_t \mathrm{e}^{-i(Et - \vec{p}\vec{x})} \overset{!}{=} -\frac{\nabla^2}{2m} \mathrm{e}^{-i(Et - \vec{p}\vec{x})}$$

$$\rightarrow E \mathrm{e}^{-i(Et - \vec{p}\vec{x})} \overset{!}{=} \frac{\vec{p}^2}{2m} \mathrm{e}^{-i(Et - \vec{p}\vec{x})} \quad \checkmark \,, \qquad (8.26)$$

where E is just the numerical value for the total energy of the free particle, which was derived in Eq. 8.22 as $E = \frac{\vec{p}^2}{2m}$. A more general solution is a linear combination

$$\Psi = A\mathrm{e}^{-i(Et - \vec{p}\vec{x})} + B\mathrm{e}^{-i(E't - \vec{p}'\vec{x})} + \dots.$$

We interpret the wave-function as a probability amplitude and therefore it must be normalized, because the total probability for finding the particle must be 100% $= 1$. This is not possible for the wave-function above which spreads out over all of space. To describe an individual free particle we have to use a suited linear combination, called a wave-packet:

$$\Psi_{WP}(\vec{x}, t) = \int dp^3 A(\vec{p}) e^{i(\vec{p}\vec{x} - Et)}, \qquad (8.27)$$

where the complex numbers $A(\vec{p})$ have to be chosen in a way that makes the wave packet normalizable. One possibility is a Gaussian
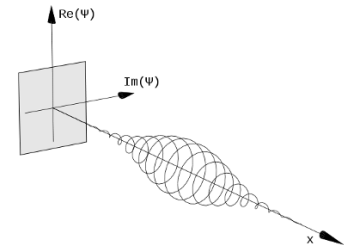

Fig. 8.1: Free wave-packet with Gaussian envelope. Figure by Inductiveload (Wikimedia Commons) released under a public domain licence. URL: `http://commons.wikimedia.org/wiki/File:Travelling_Particle_Wavepacket.svg`, Accessed: 4.5.2014
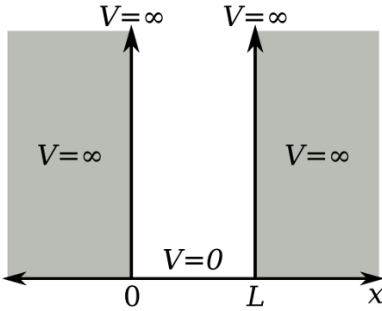
[13] Using $\sin(x) = \frac{1}{2i}\left(e^{ix} - e^{-ix}\right)$ and $\cos(x) = \frac{1}{2}\left(e^{ix} + e^{-ix}\right)$, which follows directly from the series expansion of $\cos(x), \sin(x)$ and e^{ix} as derived in Appendix B.4.1.

wave-packet, where $A(\vec{p})$ is a Gauss distribution.

$$\Psi_{GWP}(\vec{x},t) = \int dp^3 A(\vec{p})e^{i(\vec{p}\vec{x}-Et)} = \int dp^3 \psi_0 e^{i(\vec{p}-\vec{\bar{p}})^2/4\sigma^2} e^{i(\vec{p}\vec{x}-Et)} .$$

An example of such a Gaussian wave-packet is plotted in Fig. 8.1. For many computations clever tricks can be used in order to avoid working with complicated wave packets, allowing us to work with simple wave functions instead.

8.5.2 Example: Particle in a Box

Now we look at one of the standard examples of quantum mechanics: a particle confined in a box, here 1-dimensional, with infinitely high potential walls. Inside the box the potential is zero, outside it's infinite (see Fig. 8.2).

The potential is defined piece-wise

$$V(x) = \begin{cases} 0, & 0 < x < L \\ \infty, & \text{otherwise} \end{cases} \tag{8.28}$$

and therefore, we have to solve the one-dimensional Schrödinger equation

$$i\partial_t \Psi(x,t) = -\frac{\partial_x^2}{2m}\Psi(x,t) + V(x)\Psi(x,t)$$

piece-wise.

- Inside the box, the solution is equal to the free particle solution, because $V(x) = 0$ for $0 < x < L$.

- Outside, because here $V(x) = \infty$, the only possible, physical solution is $\Psi(x,t) = 0$.

We can rewrite the general free particle solution as[13]

$$\Psi(x,t) = Ae^{-i(Et-px)} + Be^{-i(Et+px)}$$

$$= \left(C\sin(px) + D\cos(px)\right)e^{-iEt},$$

which we can rewrite again using the non-relativistic energy-momentum relation (Eq. 8.22)

$$E = \frac{p^2}{2m} \qquad \rightarrow \qquad p = \sqrt{2mE}$$

$$\Psi(x,t) = \left(C\sin(\sqrt{2mE}x) + D\cos(\sqrt{2mE}x)\right)e^{-iEt} . \tag{8.29}$$

Next we use that the wave-function must be a continuous function. If there are any jumps in the wave-function, the momentum of

the particle $\hat{p}_x \Psi = -i\partial_x \Psi$ is infinite, because the derivative at the jump would be infinite. Therefore, we have the boundary conditions $\Psi(0) = \Psi(L) \overset{!}{=} 0$. Because $\cos(0) = 1$, we see that we have $D \overset{!}{=} 0$. Furthermore, we see that these conditions impose

$$\sqrt{2mE_n} \overset{!}{=} \frac{n\pi}{L}, \qquad (8.30)$$

with arbitrary integer n. This follows because for[14]

$$\Phi_n(x,t) = C\sin(\frac{n\pi}{L}x)e^{-iE_n t} \qquad (8.31)$$

both boundary conditions are satisfied

$$\to \Phi_n(L,t) = C\sin(\frac{n\pi}{L}L)e^{-iE_n t} = C\sin(n\pi)e^{-iE_n t} = 0 \quad \checkmark$$

$$\to \Phi_n(0,t) = C\sin(\frac{n\pi}{L}0)e^{-iE_n t} = C\sin(0)e^{-iE_n t} = 0 \quad \checkmark$$

The normalization constant C, can be found to be $C = \sqrt{\frac{2}{L}}$, because the probability for finding the particle anywhere inside the box must be $100\% = 1$ and the probability outside is zero, because there we have $\Psi = 0$. Therefore

$$P = \int_0^L dx \Phi_n^\star(x,t)\Phi_n(x,t) \overset{!}{=} 1$$

$$P = \int_0^L dx C^2 \sin(\frac{n\pi}{L}x)e^{+iE_n t}\sin(\frac{n\pi}{L}x)e^{-iE_n t}$$

$$= C^2\int_0^L dx \sin^2(\frac{n\pi}{L}x) = C^2\left[\frac{x}{2} - \frac{\sin(\frac{2n\pi}{L}x)}{4\frac{n\pi}{L}}\right]_0^L$$

$$= C^2\left(\frac{L}{2} - \frac{\sin(\frac{2n\pi}{L}L)}{4\frac{n\pi}{L}}\right) = C^2\frac{L}{2} \overset{!}{=} 1$$

$$\to C^2 \overset{!}{=} \frac{2}{L}.$$

We can now solve Eq. 8.30 for the energy E

$$E_n \overset{!}{=} \frac{n^2\pi^2}{L^2 2m}. \qquad (8.32)$$

We see here that the **possible energies are quantized**, which means that only a discrete set of values is allowed and not arbitrary continuous values[15]. Hence the name quantum mechanics.

Take note that we have a solution for each n and linear combinations of the form

$$\Phi(x,t) = A\Phi_1(x,t) + B\Phi_2(x,t) + \dots$$

are solutions, too. These solutions have to be normalised again because of the probabilistic interpretation[16].

[16] A probability of more than 1 = 100% doesn't make sense.

Next we can ask[17], what is the probability for measuring the particle having Energy $E = E_2 = \frac{2^2\pi^2}{L^2 2m}$. Say our particle is in the normalized state given by

[17] Instead of what's the probability to find a particle at place x.

$$\Psi(x,t) = \sqrt{\frac{3}{5}}\Phi_2(x,t) + \sqrt{\frac{2}{5}}\Phi_3(x,t).$$

The answer in the conventional interpretation of quantum mechanics is: it is the absolute value squared of the overlap between Ψ and Φ_2

$$P\left(E = \frac{2^2\pi^2}{L^2 2m}\right) = \left|\int dx \Phi_2^\star(x,t)\Psi(x,t)\right|^2,$$

where the overlap can be seen as a scalar product[18] of Φ_2 and Ψ:

[18] In fact, this is the scalar product of the Hilbert space in which our state vectors Ψ, Φ_n live.

$$(\Phi_2, \Psi) = \int dx \Phi_2^\star \Psi = \underbrace{c}_{\text{complex number}}.$$

The computation is easy, because the solutions we found are orthogonal, i.e.

$$\int dx \Phi_n^\star(x,t)\Phi_{n'}(x,t) = \delta_{nn'}.$$

[19] You can check this easily using integration by parts or something like Wolframalpha.com

For example[19],

$$\int_0^L dx \Phi_2^\star \Phi_3(x,t) = \int_0^L dx C \sin(\frac{2\pi}{L}x)\mathrm{e}^{+iEt} C \sin(\frac{3\pi}{L}x)\mathrm{e}^{-iEt}$$

$$= C^2 \int_0^L dx \sin(\frac{2\pi}{L}x)\sin(\frac{3\pi}{L}x) = 0.$$

Therefore, we get the probability for finding the energy $E = \frac{2^2\pi^2}{L^2 2m}$

$$P\left(E = \frac{2^2\pi^2}{L^2 2m}\right) = \left|\int dx \Phi_2^\star(x,t)\Psi(x,t)\right|^2$$

$$= \left|\int dx \Phi_2^\star(x,t)\left(\sqrt{\frac{3}{5}}\Phi_2(x,t) + \sqrt{\frac{2}{5}}\Phi_3(x,t)\right)\right|^2$$

$$= \left|\int dx \left(\sqrt{\frac{3}{5}}\underbrace{\Phi_2^\star(x,t)\Phi_2(x,t)}_{=1 \text{ if integrated}} + \sqrt{\frac{2}{5}}\underbrace{\Phi_2^\star(x,t)\Phi_3(x,t)}_{=0 \text{ if integrated}}\right)\right|^2$$

$$= \left(\sqrt{\frac{3}{5}}\right)^2. \tag{8.33}$$

Take note that we are able to call the functions we just found in Eq. 8.31, eigenstates of the energy operator $i\partial_t$ or equivalently of the Hamiltonian[20] $\hat{H} \equiv -\frac{\partial_x^2}{2m}$, because

[20] This follows directly from the Schrödinger equation $i\partial_t \Phi = -\frac{\partial_x^2}{2m}\Phi \equiv \hat{H}\Phi$.

$$\hat{H}\Phi_n = E_n\Phi_n. \tag{8.34}$$

If we act with the energy operator on an eigenstate, we get the same state multiplied with a constant, which we call energy of the state. In contrast, an arbitrary state is changed when the energy operator or the Hamiltonian operator $\hat{H}$ act on it. For example, if we take a look at the linear combination

$$\Psi = \sqrt{\frac{3}{5}}\Phi_2 + \sqrt{\frac{2}{5}}\Phi_3$$

we see that

$$\hat{H}\Psi = \hat{H}\left(\sqrt{\frac{3}{5}}\Phi_2 + \sqrt{\frac{2}{5}}\Phi_3\right)\underbrace{=}_{\text{Eq. 8.34}} \sqrt{\frac{3}{5}}E_2\Phi_1 + \sqrt{\frac{2}{5}}E_3\Phi_3$$

which cannot be written as multiple of Ψ because $E_2 \neq E_3$. Therefore, Ψ is not an eigenstate of the energy operator. Nevertheless, every wave function can be expressed in terms of the eigenstates Φ_n, because they form a complete basis set.

Next we want to introduce a useful notation invented by Dirac, which extremely helpful to understand the structure of quantum mechanics.

8.5.3 Dirac Notation

In the Dirac notation the state of a physical system is denoted abstractly by

$$|\Psi\rangle, \tag{8.35}$$

which is called a **ket**[21]. For example, if we prepare a particle in a box in an energy eigenstate we have

$$\hat{H}|\Phi_n\rangle = E_n|\Phi_n\rangle.$$

To each ket we can define a **bra**, denoted by $\langle\Psi|$ which is given by

$$\langle\Psi| = |\Psi\rangle^\dagger, \tag{8.36}$$

where the † symbol (called dagger) denotes the Hermitian conjugate, i.e. transposing plus complex conjugation. A bra is an object that acts on a ket. We can define an inner product as follows:

$$(|\Phi\rangle, |\Psi\rangle) \equiv \langle\Phi|\Psi\rangle.$$

If a ket is multiplied with a **bra** from the left-hand side, the result is a complex number

$$\langle\Phi|\Psi\rangle = c. \tag{8.37}$$

[21] The pun will become clear in a second.

This complex number is the probability amplitude for a physical system in the state $|\Psi\rangle$ to be measured in the state $|\Phi\rangle$. Consequently, the probability is given by $|\langle\Phi|\Psi\rangle|^2$. For example, the probability amplitude for finding a particle in the state $|\Psi\rangle$ in the interval $[x, x + dx]$ is given by

$$\langle x|\Psi\rangle \equiv \Psi(x).$$

This is the wave function we used in the last chapters. Furthermore, we could ask: What's the probability amplitude for finding the same particle with momentum in the interval $[p, p + dp]$? The answer in the Dirac notation is

$$\langle p|\Psi\rangle \equiv \Psi(p).$$

Recall that our states must fulfil a normalization condition, because we are using a probabilistic interpretation[22]. For example, we have

$$\int dx |\Psi(x,t)|^2 = \int dx \Psi^\star(x,t)\Psi(x,t) \overset{!}{=} 1 \qquad (8.38)$$

and equally

$$\int dp |\Phi(p,t)|^2 = \int dp \Phi^\star(p,t)\Phi(p,t) \overset{!}{=} 1 \qquad (8.39)$$

or written in our new notation

$$\int dx |\langle x|\Psi\rangle|^2 = \int dx (\langle x|\Psi\rangle)^\dagger \langle x|\Psi\rangle = \int dx \langle\Psi|x\rangle \langle x|\Psi\rangle \overset{!}{=} 1 \quad (8.40)$$

and

$$\int dp |\langle p|\Phi\rangle|^2 = \int dp (\langle p|\Phi\rangle)^\dagger \langle p|\Phi\rangle = \int dp \langle\Phi|p\rangle \langle p|\Phi\rangle \overset{!}{=} 1. \quad (8.41)$$

We can see here a new kind of operator: $|p\rangle \langle p|$ and $|x\rangle \langle x|$, which are called **projection operators**[23]. They are operators because they transform a ket into another ket. For example,

$$|x\rangle \underbrace{\langle x|\Psi\rangle}_{\text{=some complex function we call } \Psi(x)} = |x\rangle \Psi(x),$$

which is again a ket, because the product of a complex function with a ket is again a ket. In general, an operator is any object that acts on a ket to generate another ket. Looking at Eq. 8.40 motivates us to introduce another operator

$$\int dx \langle\Psi|x\rangle \langle x|\Psi\rangle = \langle\Psi| \underbrace{\left(\int dx |x\rangle \langle x|\right)}_{\equiv \hat{I}} |\Psi\rangle = \langle\Psi| \hat{I} |\Psi\rangle \overset{!}{=} 1. \quad (8.42)$$

From this we can conclude

$$\hat{I} |\Psi\rangle \overset{!}{=} |\Psi\rangle \qquad (8.43)$$

for an arbitrary ket $|\Psi\rangle$, because $\langle\Psi|\Psi\rangle = 1$. This follows, because the probability for a system prepared in state the $|\Psi\rangle$ to be found in the state $|\Psi\rangle$, must be, of course, $100\% = 1$. For example, if we prepare a particle to be at some point x_0, the probability for finding it at point x_0 is 1. Therefore, $\langle x_0|x_0\rangle = 1$. Because of this, $\hat{I}$ is called the unit operator and plays the same role as the number 1 in the multiplication of numbers. The result

$$\int dx\,|x\rangle\,\langle x| = \hat{I} \tag{8.44}$$

or for a discrete basis

$$\sum_i |i\rangle\,\langle i| = \hat{I} \tag{8.45}$$

are called **completeness relations**. In general, we say the component of a ket $|a\rangle$ in the basis $|i\rangle$ is

$$|i\rangle^\dagger\,|a\rangle \equiv \langle i|a\rangle \equiv a_i,$$

which is a complex number. Using the completeness relation, i.e. $\sum_i |i\rangle\,\langle i| = \hat{I}$, we can write

$$|a\rangle = \sum_i |i\rangle\,\langle i|a\rangle = \sum_i |i\rangle\,a_i.$$

This can be seen as the series expansion[24] of the ket $|a\rangle$ in terms of the basis $|i\rangle$. Consequently, the complex numbers a_i can be seen as the expansion coefficients. Equally for a continuous complete basis, we have

$$|\Psi\rangle = \int dx\,|x\rangle\,\underbrace{\langle x|\Psi\rangle}_{\equiv\Psi(x)\ \text{complex number}} = \int dx\,|x\rangle\,\Psi(x).$$

[24] Analogous to how we can write a vector in terms of a given basis: $\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3$, as explained in Appendix A.1.

The expectation value we introduced in Section 8.3.1 is in the Dirac notation given by

$$\langle\hat{O}\rangle = \langle\Psi|\,\hat{O}\,|\Psi\rangle\,.$$

Now let's return to the example of the particle in a box, and solve it using the Dirac notation.

8.5.4   Example: Particle in a Box, Again

The question we asked was: What is the probability for finding the particle having energy $E = E_2 = \frac{n^2\pi^2}{2mL^2}$? This question can be answered in the Dirac notation in a very natural way. The probability is given by

$$P(E_2) = \big|\,\langle E = \tfrac{n^2\pi^2}{2mL^2}|\Psi\rangle\,\big|^2 = \Big|\,\langle E = \tfrac{n^2\pi^2}{2mL^2}|\underbrace{\Big(\int dx\,|x\rangle\,\langle x|\Big)}_{=\hat{I}}|\Psi\rangle\,\Big|^2$$

$$= \Big| \int dx \, \langle E = \frac{n^2 \pi^2}{2mL^2} | x \rangle \, \langle x | \Psi \rangle \Big|^2 = \Big| \int dx \Phi_2^\star(x) \Psi(x) \Big|^2 ,$$

which is exactly the result we derived using the standard wave mechanics in Section 8.5.2. All of this becomes much clearer as soon as you learn more about quantum mechanics and solve some problems on your own, using both notations.

8.5.5   Spin

Now it's time to return to the operator we derived in Section 5.1.1 for a new kind of angular momentum, we called spin. So far we have two loose ends. On the one hand, we have used spin as a label for the representations of the Lorentz group. For example, if we want to describe an elementary particle with spin $\frac{1}{2}$, we have to use an object transforming according to the spin $\frac{1}{2}$ representation. On the other hand, we derived from rotational symmetry something that we called spin, too[25]. In the quantum framework, when we act with the operator that we derived in Section 5.1.1 on a given state, the result is the spin of the particle. For the scalar representation, this operator is given by $\hat{S} = 0$, which of course always yields 0 when with it on a state Ψ.

For the $(\frac{1}{2}, 0)$ representation we have to use the two-dimensional representation of the rotation generator, which was derived in Section 3.7.5

$$\hat{S}_i = \frac{\sigma_i}{2} \tag{8.46}$$

and where σ_i denotes the Pauli matrices. If we want to know the spin of a particle described by Ψ, we have to act with the spin-operator on Ψ. For example, for $\hat{S}_3$ this will give us the spin of the corresponding particle in the 3-, or more familiar called z-direction. Analogously for $\hat{S}_2$ in the y- and $\hat{S}_1$ in the x-direction.

The explicit form of the operator $\hat{S}_3$ is

$$\hat{S}_3 = \frac{\sigma_3}{2} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \tag{8.47}$$

The corresponding eigenstates are

$$v_{\frac{1}{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad v_{-\frac{1}{2}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{8.48}$$

with eigenvalues $\frac{1}{2}$ and $-\frac{1}{2}$, respectively.

[25] The conserved quantity that follows from invariance under rotations has two parts. One part follows from the invariance under rotations of the spacetime coordinates and is called orbital angular momentum. The second parts follows from the invariance under the mixing of the field components. See Section 4.5.4.
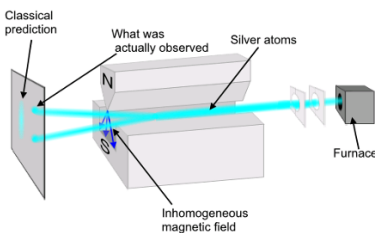

Fig. 8.3: Illustration of the Stern-Gerlach experiment. The original experiment was performed with silver atoms, whose spin behavior is dominated by the one electron in the outermost atomic orbital. The experimental result is the same as with electrons. A beam of particles is affected by an inhomogeneous magnetic field. For a classical type of angular momentum the deflection of the particles through the magnetic field should be a continuous distribution. Measured are just two deflection types, i.e. the beam splits in two parts, one corresponding to spin $\frac{1}{2}$ and one to $-\frac{1}{2}$. Figure by Theresa Knott (Wikimedia Commons) distributed under a CC BY-SA 3.0 license: http://creativecommons.org/licenses/by-sa/3.0/deed.en. URL: http://commons.wikimedia.org/wiki/File:Stern-Gerlach_experiment.PNG, Accessed: 24.5.2014.

This means a particle described by a spinor[26] has spin $\frac{1}{2}$, which can be aligned or anti-aligned to some arbitrary measurement axis. This is why we call this representation spin $\frac{1}{2}$ representation[27]. In the quantum framework this is interpreted by saying that a measurement of spin can only result in $\frac{1}{2}$ and $-\frac{1}{2}$. In Section 4.5.4 we learned that spin is something similar to orbital angular momentum, because both notions arise from rotational invariance. Here we can see that this kind of angular momentum gives quite surprising results, when measured. The most famous experiment proving this curious fact of nature is the Stern-Gerlach experiment (see Fig. 8.3).

The same is true for a measurement of spin in any direction. A measurement of the spin in the x-,y- or z-direction can only result in $\frac{1}{2}$ and $-\frac{1}{2}$.

Let's look at one concrete example of how a spin-measurement works in the quantum formalism. As mentioned above, the explicit form of the spin operator (Eq. 5.4), say for a measurement along the z-axis, is

$$\hat{S}_z = \frac{1}{2}\sigma_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}. \tag{8.49}$$

The eigenstates are $|\frac{1}{2}\rangle_z \triangleq \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|-\frac{1}{2}\rangle_z \triangleq \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, where the subscript z denotes that we are dealing with eigenstates of $\hat{S}_z$. A general spinor is not a spin-eigenstate, but a superposition $|X\rangle = a\,|\frac{1}{2}\rangle_z + b\,|-\frac{1}{2}\rangle_z$. The coefficients depend on how we prepare a given particle. If we did a measurement of the spin along the z-axis and filtered out all particles with spin $-\frac{1}{2}$, the coefficient b would be zero and a would be 1. Without any measurement and filtering the coefficients are $a = b = \frac{1}{\sqrt{2}}$, which means probability[28] $\frac{1}{2}$ for each possibility. Things get interesting if we make a measurement along the z-axis and afterwards a measurement, for example, along the x-axis. Even if we did filter out all $-\frac{1}{2}$ components along the z-axis, there will be particles with spin $-\frac{1}{2}$ along the x-axis.

Acting with the spin operator $\hat{S}_z$ on a state means measuring spin along the z-axis. For a general state $|X\rangle$ both outcomes $+\frac{1}{2}$ and $-\frac{1}{2}$ are possible and the probability is directly related to the factors a and b. If we want to know the probability for measuring $-\frac{1}{2}$, the quantum formalism tells us that the corresponding probability amplitude is

$$_z\langle -\frac{1}{2}|X\rangle = a\underbrace{_z\langle -\frac{1}{2}|\frac{1}{2}\rangle_z}_{=0} + b\underbrace{_z\langle \frac{1}{2}|-\frac{1}{2}\rangle_z}_{=1} = b \tag{8.50}$$

[26] Recall that a spinor is an object transforming according to the $(\frac{1}{2},0)$, the $(0,\frac{1}{2})$ or $(\frac{1}{2},0) \oplus (0,\frac{1}{2})$ representation.

[27] Analogous statements can be made for the spin 1 or other higher representations.

[28] The coefficients are directly related to the probability amplitude which we need to square in order to get the probability. This will be shown in a moment.

Therefore, the probability for measuring spin $-\frac{1}{2}$ along the z-axis is $P_{z=-\frac{1}{2}} = |b|^2$. If we want to measure the spin along another axis, say the x-axis, we need to expand our two states in terms of the eigenstates of $\hat{S}_x$, which reads in explicit matrix form (Eq. 5.4)

$$S_x = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \tag{8.51}$$

and the corresponding normalized eigenstates are $|\frac{1}{2}\rangle_x \mathrel{\hat{=}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $|-\frac{1}{2}\rangle_x \mathrel{\hat{=}} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. If we want to know the probability for measuring spin $-\frac{1}{2}$ along the x-axis, we need to rewrite $|\frac{1}{2}\rangle_z$ and $|-\frac{1}{2}\rangle_z$ in terms of $|\frac{1}{2}\rangle_x$ and $|-\frac{1}{2}\rangle_x$:

$$\underbrace{|\frac{1}{2}\rangle_z}_{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} = \frac{1}{\sqrt{2}} \Big(\underbrace{|\frac{1}{2}\rangle_x}_{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} + \underbrace{|-\frac{1}{2}\rangle_x}_{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}} \Big) \tag{8.52}$$

$$\underbrace{|-\frac{1}{2}\rangle_z}_{\begin{pmatrix} 0 \\ 1 \end{pmatrix}} = \frac{1}{\sqrt{2}} \Big(\underbrace{|\frac{1}{2}\rangle_x}_{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} - \underbrace{|-\frac{1}{2}\rangle_x}_{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}} \Big) . \tag{8.53}$$

And therefore

$$|X\rangle = a |\frac{1}{2}\rangle_z + b |-\frac{1}{2}\rangle_z = a \frac{1}{\sqrt{2}} \Big(|\frac{1}{2}\rangle_x + |-\frac{1}{2}\rangle_x \Big) + b \frac{1}{\sqrt{2}} \Big(|\frac{1}{2}\rangle_x - |-\frac{1}{2}\rangle_x \Big). \tag{8.54}$$

The probability amplitude for measuring $-\frac{1}{2}$ for the spin along the x-axis is then

$$_x\langle -\frac{1}{2}|X\rangle = {}_x\langle -\frac{1}{2}| \Big(a \frac{1}{\sqrt{2}} \Big(|\frac{1}{2}\rangle_x + |-\frac{1}{2}\rangle_x \Big) + b \frac{1}{\sqrt{2}} \Big(|\frac{1}{2}\rangle_x - |-\frac{1}{2}\rangle_x \Big) \Big)$$

$$= \frac{a}{\sqrt{2}} - \frac{b}{\sqrt{2}} \tag{8.55}$$

and the probability is $P_{x=-\frac{1}{2}} = |\frac{a}{\sqrt{2}} - \frac{b}{\sqrt{2}}|^2$.

Now, let's come back to the example outlined at the beginning of this computation. If we filter out all particles with spin $-\frac{1}{2}$ along the z-axis the state $|X\rangle$ is

$$|X\rangle_{\text{after } z\text{-axis filtering}} = |\frac{1}{2}\rangle_z . \tag{8.56}$$

This means $a = 1$ and $b = 0$ and we get a non-zero probability for measuring spin $-\frac{1}{2}$ along the x-axis $P_{x=-\frac{1}{2}} = |\frac{1}{\sqrt{2}} - \frac{0}{\sqrt{2}}|^2 = \frac{1}{2}$. If we now filter out all particles with spin $-\frac{1}{2}$ along the x-axis and repeat our measurement of spin along the z-axis we notice something quite remarkable. After filtering the particles with spin $-\frac{1}{2}$ along the x-axis we have the state

$$|X\rangle_{\text{after x-axis filtering}} = |\frac{1}{2}\rangle_x . \tag{8.57}$$

If we want to know the probability for measuring spin $-\frac{1}{2}$ along the z-axis, we need to write $|\frac{1}{2}\rangle_x$ in terms of $|\frac{1}{2}\rangle_z$ and $|-\frac{1}{2}\rangle_z$:

$$|X\rangle_{\text{after x-axis filtering}} = \underbrace{|\frac{1}{2}\rangle_x}_{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}} = \frac{1}{\sqrt{2}}\Big(\underbrace{|\frac{1}{2}\rangle_z}_{\begin{pmatrix}1\\0\end{pmatrix}} + \underbrace{|-\frac{1}{2}\rangle_z}_{\begin{pmatrix}0\\1\end{pmatrix}} \Big) \tag{8.58}$$

and we get the probability $P_{z=-\frac{1}{2}} = |_z\langle -\frac{1}{2}|X\rangle|^2 = \frac{1}{2}$.

To summarize, this means:

- We start with a spin measurement along the z-axis and filter out all particles with $-\frac{1}{2}$. This leaves us with a state

$$|X\rangle_{\text{after z-axis filtering}} = |\frac{1}{2}\rangle_z . \tag{8.59}$$

- If we now measure again the spin along the z-axis we get a very unsurprising result: The probability for measuring spin $-\frac{1}{2}$ is zero and for spin $+\frac{1}{2}$ the probability is 100%.

$$_z\langle \frac{1}{2}|X\rangle_{\text{after z-axis filtering}} = 1 , \tag{8.60}$$

$$_z\langle -\frac{1}{2}|X\rangle_{\text{after z-axis filtering}} = 0 . \tag{8.61}$$

- If we then measure the spin along the x-axis, for our z-filtered particle stream, we get a probability of $\frac{1}{2} = 50\%$ for a measurement of $+\frac{1}{2}$. Equivalently, we have a probability of $\frac{1}{2} = 50\%$ for a measurement of $-\frac{1}{2}$. If we then filter out all components with spin $-\frac{1}{2}$ along the x-axis we are in the state $|X\rangle_{\text{after x-axis filtering}} = |\frac{1}{2}\rangle_x$.

- Now measuring the spin along the z-axis again, gives us the surprising result that the probability for measuring spin $-\frac{1}{2}$ is $\frac{1}{2} = 50\%$. The measurement along the x-axis did change the state and therefore we are again getting components with spin $-\frac{1}{2}$ along the z-axis, even though we did filter these out in the first step!

A brilliant discussion of these matters, involving real measuring devices, can be found in the Feynman Lectures[29] Vol. 3.

[29] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics, Volume 3*. Addison Wesley, 1st edition, 1 1971. ISBN 9780201021189

8.6 Heisenberg's Uncertainty Principle

Now it's time to talk about one of the most curious features of quantum mechanics. We learned in the last section that a measurement of spin in the x-direction makes everything we knew previously about spin along the z-direction useless. This kind of thing happens for many observables in quantum mechanics. We can trace this behavior back to the fact that[30] $\hat{S}_x\hat{S}_z \neq \hat{S}_z\hat{S}_x$. This means that a measurement of spin along the z-axis followed by a measurement of spin along the x-axis is different from a measurement of spin along the x-axis followed by a measurement of spin along the z-axis. After measuring the spin along the z-axis the system is in an eigenstate of $\hat{S}_z$ and after a measurement of spin along the x-axis, in an eigenstate of $\hat{S}_x$. The eigenstates for $\hat{S}_z$ and $\hat{S}_x$ are all different and therefore this is no surprise.

[30] Recall that we identify the spin operators with the corresponding finite-dimensional representations for the rotation generators. These fulfil the commutator relation $[J_i, J_j] = J_iJ_j - J_jJ_i = i\epsilon_{ijk}J_k \neq 0 \rightarrow J_iJ_j \neq J_jJ_i$. For example, if we describe spin $\frac{1}{2}$ particles, we must use the two-dimensional representation $J_i = \frac{\sigma_i}{2}$.

We can look at this from a different perspective: **We aren't able to know the spin of a system along the z-axis and the x-axis at the same time!** Each time we measure spin along the z-axis the spin along the x-axis becomes undetermined and vice-versa. The same is true for spin along the z-axis/x-axis and spin along the y-axis. Spin may be something really strange, but we can observe the same behavior for measurements of position and momentum. Take a look again at Eq. 5.3, which we recite here for convenience:

$$[\hat{p}_i, \hat{x}_j] = \hat{p}_i\hat{x}_j - \hat{x}_j\hat{p}_i = i\delta_{ij}. \tag{8.62}$$

Following the line of thought as above tells us that a measurement of momentum in the x-direction changes what we can expect for a measurement of position on the x-axis. In other words this means that we can't know position and momentum in the same direction at the same time with arbitrary precision. Take note that the commutator is only non-zero for measurements along the same axis[31]. A measurement of momentum in the y-direction has no influence on what we can expect for the position on the x-axis.

[31] The Kronecker delta δ_{ij} is zero for $i \neq j$ and one for $i = j$ as defined in Appendix B.5.5.

Everytime we measure momentum the position becomes uncertain and vice versa. This is known as **Heisenberg's uncertainty principle**. Analogous observations can be made for angular momentum along different axes, because the commutator for the corresponding operator is non-zero, too. In general, we can check for any two physical quantities if they commute with each other. If they don't, we know that they can't be measured at the same time with arbitrary precision.

Maybe this shouldn't surprise us. Quantum mechanics uses the generators of the corresponding symmetries as measurement oper-

ators. For instance, this has the consequence that a measurement of momentum is equivalent to the action of the translation generator[32]. The translation generator moves our system a little bit and therefore the location is changed. What is more surprising is that nature actually works this way. Over the years there have been many experimental tests of the Heisenberg's uncertainty principle and all proved it to be correct.

[32] Recall that invariance under translations leads us to conservation of momentum.

8.7   Comments on Interpretations

The interpretation and notations described in this chapter are the standard ones. Nevertheless, there are other formalisms equally powerful. For example, the Feynman path integral formalism[33] is, in terms of results, completely equivalent to the wave mechanics, we described in this chapter. Yet computations in this formalism are completely different. If we want to compute the probability for a particle to get from point a to point b, we have to sum over all possible paths that a particle can go between a and b. As absurd as it sounds, this approach leads to the same result, which can be proved formally as well. Freeman Dyson once told the story[34]

[33] To learn more about this see for example, Richard P. Feynman and Albert R. Hibbs. *Quantum Mechanics and Path Integrals: Emended Edition*. Dover Publications, emended editon edition, 7 2010. ISBN 9780486477220

[34] Harry Woolf, editor. *Some Strangeness in the Proportion*. Addison-Wesley, 1st edition, 2 1981. ISBN 9780201099249

> Dick Feynman told me about his "sum over histories" version of quantum mechanics. "The electron does anything it likes," he said. "It just goes in any direction at any speed, forward or backward in time, however it likes, and then you add up the amplitudes and it gives you the wavefunction." I said to him, "You're crazy." But he isn't.

Another interpretation for the basic equations of quantum mechanics, further away from the mainstream, is Bohmian Mechanics. The starting point is to put the ansatz: Re^{St} into the Schrödinger equation. Separating the imaginary and real part results in two equations, one of which can be seen as completely analogous to the Hamilton-Jacobi equation of classical mechanics plus an additional term. This additional term can be interpreted as an extra potential, which is responsible for the strange quantum effects. Further computations are completely analogous to classical mechanics. A new force is computed from the extra potential using the gradient, which is then put into Newton's classical equation: $F = ma$. Therefore, in Bohmian mechanics one still has classical particle trajectories. The results from this approach are, as far as I know, equal to those computed by standard non-relativistic quantum mechanics. Nevertheless, this approach has fallen into disfavour because the extra potential undergoes non-local changes.

[35] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics, Volume 3*. Addison Wesley, 1st edition, 1 1971. ISBN 9780201021189

[36] David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Prentice Hall, 2nd edition, 4 2004. ISBN 9780131118928

[37] J. J. Sakurai. *Modern Quantum Mechanics*. Addison Wesley, 1st edition, 9 1993. ISBN 9780201539295

[38] Paul A. M. Dirac and Physics. *Lectures on Quantum Mechanics*. Dover Publications, 1st edition, 3 2001. ISBN 9780486417134

Further Reading Tips

- **Richard P. Feynman - The Feynman Lectures on Physics, Vol. 3**[35] is a great book to start learning about quantum mechanics. Most concepts of quantum mechanics are explained here more lucidly than anywhere else.

- **David J. Griffiths - Introduction to Quantum Mechanics**[36] is a very readable and enlightening book.

- **J. J. Sakurai - Modern Quantum Mechanics**[37] is a brilliant book, which often offers a unique perspective on the concepts of quantum mechanics.

- **Paul A. M. Dirac - Lectures on Quantum Mechanics**[38] is an old book by one of the fathers of quantum mechanics. Highly recommended, because reading about ideas by the person who discovered them is always a good idea.

- You can find a detailed discussion of the interpretation of the components of a Dirac spinor in Appendix 8.8.

8.8   Appendix: Interpretation of the Dirac Spinor Components

In this and the corresponding appendices, u and v denote two-component objects inside a Dirac spinor and u and v four-component objects. For example u_1 and u_2 describe two different four-component objects. For a general four-component object u, we denote the two two-component objects by u_1 and u_2, i.e. $u = \begin{pmatrix} \mathrm{u}_1 \\ \mathrm{u}_2 \end{pmatrix}$.

Up to this point we have been relatively vague about the two Weyl spinors inside a Dirac spinor. What do they stand for? How can they be interpreted? In addition, each such Weyl spinor inside a Dirac spinor consists of two components. How to interpret these? Now, we are finally in the position to give answers to these questions.

In short:

- The two Weyl spinors χ_L, ξ_R inside a Dirac spinor $\psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}$ describe "different particles". Nevertheless, it's conventional to call them the *same* particle, for example an electron, with different chirality:

  – χ_L describes a left-chiral electron,
  – ξ_R describes a right-chiral electron,

but the crucial point is that these are really distinct particles/fields[39], because they aren't related by a parity transformation or charge conjugation. This is why we use different symbols. There is, of course, some sort of connection between them, which is why we write them in one object. This will be discussed in detail in a moment.

[39] They are labelled by different quantum numbers and therefore behave differently in experiments!

- The two components of each Weyl spinor describe different spin configurations[40] of the corresponding particle.

  - A Weyl spinor proportional to $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ describes a particle with **spin up**

  - A Weyl spinor proportional to $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ describes a particle with **spin down**

  - Any other Weyl spinor is simply a mixture of spin up and spin down.

[40] This was discussed in Section 8.5.5. Recall that spin can be measured like angular momentum, but for a spin $\frac{1}{2}$ particle the result of such a measurement can only be $+\frac{1}{2}$ or $-\frac{1}{2}$, no matter what axis we choose. These two measurement results are commonly called **spin up** and **spin down**.

Let's see how this comes about in detail:

The important thing we learn from weak interactions and parity violation is that left-chiral and right-chiral particles are really different particles. Left-chiral particles carry weak charge (isospin) and therefore interact via the weak force. Right chiral particles do not, which was explained in Section 7.7.1.

We have for every particle in nature a corresponding antiparticle and in general, we get the antiparticle description through charge conjugation. Charge conjugation flips all particle labels, which includes isospin. Let's see what particles we can expect that are related to, say electrons. We have

- A left-chiral *electron* χ_L, with isospin $-\frac{1}{2}$ and electric charge $-e$, which is part of a doublet.

- A right-chiral *anti-left-chiral-electron* $(\chi_L)^c = \chi_R$ with isospin $\frac{1}{2}$, electric charge $+e$, which is part of a doublet, too. This property does not simply vanish through charge conjugation. This may be confusing at the moment, but so far we have only talked about the coupling of the weak force to particles. It turns out that the weak force couples to right-chiral antiparticles as well.

- A right-chiral *electron* ξ_R with isospin 0 and electric charge $-e$

- A left-chiral *anti-right-chiral-electron* $(\xi_R)^c = \xi_L$ with isospin 0 and electric charge $+e$

Therefore, when talking about electrons, we have in fact **four** different "things" we need to consider. These are all really different particles and we need to give them different names! Usually one talks about just two particles related to an electron: The electron and the positron and we will see how this comes about in a moment.

We restrict the following discussion to the rest frame of the particles in question. In other frames the discussion works analogous but is more cumbersome.

The objects inside a Dirac spinor and its charge conjugate are directly related to these four particles. The **physical electron** and the **physical positron** are commonly identified as the solutions of the Dirac equation. The Dirac equation is an equation of motion, i.e. an equation that determines the dynamics of the particles in question. The explicit solutions are Dirac spinors that evolve in time. We need such Dirac spinors with a definite time-evolution in order to describe how our particles evolve in time. As we will see this requires that we always use two of the particles listed above at once.

An explicit derivation of the solutions of the Dirac equation can be found in the appendix Section 8.9. Here we just use the results. There are four independent solutions of the Dirac equation and two are of the form

$$\psi_i = \begin{pmatrix} u_i \\ u_i \end{pmatrix} \tag{8.63}$$

[41] This is a basis choice. The only requirement is that they are linearly independent.

with for example[41] $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-imt}$ and $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-imt}$ and two solutions are of the form

$$\tilde{\psi}_i = \begin{pmatrix} -v_i \\ v_i \end{pmatrix}, \tag{8.64}$$

with for example $v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{+imt}$ and $v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{+imt}$

The Dirac equation tells us that the spin configuration of the two particles described by the upper and lower Weyl spinors inside a Dirac spinor, are directly related. In addition, their time dependence must be the same. These solutions describe what is commonly known as a **physical electron** and a **physical positron**, with different spin configurations[42].

[42] Take note that the connection between these objects is charge conjugation. This can be seen by using the explicit form of the charge conjugation operator: $\psi_1^c = i\gamma_2 \psi_1^\star$. Therefore: $(\psi_1)^c = \tilde{\psi}_2$ and $(\psi_2)^c = \tilde{\psi}_1$.

- ψ_1 is an electron with spin up

- ψ_2 is an electron with spin down

- $\tilde{\psi}_1$ is a positron with spin up

- $\tilde{\psi}_2$ is a positron with spin down

We can see nicely that a physical electron has a left-chiral (the upper two components) and a right-chiral part (the lower two components). For a physical electron the spin configuration of the left-chiral and the right-chiral part and the time-dependence must be the same for both parts. As discussed above, this left-chiral and right-chiral parts are really different, because they have different weak charge! Nevertheless, in order to describe a dynamical, physical electron we always need both parts.

Take note that these solutions do not mean that the object we use to describe a physical electron consists of exactly the same upper and lower two-component objects. Only their spin configuration and their time dependence must be equivalent. The upper object is still part of a doublet, whereas the lower object isn't. The upper object transforms under $SU(2)$ transformations and the lower doesn't. Using the notation from above we have

$$\text{physical electron} = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \propto \begin{pmatrix} u \\ u \end{pmatrix}. \tag{8.65}$$

This does not mean that $\chi_L = \xi_R$. The Weyl spinor χ_L is part of a doublet and describes a particle with isospin, whereas the particle described by ξ_R has isospin zero. In addition, we already know that the upper and lower Weyl spinor inside a Dirac spinor transform differently under Lorentz boosts. Therefore it is important that we use different symbols. In other words: The object describing a left-chiral electron χ_L carries an additional $SU(2)$ index, because χ_L transforms as part of a doublet under $SU(2)$ transformations. ξ_R has no such index and transforms as a singlet under $SU(2)$ transformations.

Equivalently we have

$$\text{physical positron} = \begin{pmatrix} -\tilde{\xi}_L \\ \chi_R \end{pmatrix} \propto \begin{pmatrix} -v \\ v \end{pmatrix} \tag{8.66}$$

which we can see through charge conjugation[43]

$$(\text{physical electron})^c = i\gamma_2 \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}^{\star} = \begin{pmatrix} -\tilde{\xi}_L \\ \chi_R \end{pmatrix} \propto i\gamma_2 \begin{pmatrix} u \\ u \end{pmatrix}^{\star} = \begin{pmatrix} -u^c \\ u^c \end{pmatrix}$$

$$= \text{physical positron} \tag{8.67}$$

[43] See Section 7.1.5 and use the explicit form of the matrix as defined in Eq. 6.13: $\gamma_2 = \begin{pmatrix} 0 & \bar{\sigma}_2 \\ \sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix}$.

The message to take away is that the physical electron we observe in nature most of the time is a mixture of two different particles: The

left-chiral electron that carries isospin and the right-chiral electron with isospin zero! Equivalently the physical positron is a mixture of an anti-left-chiral electron, which carries isospin and an anti-right-chiral electron with isospin zero.

The solutions of the Dirac equation tell us how our particles evolve in time. Let's say we start with an electron with spin up that was created in a weak interaction and is therefore purely left-chiral. How does this particle evolve in time? A purely left-chiral electron with spin up

$$
e_L^\uparrow = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{8.68}
$$

is not a solution of the Dirac equation and therefore, in order to determine its time evolution, we must rewrite this in terms of solutions of the Dirac equation.

$$
e_L^\uparrow = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{2} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right) = \Psi_1(t=0) - \tilde{\Psi}_1(t=0). \tag{8.69}
$$

We know how Ψ_1 and $\tilde{\Psi}_1$ evolve in time

$$
\Psi_1(t) - \tilde{\Psi}_1(t) = \rightarrow \frac{1}{2} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{-imt} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{imt} \right) \tag{8.70}
$$

For $t = 0$ this reduces to the left-chiral state as it should be, but as time evolves, say $t = \frac{\pi}{2m}$ we have

$$
\rightarrow \frac{1}{2} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \underbrace{e^{-i\frac{\pi}{2}}}_{=-i} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \underbrace{e^{i\frac{\pi}{2}}}_{=i} \right) = \frac{i}{2} \left(\begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right)
$$

$$
= -i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = -i e_R^\uparrow \tag{8.71}
$$

which describes a right-chiral electron with spin up! The lesson here is that as time evolves, a left-chiral particle changes into a right-chiral

particle and vice versa. To describe the time-evolution of a particle like an electron we need e_L and e_R, which is why we wrote them together in one object: the Dirac spinor. The same is true for the positron.

Recall that the two different particles e_L and e_R, carry different weak charge, i.e. isospin. Nevertheless as time evolves these two particles can transform into each other. Most of the time it will be a mixture of both and not a definite eigenstate. Isospin and chirality are therefore not conserved as time evolves, only in interactions.

We can now see that the notation with Dirac spinors is necessary, because we have a close, dynamical connection between each two particles of the four particles listed at the beginning of this section. Chirality and therefore isospin are not conserved during propagation. A propagating electron can sometimes be found as left-chiral and sometimes as right-chiral.

8.9 Appendix: Solving the Dirac Equation

As explained at the beginning of the last section, we use here the symbols u *and* v *for the two-component objects inside a Dirac spinor, and* u *and* v *for four-component objects. This means, for example* u_1 *and* u_2 *describe two different four-component objects. If we don't want to be specific and want to consider both four-component objects at the same time we simply write* u. *Then* u_1 *and* u_2 *are the two two-component objects inside such a four-component object* $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$

In this appendix we will solve the Dirac equation in the *rest frame* in the *chiral basis*. The solution for an arbitrary frame can be computed by acting with a boost transformation on the solution derived in this section. In addition to the discussion in the last section, we will use these solutions in Chapter 9, when we talk about quantum field theory. The Dirac equation is

$$(i\partial_\mu \gamma^\mu - m)\psi = 0. \tag{8.72}$$

Anticipating plane wave solutions, we make the ansatz $\Psi = u e^{-ipx}$, with some four-component object u, because the matrices γ_μ in the equation are 4×4. In the rest frame, which means momentum zero $\vec{p} = 0$, the exponent reduces to $-ipx = -i(p_0 x_0 - \vec{p}\vec{x}) = -ip_0 x_0$. Now using the relativistic energy-momentum relation $E = \sqrt{\vec{p} + m^2}$, which we derived at the beginning of this chapter, and using that

$p_0 = E$ and $x_0 = t$, we have $-ipx = -iEt = -i\sqrt{\underbrace{\vec{p}}_{=0} + m^2}\, t = -imt$.

Putting this ansatz into the Dirac equation yields

$$(i\partial_\mu \gamma^\mu - m)ue^{-imt} = 0$$
$$\rightarrow (i(\partial_0 \gamma^0 + \partial_i \gamma^i) - m)ue^{-imt} = 0$$
$$\rightarrow i((-im)\gamma^0 - m)ue^{-imt} = 0$$
$$\rightarrow (m\gamma^0 - m)u = 0$$
$$\underset{\text{dividing by } m}{\longrightarrow} \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)c = 0$$
$$\rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0$$
$$\rightarrow \begin{pmatrix} -u_1 + u_2 \\ u_1 - u_2 \end{pmatrix} = 0 \tag{8.73}$$

The 1 inside the remaining matrix here is the 2×2 unit matrix and therefore u_1 and u_2 are two-component objects. We see that our ansatz solves the equation, if $u_1 = u_2$. Therefore, we have found two linearly independent solutions of the Dirac equation

$$\Psi_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{-imt} \quad \Psi_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} e^{-imt} \tag{8.74}$$

We can find two other solutions by making the ansatz $\tilde{\Psi} = ve^{ipx}$, which analogously reduces in the rest frame to $\tilde{\Psi} = ve^{imt}$. This ansatz yields

$$(i\partial_\mu \gamma^\mu - m)ve^{imt} = 0$$
$$\rightarrow (-m\gamma^0 - m)v = 0$$
$$\rightarrow \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0$$
$$\rightarrow \begin{pmatrix} -v_1 - v_2 \\ -v_1 - v_2 \end{pmatrix} = 0. \tag{8.75}$$

We therefore conclude that we have a solution with time dependence e^{imt}, if the upper and lower two-component objects in the Dirac spinor are related by $-v_1 = v_2$. Two linearly independent

solutions following from this ansatz are

$$\tilde{\Psi}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} e^{imt} \quad \tilde{\Psi}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} e^{imt}. \tag{8.76}$$

8.10 Appendix: Dirac Spinors in Different Bases

In the Lagrangian the Dirac spinors ψ appear always in combination with the matrices γ_μ. This can be used to simplify computations, by switching to a different basis. This works, because we can add terms of the form $1 = N^{-1}N$, with some arbitrary invertible matrix N, between ψ and γ_μ and then redefine both. For example

$$\partial_\mu \bar{\psi} \gamma_\mu \psi = \partial_\mu \bar{\psi} \underbrace{N^{-1}N}_{=1} \gamma_\mu \underbrace{N^{-1}N}_{=1} \psi = \partial_\mu \underbrace{\bar{\psi}N^{-1}}_{\equiv\bar{\psi}'} \underbrace{N\gamma_\mu N^{-1}}_{\equiv\gamma'_\mu} \underbrace{N\psi}_{\equiv\psi'} = \partial_\mu \bar{\psi}'\gamma'_\mu\psi'. \tag{8.77}$$

The basis we worked with in this text so far is called the **chiral basis** or **Weyl basis**. Conventionally the Dirac equation is solved in another basis, called **mass basis** or **Dirac basis**. In the chiral basis we worked with so far, the Dirac Lagrangian

$$\mathcal{L}_D = i\chi_L^\dagger \sigma^\mu \partial_\mu \chi_L + i\xi_R^\dagger \bar{\sigma}^\mu \partial_\mu \xi_R - m\chi_L^\dagger \xi_R - m\xi_R^\dagger \chi_L \tag{8.78}$$

has non-diagonal mass terms, i.e. mass terms that mix different states. We can use the freedom to choose a basis to pick a basis where the mass terms are **diagonal**, which is then called mass basis.

This means we want a mass term $\psi^\dagger m\psi$, with $m = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$, which gives us mass terms of the form

$$\bar{\psi}'M'\psi' = \psi^\dagger \gamma'_0 M'\psi' = \begin{pmatrix} u' \\ v' \end{pmatrix}^\dagger \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix} = (u')^\dagger m_1 u' + (v')^\dagger m_2 v', \tag{8.79}$$

whereas at the moment we are dealing with

$$\bar{\psi}M\psi = \psi^\dagger \gamma_0 M\psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}^\dagger \begin{pmatrix} 0 & m \\ m & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = m\chi_L^\dagger \xi_R + m\xi_R^\dagger \chi_L. \tag{8.80}$$

The latter basis, which we worked with so far, makes it easy to interpret things in terms of chirality, whereas it's easier for Dirac spinors in the mass basis to make the connection to **physical** propagating **particles**.

To find the connection between the second and the first form, we need to diagonalize the matrix $M = \begin{pmatrix} 0 & m \\ m & 0 \end{pmatrix} = m \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The matrix is diagonalized through the matrix $N = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$:

$$N^{-1} \underbrace{\begin{pmatrix} -m & 0 \\ 0 & m \end{pmatrix}}_{\equiv M'} N = M \tag{8.81}$$

$$\rightarrow m \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = m \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{8.82}$$

and therefore we redefine the Dirac spinors accordingly

$$\begin{aligned} \bar{\psi} M \psi &= \bar{\psi} \underbrace{N N^{-1}}_{=1} M \underbrace{N N^{-1}}_{=1} \psi \\ &= \underbrace{\bar{\psi} N}_{\equiv \bar{\psi}'} \underbrace{N^{-1} M N}_{\equiv M'} \underbrace{N^{-1} \psi}_{\equiv \psi'} \\ &= \bar{\psi}' M' \psi'. \end{aligned} \tag{8.83}$$

It is instructive to have a look at the chiral projection operators $P_L = \frac{1 - \gamma_5}{2}$ in this basis. We need to calculate how the matrix γ_5 looks like in this new basis

$$\begin{aligned} \tilde{\gamma}_5 = N^{-1} \gamma_5 N &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \end{aligned} \tag{8.84}$$

The corresponding eigenvectors are $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. This means a chiral eigenstate is now described by a Dirac spinor with upper and lower components. For example, a left-chiral state is in this basis of the form $\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. In contrast, in the chiral basis γ_5

was diagonal and a left-chiral eigenstate was given by a Dirac spinor with upper components only $\psi_L = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix}$, and a right-chiral Dirac spinor with lower components only $\psi_R = \begin{pmatrix} 0 \\ \xi_R \end{pmatrix}$.

The chiral projection operator is in this basis

$$P_L = \frac{1 - \gamma_5}{2} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{8.85}$$

8.10.1 Solutions of the Dirac Equation in the Mass Basis

We can solve the Dirac equation in the mass basis

$$\left(i\gamma_\mu \partial^\mu - m\right)\Psi = 0 \tag{8.86}$$

by making the ansatz $\psi = u e^{-ipx}$, which yields

$$\left(\gamma_\mu p^\mu - m\right) u e^{-ipx} = 0$$
$$\rightarrow \left(\left(\gamma_\mu p^\mu - m\right) u = 0\right.. \tag{8.87}$$

Equivalently we can make the ansatz $\psi = v e^{ipx}$, which yields

$$\left(-\gamma_\mu p^\mu - m\right) v e^{ipx} = 0$$
$$\rightarrow \left(-\gamma_\mu p^\mu - m\right) v = 0.$$

Analogously to our solution in the chiral basis, we work in the rest frame, i.e. $\vec{p} = 0$. We are allowed to make such a choice, because physics is the same in all frames of reference and therefore we can pick one that fits our needs best. In this frame of reference, with $p_i = 0$, we have

$$\rightarrow \left(\gamma_0 p^0 - m\right) u = 0$$
$$\rightarrow \left(-\gamma_0 p^0 - m\right) v = 0.$$

In addition, we have $p_0 = E$ and we can use the relativistic energy-momentum relation, which we derived at the beginning of this chapter (Eq. 8.2). In the rest frame we have $E = \sqrt{(p_i)^2 + m^2} = m$. We now use the explicit form of γ_0 in the mass basis, which can be computed using the matrix N from above and the transformation law $\gamma_0' = N^{-1}\gamma_0 N$. Remember that we have an implicit unit matrix behind m and therefore

$$\rightarrow \left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} m - m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) u = 0$$

$$\rightarrow \left(-\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} m - m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) v = 0$$

$$\rightarrow \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} u = 0$$

$$\rightarrow \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} v = 0.$$

Recalling that each Dirac spinor consists of two two-component objects, we conclude that the lower two-component object of u and the upper two-component object of v must be zero:

$$\rightarrow \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -2u_2 \end{pmatrix} = 0 \rightarrow u_2 = 0$$

$$\rightarrow \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -2v_1 \\ 0 \end{pmatrix} = 0 \rightarrow v_1 = 0.$$

We can see that in this basis the physical propagating particles (=solutions of the Dirac equation) are described by spinors with upper components only or equivalently for antiparticles with lower components only. We therefore have again four linearly independent solutions

$$\Psi_1' = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{-imt} \quad \Psi_2' = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} e^{-imt} \tag{8.88}$$

and

$$\tilde{\Psi}_1' = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{imt} \quad \tilde{\Psi}_2' = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} e^{imt}. \tag{8.89}$$

A general solution in this frame, in this basis, is a linear combination

$$\psi = u e^{-ipx} + v e^{ipx} = \begin{pmatrix} u_1 \\ 0 \end{pmatrix} e^{-ipx} + \begin{pmatrix} 0 \\ v_1 \end{pmatrix} e^{ipx} \tag{8.90}$$

and we get the solution in an arbitrary frame by transforming this solution with a Lorentz boost. In addition, the most general solution is a superposition of all possible momenta and spin configurations[44]

[44] Recall that the two components of a Weyl spinor represent different spin states.

$$\Psi = \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3 p}{\sqrt{E_p}} \left(c_r(p) u_r(p) e^{-ipx} + d_r^\dagger(p) v_r(p) e^{+ipx} \right) \tag{8.91}$$

9

Quantum Field Theory

Summary

In this chapter the framework of quantum field theory is introduced. Starting with the equation derived in Chapter 5

$$[\Phi(x), \pi(y)] = i\delta(x - y),$$

we are able to see that the fields themselves are operators. The solutions of the equations of motion for spin 0, $\frac{1}{2}$ and 1 are written in terms of their **Fourier expansions**[1]. Using the commutation relation, cited above, we discover that the Fourier coefficients are now operators. Afterwards, we will see how these **operators**, and with them of course the fields, **create and annihilate particles**. Using the Lagrangian for the corresponding fields, we are able to derive the Hamiltonian operator representing energy.

Afterwards, we start developing **interaction theory**, which is what quantum field theory is all about. We will see that in interaction theory, the Hamiltonian is given by a linear combination of the Hamiltonian for the free field plus an **interaction Hamiltonian**. This insight can then be used in the **interaction picture**, where the time evolution of the fields is governed by the free Hamiltonian and the time evolution of the states by the interaction Hamiltonian. Using this picture, we are able to derive the probability amplitudes for scattering processes. These are denoted in the Dirac notation

$$\langle f | \hat{S} | i \rangle \, ,$$

where $\hat{S}$ denotes the operator describing the scattering process, $|i\rangle$ is the initial state and $\langle f |$ is the final state. We will discover that the operator $\hat{S}$ can be written in terms of the interaction Hamiltonian H_I:

$$\hat{S}(t, t_i) = e^{-i \int_{t_i}^{t} dt' H_I}.$$

[1] The idea behind the Fourier transform is explained in Appendix D.1.

This can't be solved and therefore we evaluate the exponential in terms of its series expansions. For most experiments the first few terms suffice to get an accurate description.

Each term in this series expansion can be interpreted physically as describing a different kind of scattering process. The interaction Hamiltonian contains linear combinations of the fields, which create and annihilate particles as mentioned above. For the first non-trivial order, we get 8 terms and we will see that the first term describes a scattering process of the form $e^- e^+ \to \gamma$. This means we start with an initial state $|e^- e^+\rangle$ consisting of an electron and a positron, which are annihilated by the field operators of the spin $\frac{1}{2}$ fields and afterwards a photon $\langle \gamma |$ is created by the photon field. The other terms result in zero when operating on this initial state $|e^- e^+\rangle$.

The next order in the series expansion consists of many, many terms and we will take a look at just one of them. Again, we start with an initial state $|e^- e^+\rangle$ and we will see that **one** term describes the process $e^- e^+ \to \gamma \to e^- e^+$, where the initial and final electron and positron have, in general, completely different momenta.

In exactly the same way, all terms can be interpreted for all interaction Hamiltonians. A pictorial way to simplify these kinds of computations are the famous Feynman diagrams. Each line and vertex in such a diagram represents a factor of the kind mentioned above.

9.1   Field Theory Identifications

In this section, we want to understand how the Lagrangians that we derived from symmetry constraints can be used in a field theoretical framework. The first step in deriving a field theory describing nature is combining the Lagrangians we found with the result from Chapter 5, specifically Eq. 5.5, which we recite here for convenience:

$$[\Phi(x), \pi(y)] = i\delta(x - y), \tag{9.1}$$

where the conjugate momentum density $\pi(y)$ is given by

$$\pi(y) = \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi(y))}. \tag{9.2}$$

9.2 Free Spin 0 Field Theory

"Every act of creation is first of all an act of destruction."

- Pablo Picasso[2]

Again, let's start with the simplest possible case: free spin 0 fields. We describe such fields with scalars, which are objects that do not change at all under Lorentz transformations[3]. We already derived in Chapter 6.2 the corresponding Lagrangian

[3] This was derived in Section 3.7.4

$$\mathscr{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \tag{9.3}$$

and the equation of motion, called Klein-Gordon equation

$$(\partial_\mu \partial^\mu + m^2)\Phi = 0. \tag{9.4}$$

Using the Lagrangian in Eq. 9.3, we can calculate the conjugate momentum

$$\pi(x) = \frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi(x))} = \frac{\partial}{\partial(\partial_0 \Phi(x))} \frac{1}{2}(\partial_\mu \Phi(x) \partial^\mu \Phi(x) - m^2 \Phi^2(x)) = \partial_0 \Phi(x).$$

The most general solution of the Klein-Gordon equation can be written in terms of its Fourier-expansion[4]

[4] See the appendix in Section 9.6 at the end of this chapter for a detailed computation of the integral measure and a justification for writing the solution like this.

$$\Phi(x) = \int dk^3 \frac{1}{(2\pi)^3 2\omega_k} \left(a(k)e^{-i(kx)} + b(k)e^{i(kx)} \right), \tag{9.5}$$

with $(\omega_k)^2 \equiv \vec{k}^2 + m^2$. For **real** scalar fields we can rewrite this as

$$\Phi(x) = \int dk^3 \frac{1}{(2\pi)^3 2\omega_k} \left(a(k)e^{-i(kx)} + a^\dagger(k)e^{i(kx)} \right) \tag{9.6}$$

because $c + c^\dagger = \underbrace{\text{Re}(c) + i \cdot \text{Im}(c)}_{c} + \underbrace{\text{Re}(c) - i \cdot \text{Im}(c)}_{c^\dagger} = 2\text{Re}(c).$

Now we are having a look at the implications of Eq. 9.1, i.e. the non vanishing commutator $[\Phi(x), \pi(y)] \neq 0$. This means that $\Phi(x)$ and $\pi(y)$ cannot be ordinary functions, but must be **operators**, because ordinary functions commute: $(3+x)(7xy) = (7xy)(3+x)$. By looking at Eq. 9.6, we conclude that the Fourier-coefficients $a(k)$ and $a(k)^\dagger$ are operators, because $e^{\pm i(kx)}$ is just a complex number and complex numbers commute.

Using Eq. 9.1 we can compute[5]

$$[a(k), a^\dagger(k')] = (2\pi)^3 \delta^3(\vec{k} - \vec{k}') \tag{9.7}$$

and

$$[a(k), a(k')] = 0 \tag{9.8}$$

$$[a^\dagger(k), a^\dagger(k')] = 0. \tag{9.9}$$

Now that we know that our field itself is an operator, the logical next thing to ask is: **What does it operate on?** In a particle theory, we identify the dynamical variables as operators acting on something describing a particle (the wavefunction, an abstract Dirac vector, etc.). In a field theory we have up to now nothing to describe a particle. At this point, it is completely unclear how particles appear in a field theory. Nevertheless, let's have a look at how our field coefficents $a(k)$ and $a^\dagger(k)$ act on something abstract and by doing this, we learn how the fields act on something abstract. To get some intuition about what is going on here, let's first have a look at something we are familiar with: energy.

The energy E of a scalar field is given by[6]

$$E = \int d^3x\, T^{00}$$

$$= \int d^3x \left(\frac{\partial \mathscr{L}}{\partial(\partial_0 \Phi)} \underbrace{\frac{\partial \Phi}{\partial x_0}}_{=\partial_0 \Phi} - \mathscr{L} \right)$$

$$= \int d^3x \left((\partial_0 \Phi)^2 - \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \right)$$

$$\underbrace{=}_{\partial_\mu \partial^\mu = \partial_0 \partial_0 - \partial_i \partial_i} \frac{1}{2} \int d^3x \left((\partial_0 \Phi)^2 + (\partial_i \Phi)^2 + m^2 \Phi^2 \right) \tag{9.10}$$

By substituting Eq. 9.6 into Eq. 9.10 and using the commutation relations (Eq. 9.7-9.9), we can write

$$E = \frac{1}{2} \int dk^3 \frac{1}{(2\pi)^3} \omega_k \left(a^\dagger(k)a(k) + a(k)a^\dagger(k) \right)$$

$$\underbrace{=}_{\text{Eq. 9.7}} \int dk^3 \frac{1}{(2\pi)^3} \omega_k \left(a^\dagger(k)a(k) + \frac{1}{2}(2\pi)^3 \delta^3(0) \right) \tag{9.11}$$

At this point we can see that our theory explodes. The second term in the integral is infinite. We could stop at this point and say that this kind of theory does not work. Nevertheless, some brave physicists dug deeper, ignoring this infinite term and discovered a theory describing nature very accurately. There is no explanation for this and the standard way of continuing from here on is to ignore the second term. The crux here is that this term appears in the energy of every system and we are only able to measure energy differences. Therefore, this constant infinite term appears in none of our measurements.

Conventionally, the energy written as an operator is called Hamiltonian $\hat{H}$. We can compute the commutator[7] of $\hat{H}$ and the Fourier coefficents $a(k)$ and $a^\dagger(k)$. We get

$$[\hat{H}, a^\dagger(k')] = \int dk^3 \frac{1}{(2\pi)^3} \omega_k [a^\dagger(k)a(k), a^\dagger(k')]$$

$$\underbrace{=}_{[a^\dagger(k), a^\dagger(k')]=0} \int dk^3 \frac{1}{(2\pi)^3} \omega_k a^\dagger(k) [a(k), a^\dagger(k')]$$

$$= \int dk^3 \omega_k a^\dagger(k) \delta^3(k - k')$$

$$\underbrace{=}_{\text{See Appendix D.2}} \omega_{k'} a^\dagger(k')$$

$$\tag{9.12}$$

and equally

$$[\hat{H}, a(k')] = -\omega_{k'} a(k') . \tag{9.13}$$

The quantum formalism works by operating with operators on something that describes the physical system[8]. Here, if we act with the energy operator, i.e. the Hamiltonian $\hat{H}$, on something abstract $|?\rangle$ describing our physical system, we get the energy of the system:

$$\hat{H} |?\rangle = E |?\rangle \tag{9.14}$$

Now we return to our starting question: **How does a field[9] act on our system**? Let's have a look at the effect of the first Fourier coefficient, now an operator, on the energy E of the system[10]:

$$\hat{H}\left(a(k')|?\rangle\right) = \left(a(k')\hat{H} + \underbrace{\hat{H}a(k') - a(k')\hat{H}}_{[\hat{H}, a(k')]}\right)|?\rangle$$

$$= a(k')\underbrace{\hat{H}|?\rangle}_{=E|?\rangle} + [\hat{H}, a(k')]|?\rangle$$

$$= \left(a(k')E + [\hat{H}, a(k')]\right)|?\rangle$$

$$\underbrace{=}_{\text{Eq. 9.13}} \left(a(k')E - \omega_{k'}a(k')\right)|?\rangle$$

$$= \left(E - \omega_{k'}\right)\left(a(k')|?\rangle\right) \tag{9.15}$$

and equally for the second Fourier coefficent

$$\hat{H}a^\dagger(k')|?\rangle = (E + \omega_{k'})a^\dagger(k')|?\rangle . \tag{9.16}$$

How can we interpret this? We see that $a(k')|?\rangle$ can be interpreted as a new system with energy $E - \omega_k$. To make this more concrete we define

$$|?_2\rangle \equiv a(k')|?\rangle$$

with

$$\hat{H} \ket{?_2} \underbrace{=}_{\text{Using Eq. 9.15}} (E - \omega_{k'}) \ket{?_2} \,.$$

This suggests how we should interpret what the field does. Imagine a completely empty system $\ket{0}$, with, by definition $H \ket{0} = 0 \ket{0}$. If we now act with $a^\dagger(k')$ on $\ket{0}$, we know that this transforms our empty system into a system having energy $\omega_{k'}$

$$\hat{H} a^\dagger(k') \ket{0} \underbrace{=}_{\text{Using Eq. 9.16}} \omega_{k'} a^\dagger(k') \ket{0} \,. \tag{9.17}$$

We see that $a^\dagger(k')$ creates something in the completely empty system with energy $\omega_{k'}$, **which is what we call a particle with momentum k'**! If we act a second time on this system with $a^\dagger(k')$ we create a second particle with the same momentum. If we act on it with $a^\dagger(k'')$ we create a particle with momentum k'' and so on. Therefore, we call $a^\dagger(k')$ a **creation operator**. Similarly to $a^\dagger(k')$, we can interpret $a(k')$: $a(k')$ destroys or annihilates a particle of energy $\omega_{k'}$ and is therefore called **annihilation** or **destruction operator**. To make this more concrete we introduce a new notation for particle states

$$a^\dagger(k) \ket{0} \equiv \ket{1_k} \tag{9.18}$$

$$a^\dagger(k) \ket{1_k} \equiv \ket{2_k} \tag{9.19}$$

$$a^\dagger(k') \ket{2_k} \equiv \ket{2_k, 1_{k'}} \tag{9.20}$$

Take a look at the energy again:

$$E = \int dk^3 \frac{1}{(2\pi)^3} \omega_k a^\dagger(k) a(k) \,.$$

What happens if this operator acts on a state like $\ket{2k_1, k_2}$? The result should be

$$E = 2\omega_{k_1} + \omega_{k_2},$$

which is the energy of two particles with energy ω_{k_1} and one particle with energy ω_{k_2}. Therefore, the operator

$$N(k) \equiv a^\dagger(k) a(k) \tag{9.21}$$

appearing here is a number operator, denoted $N(k)$ that extracts the number of particles with momentum k from a state:

$$N(k) \ket{n_k, n'_{k'}, \ldots} = n_k \ket{n_k, n'_{k'}, \ldots} \tag{9.22}$$

The energy operator can then be written as

$$E = \int dk^3 \frac{1}{(2\pi)^3} \omega_k N(k).$$

Furthermore, take note that there are physical systems, where the momentum spectrum is not continuous, but discrete[11]. For such systems all integrals change to sums, for example, the energy is then of the form

$$E = \sum_k \omega_k N(k)$$

and the commutation relation changes to

$$[a(k), a^\dagger(k')] = \delta_{k,k'}. \tag{9.23}$$

Take note that quantum field theory is, like quantum mechanics, a theory making probabilistic predictions. Therefore, our states need to be normalized $\langle k, k', .. | k, k', .. \rangle \overset{!}{=} 1$, because a probability of more than $100\% = 1$ doesn't make sense. If we act with an operator like $a(k)$ on a ket, the new ket does not necessarily have unit norm[12]. Therefore, we write

$$a^\dagger(k) |n_k\rangle = C |n_k + 1\rangle, \tag{9.24}$$

where n_k denotes the number of particles with momentum k and C is some number. From this we get[13]

$$(a^\dagger(k) |n_k\rangle)^\dagger = (C |n_k + 1\rangle)^\dagger$$

$$\rightarrow \langle n_k | a(k) = \langle n_k + 1 | C^\dagger. \tag{9.25}$$

We can therefore write

$$\underbrace{\langle n_k | a(k)}_{9.25} \underbrace{a^\dagger(k) |n_k\rangle}_{9.24} = \langle n_k + 1 | \underbrace{C^\dagger C}_{\text{a number and no operator}} |n_k + 1\rangle = C^\dagger C \underbrace{\langle n_k + 1 | n_k + 1\rangle}_{=1}$$

$$\tag{9.26}$$

or using the discrete commutation relation (Eq. 9.23)

$$\langle n_k | a(k) a^\dagger(k) |n_k\rangle = \langle n_k | \left(\underbrace{a^\dagger(k) a(k)}_{=N(k) \text{ Eq. 9.21}} + \underbrace{\delta_{k,k}}_{=1} \right) |n_k\rangle$$

$$\overset{\text{Eq. 9.22}}{=} \langle n_k | \underbrace{(n_k + 1)}_{\text{a number and no operator}} |n_k\rangle = (n_k + 1) \underbrace{\langle n_k | n_k\rangle}_{=1}. \tag{9.27}$$

Putting Eq. 9.26 and Eq. 9.27 together yields

$$C^\dagger C = n_k + 1 \rightarrow C = \sqrt{n_k + 1} \tag{9.28}$$

and we therefore have

$$a^\dagger(k) |n_k\rangle = \sqrt{n_k + 1} |n_k + 1\rangle. \tag{9.29}$$

Following the same steps we can derive

$$a(k) |n_k\rangle = \sqrt{n_k} |n_k - 1\rangle. \tag{9.30}$$

Two questions may pop up at this point.

[11] Remember the particle in a box example. It is an often used trick in quantum field theory to assume the system in question is restricted to a volume V. This results in a discrete momentum spectrum. At the end of the computation the limit $\lim_{V \to \infty}$ is taken.

[12] Exactly the same line of though was discussed in Section 3.6.1, where we derived the finite-dimensional $\mathfrak{su}(2)$ representations. There, we also introduced "ladder operators" $J_\pm$ and derived the constants that we emerge when we act with these operators on a given vector.

[13] Recall that $|n_k\rangle^\dagger = \langle n_k|$ and we have of course $(a^\dagger)^\dagger = a$.

1. What happens if we want to annihilate a particle in a completely empty system?

2. What about energy or charge conservation? How can we create something from nothing without violating conservation laws?

The conservation laws are, of course, never violated, but how this comes about will only become clear when we develop the theory further. Maybe it helps to see that at this point Richard Feynman had the same problem[14]

> I remember that when someone had started to teach me about creation and annihilation operators, that this operator creates an electron, I said, "how do you create an electron? It disagrees with the conservation of charge", and in that way, I blocked my mind from learning a very practical scheme of calculation.

Secondly, we are never able to destroy something which is not there in the first place. If we act with the destruction operator $a(k)$ on a completely empty state $|0\rangle$ we get, using Eq. 9.30

$$a(k)\,|0_k\rangle = \sqrt{0}\,|0_k - 1_k\rangle = 0 \tag{9.31}$$

or equally

$$a(k')\,|1_k\rangle = \sqrt{0}\,|1_k, 0_{k'} - 1k'\rangle = 0. \tag{9.32}$$

We can see that if we act with an annihilation operator $a(k')$ on a ket, like $|k\rangle$ that does not include a particle with this momentum k', the theory produces a zero. The creation and annihilation operators appear in the Fourier expansion of the fields, which includes an integral (or sum) over all possible momenta. Therefore, if these fields act on a ket like $|k\rangle$, only one annihilation operator will result in something non-zero. This will be of great importance when we try to describe interactions using quantum field theory.

Before we move on to interactions, we take a brief look at free spin $\frac{1}{2}$ and spin 1 fields.

9.3   Free Spin $\frac{1}{2}$ Field Theory

The equation of motion for free spin $\frac{1}{2}$ fields is the Dirac equation[15]

$$(i\gamma_\mu \partial^\mu - m)\Psi = 0.$$

The general solution of the Dirac equation can be written in the form[16]

$$\Psi = \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{w_p}} \left(c_r(p) u_r(p) e^{-ipx} + d_r(p) v_r(p) e^{+ipx} \right)$$
$$= \Psi^+ + \Psi^-, \tag{9.33}$$

but this time we do not restrict to real fields, because we saw in Section 6.3 that a Lorentz invariant Lagrangian needs complex spin $\frac{1}{2}$ fields. In addition, we follow the standard convention and write the solutions as

$$\Psi = \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{w_p}} \left(c_r(p) u_r(p) e^{-ipx} + d_r^\dagger(p) v_r(p) e^{+ipx} \right), \tag{9.34}$$

because in this way, $d_r^\dagger(p)$ can be seen to create an anti-particle. If we would name it $d_r(p)$ in the solution, this could lead to confusion, because for particles $c_r^\dagger(p)$ creates and $c_r(p)$ annihilates. Naming the Fourier coefficient $d_r^\dagger(p)$ instead of $d_r(p)$, leads to an analogous interpretation for anti-particles: $d_r^\dagger(p)$ creates and $d_r(p)$ annihilates.

Analogously, we have for the adjoint Dirac equation

$$(i\partial^\mu \bar\Psi \gamma_\mu + m\bar\Psi) = 0,$$

which was also derived in Section 6.3, the solution

$$\bar\Psi = \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{w_p}} \left(c_r^\dagger(p) \bar u_r(p) e^{+ipx} + d_r(p) \bar v_r(p) e^{-ipx} \right)$$
$$\equiv \bar\Psi^+ + \bar\Psi^-. \tag{9.35}$$

In these solutions u_1, u_2, v_1, v_3 denote the "basis spinors" in an arbitrary frame[17]

[17] We derived these spinors in the rest frame ($p_i = 0$) in Section 8.10. The basis spinors in an arbitrary frame can be computed from the basis spinors in the rest frame by a boost transformation.

$$u_1 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} 1 \\ 0 \\ \frac{p_3}{E+m} \\ \frac{p_1+ip_2}{E+m} \end{pmatrix} \qquad u_2 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} 0 \\ 1 \\ \frac{p_1-ip_2}{E+m} \\ \frac{-p_3}{E+m} \end{pmatrix} \tag{9.36}$$

$$v_1 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \frac{p_1-ip_2}{E+m} \\ \frac{-p_3}{E+m} \\ 0 \\ 1 \end{pmatrix} \qquad v_2 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \frac{p_3}{E+m} \\ \frac{p_1+ip_2}{E+m} \\ 1 \\ 0 \end{pmatrix}. \tag{9.37}$$

The rest of the theory for free spin $\frac{1}{2}$ fields can be developed similarly to the scalar theory, but there is one small difference[18]. Let's try

[18] With incredible huge consequences! In fact, nothing in this universe would be stable if the spin $\frac{1}{2}$ theory would work exactly like the scalar theory.

what we get when we use the formula

$$[\Psi(x), \pi(y)] = i\delta(x - y) \tag{9.38}$$

for spin $\frac{1}{2}$ fields, too. In the general solution for the equation of motion, i.e. the Dirac equation, we have two different coefficients: $c^\dagger$ creates particles, whereas $d^\dagger$ creates anti-particles. When we now compute the Hamiltonian for a spin $\frac{1}{2}$ field, using the *commutation* relation[19], we get something of the form

$$H \sim \int c^\dagger c - d^\dagger d \,.$$

This shows that the energy of anti-particles is then *negative*. This is a serious problem because every state could forever decay to lower energy states and nothing in this universe would be stable. The Hamiltonian is not bounded from below, in contrast to the scalar case.

However, when we use instead the *anticommutation* relation

$$\{\Psi(x), \pi(y)\} = \Psi(x)\pi(y) + \pi(y)\Psi(x) = -i\delta(x - y),$$

the term involving the anti-particle creation and annihilation operators comes out positive and the problem is therefore solved.

This has very interesting consequences. For example, we then have[20]

$$\{c^\dagger(p), c^\dagger(p')\} = 0$$

from which we can conclude

$$\{c^\dagger(p), c^\dagger(p)\} = c^\dagger(p)c^\dagger(p) + c^\dagger(p)c^\dagger(p) = 2c^\dagger(p)c^\dagger(p) = 0.$$

$$\Rightarrow c^\dagger(p)c^\dagger(p) = 0 \tag{9.39}$$

The action of two equal creation operators[21] always results in a zero! This means we cannot create two equal spin $\frac{1}{2}$ particles and this is famously called **Pauli-exclusion principle**.

The other anti-commutation relations for the Fourier coefficients can be derived from the anti-commutator of the field and the conjugate momentum to be:

$$\{c_r(p), c_s^\dagger(p')\} = \delta_{rs}\delta(p - p') \qquad \{d_r(p), d_s^\dagger(p')\} = \delta_{rs}\delta(p - p') \tag{9.40}$$

and all other possible combinations equal zero. Therefore, these coefficients can be seen to have the same properties as those we derived in the last section for the spin 0 field. They create and destroy, with

[19] $[\Psi(x), \pi(y)]$
$= \Psi(x)\pi(y) - \pi(y)\Psi(x) = i\delta(x - y)$

[20] Analogous to Eq. 9.9 for scalars, but now with the anticommutator instead of the commutator $[,] \rightarrow \{,\}$.

[21] For spin 0 fields the corresponding equation isn't very surprising, because there we use the commutator: $[c^\dagger(p), c^\dagger(p)] = c^\dagger(p)c^\dagger(p) - c^\dagger(p)c^\dagger(p) = 0$.

the difference that acting twice with the same operator on a state results in a zero. The Hamiltonian for a free spin $\frac{1}{2}$ field can be derived analogous to the Hamiltonian for a free spin 0 field

$$H^{\frac{1}{2}}_{\text{free}} = \int d^3x (-i\bar{\Psi}\gamma_i \partial^i \Psi - m\bar{\Psi}\Psi) \qquad (9.41)$$

or expressed in terms of the Fourier coefficients

$$\mathscr{H}^{\frac{1}{2}}_{\text{free}} = \sum_r \int d^3p \; w_p \left(c_r^\dagger(p)c_r(p) + d_r^\dagger(p)d_r(p) + \text{const} \right), \qquad (9.42)$$

where again the constant term leads to an infinite term and we choose to ignore this.

9.4 Free Spin 1 Field Theory

The solution for the equation of motion for free spin 1 fields, the Proca equation[22]

$$m^2 A^\rho = \frac{1}{2}\partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma) \qquad (9.43)$$

is, analogous to the spin 0 field solution, of the form

$$A_\mu = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \left(\epsilon_{r,\mu}(k)a_r(k)e^{-ikx} + \epsilon_{r,\mu}(k)a_r^\dagger(k)e^{ikx} \right)$$
$$\equiv A_\mu^+ + A_\mu^- \qquad (9.44)$$

where $\epsilon_{r,\mu}(k)$ are **basis vectors**, called **polarization vectors**. For spin 1 fields we are again able to use the commutator instead of the anti-commutator. We are therefore able to derive that the coefficients $a_r, a_r^\dagger$ have the same properties as the coefficients for spin 0 fields.

9.5 Interacting Field Theory

The next step is to look at interactions between fields of different spin. The corresponding Lagrangians were derived in earlier chapters from Lorentz and gauge symmetry. For example, for the interaction between a massive spin $\frac{1}{2}$ field and one massless spin 1 field, we have the Lagrangian (Eq. 7.17)

$$\mathscr{L}_{\text{Dirac+Extra-Term}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu \partial^\mu \Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi \qquad (9.45)$$

from which we derive the corresponding Hamiltonian[23]

[22] We derived the Proca equation in Section 6.4.

[23] We will use this Hamiltonian in a moment!

$$H = \int d^3x \, T^{00}$$

$$= \int d^3x \left(\underbrace{\frac{\partial \mathcal{L}}{\partial(\partial_0 \Psi)}}_{=i\bar{\Psi}\gamma_0} \partial_0 \Psi - \mathcal{L} \right)$$

$$= \int d^3x \left(i\bar{\Psi}\gamma_0\partial_0\Psi + m\bar{\Psi}\Psi - i\bar{\Psi} \underbrace{\gamma_\mu \partial^\mu}_{=\gamma_0\partial_0 - \gamma_i\partial_i} \Psi - gA_\mu\bar{\Psi}\gamma^\mu\Psi \right)$$

$$= \underbrace{\int d^3x (m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_i\partial_i\Psi)}_{=H^{\frac{1}{2}}_{\text{free}}} - \underbrace{\int d^3x \left(gA_\mu\bar{\Psi}\gamma^\mu\Psi \right)}_{\equiv -H_I}$$

$$= H^{\frac{1}{2}}_{\text{free}} + H_I \tag{9.46}$$

9.5.1   Scatter Amplitudes

One of the main goals in quantum field theory is to compute the probability for a fixed number n of particles with defined[24] momenta $p_1, p_2, .., p_n$ to transform into a (possibly different) number of (possibly different) particles n' with momenta $q_1, q_2, .., q_{n'}$. We can write this using Dirac's notation

$$\langle q_1, q_2, .., q_{n'} | \hat{S} | p_1, p_2, .., p_n \rangle \, , \tag{9.47}$$

where $\hat{S}$ is the operator describing the scattering. We will derive how this operator looks concretely in the next section. Formulated differently, we have some particles at initial time t_i at some points in space x_i and after the interaction, i.e. at the final time t_f, (possibly different) particles at a possibly different location x_f.

Therefore, the first question we have to answer is: What can we say about the time evolution of states in quantum field theory?

9.5.2   Time Evolution of States

To answer this question we observe that the energy operator is given on the one hand by the identification with the generator of time-translations $i\partial_0$ and on the other hand by the Hamiltonian. For example, for the free spin 0 theory we derived (Eq. 9.10)

$$H^0_{\text{free}} = \frac{1}{2} \int d^3x \left((\partial_0\Phi)^2 + (\partial_i\Phi)^2 + m\Phi^2 \right) , \tag{9.48}$$

while for the free spin $\frac{1}{2}$ theory we have (Eq. 9.41)

$$H^{\frac{1}{2}}_{\text{free}} = \int d^3x (m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_i\partial^i\Psi). \tag{9.49}$$

Both identifications are operators in a field theory, representing energy, and we therefore write

$$i\partial_0 \ket{?(t)} = H \ket{?(t)}, \tag{9.50}$$

which is the equation governing the time evolution of a state in quantum field theory. We can use this equation to define a time-evolution operator U that transforms the state from one point in time to another. If we choose, for brevity, the start time $t = 0$ we have[25]

$$\ket{?(t)} = U(t) \ket{?(0)}. \tag{9.51}$$

Putting this ansatz into Eq. 9.50 yields

$$i\partial_0 U(t) \ket{?(0)} = HU(t) \ket{?(0)}. \tag{9.52}$$

This equation holds for arbitrary $\ket{?(0)}$ and therefore

$$i\partial_0 U(t) = HU(t). \tag{9.53}$$

The general solution of this equation is

$$U(t) = e^{-i \int_0^t dx_0 H} \tag{9.54}$$

as we can check

$$i\partial_0 U(t) = HU(t) \rightarrow i\partial_0 e^{-i \int_0^t dx_0 H} = He^{-i \int_0^t dx_0 H}$$
$$\rightarrow i(-iH)e^{-i \int_0^t dx_0 H} = He^{-i \int_0^t dx_0 H}$$
$$\rightarrow He^{-i \int_0^t dx_0 H} = He^{-i \int_0^t dx_0 H} \quad \checkmark \tag{9.55}$$

In experiments we never measure a ket $\ket{?(t)}$, but always the combinations of a bra with a ket. In general, we have objects of the form

$$\bra{f(t)} \hat{O} \ket{i(t)}, \tag{9.56}$$

with some operator $\hat{O}$, initial state $\ket{i(t)}$ and final state $\bra{f(t)}$. Here the states evolve in time, as described by the operator U, and the operators are time independent. This somewhat arbitrary choice is called Schrödinger picture and in the following we will talk about different possible choices, i.e. different "pictures".

Writing Eq. 9.56 in terms of the time-evolution operator leads us to[26]

$$\bra{f(0)} U^\dagger(t)\hat{O}U(t) \ket{i(0)}. \tag{9.57}$$

An important idea is that we can switch our perspective and say the operator $\hat{O}$ evolves in time, according to the rule $U^\dagger(t)\hat{O}U(t)$ and

[25] In general $\ket{?(t)} = U(t - t') \ket{?(t')}$.

[26] Remember $\bra{f} = \ket{f}^\dagger$

the bra and ket are time-independent. This new perspective is called **Heisenberg picture**.

There is a similar, very useful trick that is helpful in interaction theory. The Hamiltonian is always the sum of the free Hamiltonian and the interaction Hamiltonian (Eq. 9.46)

$$H = H_{\text{free}} + H_I. \tag{9.58}$$

The trick is to use a mixture of the two perspectives introduced above. We let the states evolve according to H_I and the operators[27] according to H_{free}. This is incredibly useful, because then we can reuse all results we already derived for free fields. This type of perspective is called **interaction picture**. We define a state in the interaction picture

$$|i(t)\rangle_I \equiv U_{\text{free}}^\dagger |i(t)\rangle_S, \tag{9.59}$$

where $U_{\text{free}} = \mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}}$ and the index S stands for **Schrödinger picture**, which is the name for the standard perspective where the states evolve according to the full Hamiltonian and the operators are time-independent.

Putting now Eq. 9.59 into Eq. 9.50 we get[28] the time-evolution equation in the interaction picture:

$$i\partial_0 |i(t)\rangle_S = H |i(t)\rangle_S$$

$$\rightarrow i\partial_0 U_{\text{free}} |i(t)\rangle_I = H U_{\text{free}} |i(t)\rangle_I$$

$$\rightarrow i\partial_0 \mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I = (H_{\text{free}} + H_I)\mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I$$

$$\underbrace{\rightarrow}_{\text{product rule}} H_{\text{free}}\cancel{\mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I} + i\mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}}\partial_0 |i(t)\rangle_I = (\cancel{H_{\text{free}}} + H_I)\mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I$$

$$\rightarrow i\mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}}\partial_0 |i(t)\rangle_I = H_I \mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I$$

$$\rightarrow i\partial_0 |i(t)\rangle_I = \underbrace{\mathrm{e}^{i \int_0^t dx_0 H_{\text{free}}} H_I \mathrm{e}^{-i \int_0^t dx_0 H_{\text{free}}}}_{=H_I^{\text{int}} = \text{the interaction Hamiltonian in the interaction picture.}} |i(t)\rangle_I$$

$$\rightarrow i\partial_0 |i(t)\rangle_I = H_I^{\text{int}} |i(t)\rangle_I \tag{9.60}$$

We conclude that the time-evolution of the states is now indeed governed by the interaction Hamiltonian H_I^{int}. This little equation will be incredibly important for everything that follows.

Now we are able to return to the starting question: How can we compute scattering processes

$$\langle f(t_f)| \hat{S}(t_f, t_i) |i(t_i)\rangle \quad ? \tag{9.61}$$

---

[27] Recall that in quantum field theory the fields are our operators.

[28] The equation holds in this form for the Schrödinger picture. Therefore, we have to solve $|i(t)\rangle_I \equiv U_{\text{free}}^\dagger |i(t)\rangle_S$ for $|i(t)\rangle_S$. We multiply the equation with U_{free} and use $U_{\text{free}}U_{\text{free}}^\dagger = 1$. This yields $|i(t)\rangle_S = U_{\text{free}} |i(t)\rangle_I$, which we can then put into Eq. 9.50.

The operator $\hat{S}(t_f, t_i)$ transforms the initial state $|i(t_i)\rangle$ at time t_i into the final state $|\Psi(t_f)\rangle$ at time (t_f). In general, the final state is not one specific particle state, but a linear combination of many possible outcomes. If this would be not the case particle physics would be boring. An interaction would always result in one specific outcome. We are more general here and allow arbitrary linear combinations. After we specify the operator $\hat{S}$ we will see that this is actually the case.

Concretely this means that $\hat{S}$ transforms an initial state into a linear combination of final states. To avoid confusion we will call the final time in the following simply t instead of t_f. Then we have[29]

$$\hat{S}(t, t_i)\,|i(t_i)\rangle = |\Psi(t)\rangle = \sum_f \underbrace{S_{fi}}_{\text{complex numbers}} |f(t)\rangle. \qquad (9.62)$$

The multiplication with one specific $\langle f'(t)|$ from the left-hand side terminates all but one term of the sum:

$$\langle f'(t)| \sum_f S_{fi}\,|f(t)\rangle = \sum_f S_{fi} \underbrace{\langle f'(t)|f(t)\rangle}_{=\delta_{ff'}\ \text{because basis states are orthogonal}} = \sum_f S_{fi}\delta_{ff'} = S_{f'i}. \qquad (9.63)$$

Therefore, the probability for this process to happen is $|S_{f'i}|^2$.

Now we specify the scatter operator $\hat{S}$. For this purpose we take a look again at the time-evolution equation that we derived above[30]

$$i\partial_t\,|\Psi(t)\rangle_I = H_I\,|\Psi(t)\rangle_I. \qquad (9.64)$$

We can rewrite this in terms of our initial state and the operator $\hat{S}$ using Eq. 9.62

$$\rightarrow i\partial_t\hat{S}(t, t_i)\,|i(t_i)\rangle_I = H_I\hat{S}\,|i(t_i)\rangle_I$$

$$\underbrace{\rightarrow}_{\text{product rule}} i(\partial_t\hat{S}(t, t_i))\,|i(t_i)\rangle_I + i\hat{S}(t, t_i)\underbrace{\partial_t\,|i(t_i)\rangle_I}_{=0\ \text{because}\ |i(t_i)\rangle_I\ \text{does not depend on}\ t} = H_I\hat{S}\,|i(t_i)\rangle_I.$$

Now using that this equation holds for arbitrary initial states we can write

$$i\partial_t\hat{S}(t, t_i) = H_I\hat{S} \qquad (9.65)$$

with the general solution[31]

$$\hat{S}(t, t_i) = e^{-i\int_{t_i}^{t} dt'\, H_I} \qquad (9.66)$$

At a first glance the problem is easy now. We know H_I and "just have to solve the integral". This gives the operator $\hat{S}(t, t_i)$ and by acting with it on our initial states we get the probability amplitudes for different processes. Unfortunately, the integral cannot be solved

and we therefore have to use an approximation method. To simplify the problem mathematically, the initial time is taken to be $t_i = -\infty$ and the final time $t = \infty$. By doing this we avoid our probability amplitudes depending on time. For example, if we scatter particles, the probability that a certain process has happened 10^{-24} seconds after the assumed "collision" is different than $2 \cdot 10^{-24}$ seconds after the "collision", because the interaction is still happening! By taking the time values to be very large we avoid these kind of complications. What we measure is usually always a result after all the interactions have happened.

9.5.3 Dyson Series

[32] This is derived in Appendix B.4.1:
$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$

Because there is no analytic solution for the exponentiated integral, we expand it in a Taylor series[32]

$$
\begin{aligned}
\hat{S}(\infty, -\infty) &= e^{-i \int_{-\infty}^{\infty} dt' H_I(t')} \\
&= 1 - i \int_{-\infty}^{\infty} dt_1 H_I(t_1) \\
&\quad - \frac{1}{2!} \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) + \dots
\end{aligned} \tag{9.67}
$$

This is called the **Dyson series**. We need to take a careful look at the third term. $H_I(t_1)$ and $H_I(t_2)$ are not just numeric values, but operators acting on a ket to their right. Therefore, we need to make sure that the earlier time operator acts on the ket before the later time operator. The operators need to operate on the ket in a time ordered manner. It makes no sense if we act on a state with $H_I(t = 5 \text{ s})$ and after that with $H_I(t = 2 \text{ s})$. In the series above, we need for $t_1 < t_2$

$$
\left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right)
$$

and for $t_2 < t_1$

$$
\left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right)
$$

For this purpose an abstract time-ordering operator T is introduced, which is defined by

$$
T\{A(x)B(y)\} := \begin{array}{ll} A(t_1)B(t_2) & \text{if} \quad t_1 > t_2 \\ B(t_2)A(t_1) & \text{if} \quad t_1 < t_2. \end{array} \tag{9.68}
$$

Therefore we write, giving a name to each term of the series expansion,

$$\hat{S}(\infty, -\infty) = \underbrace{1}_{S^{(0)}} \underbrace{-i \int_{-\infty}^{\infty} dt_1 H_I(t_1)}_{S^{(1)}}$$

$$\underbrace{-\frac{1}{2!} T\left\{ \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) \right\}}_{S^{(2)}} + ... \quad (9.69)$$

or written as a sum

$$\hat{S}(\infty, -\infty) = \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} T\left\{ \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) ... \left(\int_{-\infty}^{\infty} dt_n H_I(t_n) \right) \right\}$$

$$= \sum_{n=0}^{\infty} S^{(n)}. \quad (9.70)$$

This is useful because H_I has a numerical factor[33] in it, the coupling constant of the corresponding interaction, i.e. $H_I \propto g$. This coupling constant is, for example, for electromagnetic interactions, smaller than one. Therefore, the second term in the expansion $S^{(2)} \propto (H_I)^2 \propto g^2$ contributes less than the first term $S^{(1)} \propto g$. The higher order terms in the expansion contribute even less: $g > g^2 > g^3 > ...$. To describe the system in question it often suffices to evaluate the first few terms of the series expansion. Higher order terms often deliver corrections that lie outside the possibility of measurement.

Unfortunately, going further from this point needs many pages of heavy algebra. The first step is Wick's-Theorem which enables one to express the time ordering in terms of commutators. Furthermore, these commutators need to be computed, which results in the famous Feynman propagators. Nevertheless, we want to go further[34], so we are going to use these results without proofs. The interested reader is referred to the standard texts on quantum field theory[35].

[33] See Eq. 9.46.

[34] Because now, the fun is about to begin!

[35] Some recommended books will be listed in the last section of this chapter.

9.5.4 Evaluating the Series

We will now return to the example introduced at the beginning of this chapter: The interaction between a massive spin $\frac{1}{2}$ field and a massless spin 0 field. The corresponding interaction Hamiltonian is (Eq. 9.46)

$$H_I = - \int d^3x g A_\mu \bar{\Psi} \gamma^\mu \Psi.$$

As explained above, we will look at each term of the series in Eq. 9.70 individually.

The first term of the series is trivial as it is simply the identity operator

$$S^{(0)} = I \tag{9.71}$$

The second term is more exciting:

$$S^{(1)} = -i \int_{-\infty}^{\infty} dt H_I = ig \int_{-\infty}^{\infty} d^4x A_\mu \bar{\Psi} \gamma^\mu \Psi \tag{9.72}$$

which we rewrite, recalling Eq. 9.33 and Eq. 9.44, as

$$S^{(1)} = ig \int_{-\infty}^{\infty} d^4x (A_\mu^+ + A_\mu^-)(\bar{\Psi}^+ + \bar{\Psi}^-)\gamma^\mu(\Psi^+ + \Psi^-). \tag{9.73}$$

We can see this second term is actually 8 terms. Let us have a look at how one of these terms, we call it $S_1^{(1)}$, acts on a state consisting of, for example, one electron and one positron with prepared momenta $|e^+(p_1), e^-(p_2)\rangle$

$$ig \int_{-\infty}^{\infty} d^4x A_\mu^+ \bar{\Psi}^+ \gamma^\mu \Psi^+ |e^+(p_1), e^-(p_2)\rangle .$$

Ψ^+ consists of destruction[36] operators for particles for all[37] possible momenta, multiplied with constants and a term of the form e^{-ipx_1}:

$$\Psi^+ \propto \int d^3p \, c_r(p) e^{-ipx} .$$

For each momentum this destroys the ket, which means we get a zero, because we are trying to destroy something which is not there, except for $c_r(p) = c_r(p_2)$. Therefore, operating with Ψ^+ results in

$$\Psi^+ |e^+(p_1), e^-(p_2)\rangle \propto e^{-ip_2x} |e^+, 0\rangle$$

In the same way, operating with $\bar{\Psi}^+$ on the ket results in

$$\bar{\Psi}^+ e^{-ip_2x} |e^+(p_1), 0\rangle \propto e^{-ip_2x} e^{-ip_1x} |0, 0\rangle .$$

Therefore, we are left with the pure vacuum state, multiplied with lots of constants. The last term operating on the ket is A_μ^+, which creates photons of all momenta.

Qualitatively we have, when we want the contribution of this one term to the probability amplitude for the creation of a photon $\langle \gamma|$ with momentum k'

$$\langle f| S_1^{(1)} |i\rangle = \langle \gamma_{k'}| S_1^{(1)} |e^+(p_1), e^-(p_2)\rangle \tag{9.74}$$

$$= \int_{-\infty}^{\infty} d^4x \, \langle \gamma_{k'}| \sum_k \text{constant}(k) |\gamma_k\rangle \, e^{-ix(p_1+p_2-k)}$$

[36] Remember the operators with the † are those who create and the operators without are those who destroy. Ψ^+ is defined in Eq. 9.33.

[37] We integrate over all possible momenta!
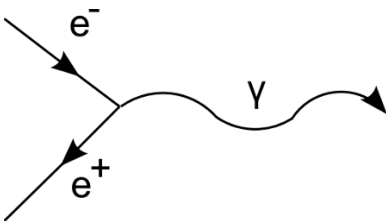
Fig. 9.1: Feynman graph for the process $e^+e^- \to \gamma$

$$= \int_{-\infty}^{\infty} d^4x \sum_k \text{constant}(k) \underbrace{\langle \gamma_{k'} | \, | \gamma_k \rangle}_{=\delta_{kk'}} e^{-ix(p_1+p_2-k)}$$

$$= \int_{-\infty}^{\infty} d^4x \, \text{constant}(k') e^{-ix(p_1+p_2-k')}.$$

The integration over x results in a delta function $\delta(p_1 + p_2 - k)$ that represents 4-momentum conservation[38]. In experiments we are never able to measure or prepare a system in one defined momentum, but only in a range. Therefore, at the end of our computation, we have to integrate over the relevant momentum range.

Take note that the seven other terms contributing to $\hat{S}^{(1)}$ result in a zero, because they destroy for example, a photon, which is not there in the beginning. If we had started with particles other than an electron and a positron, for example, a photon and a positron $|\gamma, e^+\rangle$, the first term results in a zero and some other term is non-zero.

Next we take a very quick, qualitative look at the third term of $\hat{S}$:

$$S^{(2)} = -\frac{1}{2!} T\left\{ \left(\int_{-\infty}^{\infty} d^4x_1 \mathcal{H}_I(x_1) \right) \left(\int_{-\infty}^{\infty} d^4x_2 \mathcal{H}_I(x_2) \right) \right\}$$

$$= -\frac{1}{2!} T\left\{ \left(\int_{-\infty}^{\infty} d^4x_1 g A_\mu(x_1) \bar{\Psi}(x_1) \gamma^\mu \Psi(x_1) \right) \left(\int_{-\infty}^{\infty} d^4x_2 g A_\mu(x_2) \bar{\Psi}(x_2) \gamma^\mu \Psi(x_2) \right) \right\},$$

$$(9.75)$$

where the time-ordering can be rewritten using Wick's Theorem into a sum of normal-ordered, denoted $N\{\}$, terms with commutators in it. Normal ordering means, putting all creation operators to the left, and all annihilation operators to the right. For instance, $N\{aa^\dagger a^\dagger a\} = a^\dagger a^\dagger aa$. One of the terms of this sum, for example, is

$$-\frac{1}{2!} g^2 \int \int d^4x_1 d^4x_2 N\left\{ \bar{\Psi}(x_1) \gamma^\mu \Psi(x_1) [A_\mu(x_1), A_\mu(x_2)] \bar{\Psi}(x_2) \gamma^\mu \Psi(x_2) \right\}$$

where the computation of the commutator can be done using the explicit solution for A_μ and the result is called Photon propagator[39] $[A_\mu(x_1), A_\mu(x_2)] \equiv iD_\mu(x_1 - x_2)$.

From this term we get again many, many terms, because every Ψ, $\bar{\Psi}$ etc. is actually a sum of two terms, and we will take a look at just one of them. Therefore, qualitatively we have for one[40] of these many many terms, if we start again with an electron and a positron, something of the form

$$-\frac{1}{2!} g^2 \int \int d^4x_1 d^4x_2 \bar{\Psi}^-(x_1) \Psi^-(x_1) D_\mu(x_1 - x_2) \bar{\Psi}^+(x_2) \Psi^+(x_2) |e^+, e^-\rangle$$

$$(9.76)$$

[38] The 4-momentum (=energy $p_0 = E$ and ordinary momentum p_i) of e^- plus the 4-momentum of e^+ must be equal to the 4-momentum of the photon γ: $p_1 + p_2 - k \overset{!}{=} 0$. Otherwise $\delta(p_1 + p_2 - k) = 0$ as explained in Appendix D.2.

[39] The propagators are one of the most complicated things to derive in quantum field theory. For instance, considering the scalar propagator, the starting point is $i\Delta \equiv \langle 0| T\{\Phi(x)\Phi^\dagger(y)\} |0\rangle$, which creates a particle from the vacuum at y and destroys it again at x. This can be rewritten as the commutator $i\Delta = \langle 0| [\Phi^{+\dagger}(y), \Phi^-(x)] |0\rangle$ and after many pages of math we get an expression of the form $i\Delta = \frac{-i}{2\pi^3} \int \frac{d^3k}{\omega_k} e^{-ik(x-y)}$. This final expression is what we can use in computations. This is just a sketch and many things here are not completely correct. Nevertheless it should give you a rough idea of what is going on.

[40] We pick again the term not resulting in a zero for the collision of an electron with a positron.

This can be understood physically:

- The two particles we start with are destroyed at x_2 by $\bar{\Psi}^+(x_2)\Psi^+(x_2)$.

- Then the propagator creates a "virtual" photon at x_2 and propagates it to x_1 where it is destroyed.

- Finally $\bar{\Psi}^-(x_1)\Psi^-(x_1)$ again create at x_1 an electron and a positron.

We can therefore compute with this term the probability amplitude for a reaction $e^+e^- \rightarrow e^+e^-$, where of course the individual momenta of the incoming and outgoing particles can be completely different, but not their sum[41]. In the same way, all the other terms can be interpreted as some reaction between massive[42] spin $\frac{1}{2}$ and massless[43] spin 1 fields.

The probability amplitudes we get from computations like this can then be directly checked in experiments, because the probability amplitude is directly connected to a quantity that can be measured in experiments: the cross section.

The techniques outlined in this chapter can be used to derive many important results of quantum field theory. The other interaction terms we derived can be put into the interaction Hamiltonian and the corresponding probability amplitudes follow analogously. Take not that the method we discussed here, only works if the coupling constant is smaller than 1. If the coupling constant is bigger than 1, higher order terms in the sum in Eq. 9.70 are bigger than lower order terms and therefore it is by no means justified to use just the first terms of the series, to get an approximation.

Further Reading Tips

- **Robert D. Klauber - Student Friendly Quantum Field Theory**[44] is, in my humble opinion, the best introduction to quantum field theory. All chapters are pedagogically brilliant, because the author spent a lot of time thinking about what problems someone learning quantum field theory faces.

- **Francis Halzen, Alan D. Martin - Quarks and Leptons: An Introductory Course in Modern Particle Physics**[45] is a great book which focusses on the applications of the computational schemes of quantum field theory.

- **Anthony Zee - Quantum Field Theory in a Nutshell**[46] has some brilliant and unique explanations, but some chapters are simply too short to understand as a beginner. Highly recommended *after* learning some quantum field theory from another book.

[41] This must be the case, because we have conservation of momentum.

[42] Here we looked at electrons and positrons. Other possibilities are the quarks or the other two leptons μ and τ.

[43] Here photons

Fig. 9.2: Feynman graph for the process $e^+e^- \rightarrow \gamma \rightarrow e^+e^-$

[44] Robert D. Klauber. *Student Friendly Quantum Field Theory*. Sandtrove Press, 2nd edition, 12 2013. ISBN 9780984513956

[45] Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1st edition, 1 1984. ISBN 9780471887416

[46] Anthony Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 1st edition, 3 2003. ISBN 9780691010199

- **Franz Mandl, Graham Shaw - Quantum Field Theory**[47] is a very good starting point regarding weak and strong interaction theory

- **Michele Maggiore - A Modern Introduction to Quantum Field Theory**[48] is a great introduction with a strong focus on group theoretical concepts.

- **Matthew Schwartz - Quantum Field Theory and the Standard Model**[49] offers illuminating explanations on many advanced topics.

[47] Franz Mandl and Graham Shaw. *Quantum Field Theory*. Wiley, 2nd edition, 5 2010. ISBN 9780471496847

[48] Michele Maggiore. *A Modern Introduction to Quantum Field Theory*. Oxford University Press, 1st edition, 2 2005. ISBN 9780198520740

[49] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 1 edition, 12 2013. ISBN 9781107034730

9.6 Appendix: Most General Solution of the Klein-Gordon Equation

It is not too hard to find **one** solution of the Klein-Gordon equation. Certainly plane waves

$$\Phi(x) = a e^{i(px - Et)} = a e^{-i(p_\mu x^\mu)}$$

do the job, because of the energy-momentum relation of special relativity (Eq. 8.2)

$$E^2 = \vec{p}^2 + m^2 \rightarrow p_\mu p^\mu = m^2$$

as we can check:

$$
\begin{aligned}
(\partial_\mu \partial^\mu + m^2)\Phi &= (\partial_\mu \partial^\mu + m^2)e^{-i(p_\mu x^\mu)} \\
&= (-p_\mu p^\mu + m^2)e^{-i(p_\mu x^\mu)} \\
&= (-m^2 + m^2)e^{-i(p_\mu x^\mu)} \\
&= 0 \quad \checkmark
\end{aligned}
\tag{9.77}
$$

Because we differentiate the field twice the sign in the exponent does not matter. Equally

$$\Phi^\dagger(x) = a^\dagger e^{-i(px - Et)} = a^\dagger e^{i(p_\mu x^\mu)}$$

is a solution. Further solutions can be built as linear combinations. A general solution is given by superposition of all possible solutions, which can be seen as Fourier expansion[50]

$$\Phi(x) = \int \frac{dp^4}{(2\pi)^4}\left(a(p)e^{-i(p_\mu x^\mu)} + a^\dagger(p)e^{i(p_\mu x^\mu)}\right)$$

Take note that we wrote $a = a(p)$ because we can have a different multiplication factor for each value of p and each term in the summation is a solution on its own.

In this context it is conventional to work with the wave number $k_i \equiv \frac{p_i}{\hbar}$ and the frequency $k_0 = w \equiv \frac{E}{\hbar}$ instead of the energy and

[50] That's were the factors of 2π come from. Another way of seeing this is demanding the solutions to form an orthonormal set: $\int dk e^{ikx}e^{-ikx'} = \int dk e^{ik(x-x')} = 2\pi\delta(x - x')$. Therefore, the factors of 2π are normalisation constants.

momentum. Because we work with $\hbar = 1$ we simply have to rename our variables to get the standard textbook expressions. Furthermore it's conventional to abbreviate $kx \equiv k_\mu x^\mu$. We therefore write

$$\Phi(x) = \int \frac{dk^4}{(2\pi)^4} \left(a(k)e^{-i(kx)} + a^\dagger(k)e^{i(kx)} \right)$$

Take note that not all solutions of the Klein-Gordon equation are suited to describe nature, because we have the **"mass-shell" condition**

$$p_\mu p^\mu = k_\mu k^\mu = m^2 \rightarrow k_0^2 - k_i^2 \overset{!}{=} m^2 \rightarrow k^2 \overset{!}{=} m^2. \quad (9.78)$$

Only solutions satisfying the mass-shell condition are in agreement with the relativistic energy-momentum relation (Eq. 8.2). We can build this condition into our equation, excluding all non-physical (off-shell) solutions with a delta distribution[51]

[51] This is explained in Appendix D.2.

$$\Phi(x)_{\text{physical}} = \int \frac{dk^4}{(2\pi)^4} 2\pi\delta(k^2 - m^2) \left(a(k)e^{-i(kx)} + a^\dagger(k)e^{i(kx)} \right)$$

Besides that, only positive energy solutions are physical, because as explained earlier, otherwise the energy would be unbounded from below and nothing would be stable. We can built this constraint on our solutions into the equation by using a Heaviside function $\theta(k_0)$, being zero for $k_0 < 0$ and 1 for $k_0 \geq 0$. Our integral then reads

$$\Phi(x)_{\text{physical}} = \int \frac{1}{(2\pi)^3} \underbrace{dk^4 \delta(k^2 - m^2)\theta(k_0)}_{\text{measure}} \left(a(k)e^{-i(kx)} + a^\dagger(k)e^{i(kx)} \right)$$

where we can rewrite the measure as follows

$$dk^4 \delta(k^2 - m^2)\theta(k_0) = dk^4 \delta(k_0^2 \underbrace{-\vec{k}^2 - m^2}_{\equiv -\omega_k^2 \text{ (definition)}})\theta(k_0)$$

$$= dk^4 \delta(k_0^2 - \omega_k^2)\theta(k_0)$$

$$= dk^4 \delta\big((k_0 - \omega_k)(k_0 + \omega_k)\big)\theta(k_0)$$

$$\underbrace{=}_{} dk^4 \frac{1}{2k_0}\big(\delta(k_0 - \omega_k) + \delta(k_0 + \omega_k)\big)\theta(k_0)$$

using $\delta\big(f(x)\big) = \sum_i \frac{\delta(x - a_i)}{\left|\frac{df}{dx}(a_i)\right|}$ where a_i denotes the roots, i.e. $f(a_i) = 0$, of the function $f(x)$

$$\underbrace{=}_{} dk^4 \frac{1}{2k_0}\delta(k_0 - \omega_k)$$

because the argument of $\delta(k_0 + \omega_k)$ never becomes zero with $k_0 \geq 0$

$$= dk^3 dk_0 \frac{1}{2k_0}\delta(k_0 - \omega_k)$$

$$\underbrace{=}_{\text{Integrating over } k_0} dk^3 \frac{1}{2\omega_k}. \quad (9.79)$$

So finally the general and physical solution of the Klein-Gordon equation reads

$$\Phi(x) = \int dk^3 \frac{1}{(2\pi)^3 2\omega_k} \left(a(k)e^{-i(kx)} + a^\dagger(k)e^{i(kx)} \right) . \qquad (9.80)$$

10

Classical Mechanics

In this chapter we want to explore the connection between quantum and classical mechanics. We will see that the time derivative of the expectation value for the momentum operator gives us exactly Newton's second law, which is one of the foundations of classical mechanics.

Starting with the expectation value for an operator (Eq. 8.14)

$$\langle \hat{O} \rangle = \int d^3 x \Psi^\star \hat{O} \Psi$$

and the Schrödinger equation for a particle in an external potential (Eq. 8.23)

$$(i\frac{d}{dt} - \frac{\nabla^2}{2m})\Psi - V\Psi = 0$$

$$\rightarrow i\frac{d}{dt}\Psi = \underbrace{\left(\frac{\nabla^2}{2m} + V \right)}_{=:H} \Psi$$

$$\rightarrow \frac{d}{dt}\Psi = \frac{1}{i}H\Psi$$

$$\underset{\substack{\text{Because } H^\dagger = H}}{\rightarrow} \frac{d}{dt}\Psi^\star = -\frac{1}{i} \underset{\substack{\text{Here } \Psi^\dagger = \Psi^\star}}{\Psi^\star} H \, .$$

Taking the time derivative of the expectation value yields

$$\frac{d}{dt}\langle \hat{O} \rangle = \int d^3 x \left(\left(\frac{d}{dt}\Psi^\star \right) \hat{O}\Psi + \Psi^\star \left(\frac{d}{dt}\hat{O} \right) \Psi + \Psi^\star \hat{O} \left(\frac{d}{dt}\Psi \right) \right) \, .$$

We use now $\frac{d}{dt}\hat{O} = 0$, which is true for most operators. For example, for $\hat{O} = \hat{\vec{p}} = -i\vec{\nabla} \neq \hat{O}(t)$. In addition, we use the Schrödinger equation to rewrite the time derivatives of the wave function and its

complex conjugate. This yields

$$
\begin{aligned}
\frac{d}{dt} \langle \hat{O} \rangle &= \int d^3x \left(\left(-\frac{1}{i} \Psi^\star H \right) \hat{O} \Psi + \Psi^\star \hat{O} \left(\frac{1}{i} H \Psi \right) \right) \\
&= \frac{1}{i} \int d^3x \left(-\Psi^\star H \hat{O} \Psi + \Psi^\star \hat{O} H \Psi \right) \\
&= \frac{1}{i} \int d^3x \Psi^\star [\hat{O}, H] \Psi \\
&= \frac{1}{i} \langle [\hat{O}, H] \rangle,
\end{aligned}
\tag{10.1}
$$

which is known as **Ehrenfest theorem**. If we specify $\hat{O} = \hat{p}$ and use $H = \frac{\hat{p}^2}{2m} + V$, we get

$$
\begin{aligned}
\frac{d}{dt} \langle \hat{p} \rangle &= \frac{1}{i} \langle [\hat{p}, H] \rangle \\
&= \frac{1}{i} \langle [\hat{p}, \frac{\hat{p}^2}{2m} + V] \rangle \\
&= \frac{1}{i} \langle \underbrace{[\hat{p}, \frac{\hat{p}^2}{2m}]}_{=0} + [\hat{p}, V] \rangle \\
&= \frac{1}{i} \langle [\hat{p}, V] \rangle \\
&= \frac{1}{i} \int d^3x \Psi^\star [\hat{p}, V] \Psi \\
&= \frac{1}{i} \int d^3x \Psi^\star \hat{p} V \Psi - \frac{1}{i} \int d^3x \Psi^\star V \hat{p} \Psi \\
&= \frac{1}{i} \int d^3x \Psi^\star (-i\nabla) V \Psi - \frac{1}{i} \int d^3x \Psi^\star V (-i\nabla) \Psi \\
&\underbrace{=}_{\text{Product rule}} - \int d^3x \Psi^\star (\nabla V) \Psi - \int d^3x \Psi^\star V \nabla \Psi + \int d^3x \Psi^\star V \nabla \Psi \\
&= - \int d^3x \Psi^\star (\nabla V) \Psi \\
&= \langle -\nabla V \rangle = \langle F \rangle.
\end{aligned}
\tag{10.2}
$$

In words this means that the time derivative (of the expectation value) of the momentum equals the (expectation value of the) negative gradient of the potential, which is known as force. This is exactly **Newton's second law**[1]. This equation can be used to compute the trajectories of macroscopic objects. Historically the force laws were deduced phenomenologically from experiments. All forces acting on an object were added linearly on the right-hand side of the equation. By using the phenomenologically deduced momentum $p_{\text{mak}} = mv$, we can write the left-hand side as $\frac{d}{dt} p_{\text{mak}} = \frac{d}{dt} mv$, which equals for an object with constant mass $m \frac{d}{dt} v$. The velocity is the time-derivative of the location[2] and we therefore have

$$
m \frac{d^2}{dt^2} x = F_1 + F_2 + \ldots
\tag{10.3}
$$

[1] Recall that we used this equation, without a derivation, to illustrate the conserved quantities following from Noether's theorem in Section 4.5. Here we deliver the derivation, as promised.

[2] In other words, the velocity $v = \frac{d}{dt} x(t) = \dot{x}(t)$ is the change-rate of the position of the object and equally the acceleration $a \frac{d}{dt} \frac{d}{dt} x(t) = \frac{d}{dt} v = \ddot{x}(t)$ is the change-rate of the velocity.

This is the differential equation one must solve for $x = x(t)$ to get the trajectory of the object in question. An example for such a classical force will be derived in the next chapter.

10.1 Relativistic Mechanics

Using the Lagrangian formalism, we can look at classical mechanics from quite a different perspective. We need an equation that describes the motion of individual particles. As always in this book, we assume that we can derive the correct equation if we minimize *something*. We already discussed at the beginning of Chapter 4 that this *something* must be invariant under all Lorentz transformations, because otherwise we don't get the same equations of motion in all frames of reference.

Luckily, we already know something that is invariant under all Lorentz transformations: the invariant of special relativity, which was derived in Section 2.1:

$$(ds)^2 = (cd\tau)^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2, \qquad (10.4)$$

where τ is the proper time as explained in Section 2.2. Equally the square root of this is invariant and therefore the simplest possible thing we can minimize is

$$S = \int C d\tau \qquad (10.5)$$

with some constant C and

$$d\tau = \frac{1}{c}\sqrt{(cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2}. \qquad (10.6)$$

The correct constant turns out to be $C = -mc^2$ and therefore we need to minimize

$$S = -mc^2 \int d\tau. \qquad (10.7)$$

For brevity we will restrict the following discussion to one dimension. Then we can write

$$d\tau = \frac{1}{c}\sqrt{(cdt)^2 - (dx)^2} = \frac{1}{c}\sqrt{(cdt)^2\left(1 - \frac{(dx)^2}{c^2(dt)^2}\right)}$$

$$= \frac{1}{c}(cdt)\sqrt{1 - \frac{1}{c^2}\left(\frac{dx}{dt}\right)^2} \underbrace{=}_{\frac{dx}{dt} = \dot{x} \text{ is the velocity of the particle in question}} dt\sqrt{1 - \frac{\dot{x}^2}{c^2}}. \qquad (10.8)$$

Putting this into Eq. 10.7 yields

$$S = \int \underbrace{-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}}}_{\equiv \mathcal{L}} \, dt. \tag{10.9}$$

As always, we can find the extremal action by putting the Lagrangian into the Euler-Lagrange equation (Eq. 4.7):

$$0 = \frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right)$$

$$\rightarrow 0 = \underbrace{\frac{\partial}{\partial x}\left(-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}}\right)}_{=0} - \frac{d}{dt}\left(\frac{\partial}{\partial \dot{x}}\left(-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}}\right)\right)$$

$$\rightarrow 0 = c^2 \frac{d}{dt}\left(\frac{-m\frac{\dot{x}}{c^2}}{\sqrt{1 - \frac{\dot{x}^2}{c^2}}}\right)$$

$$\rightarrow 0 = \frac{d}{dt}\left(\frac{m\dot{x}}{\sqrt{1 - \frac{\dot{x}^2}{c^2}}}\right). \tag{10.10}$$

This is the correct relativistic equation for a *free* particle. If the particle moves in an external potential $V(x)$, we must simply add this potential to the Lagrangian[3]

$$\mathcal{L} = -mc^2 \sqrt{1 - \frac{(\dot{x})^2}{c^2}} - V(x). \tag{10.11}$$

For this Lagrangian the Euler-Lagrange equation yields

$$\frac{d}{dt}\left(\frac{m\dot{x}}{\sqrt{1 - \frac{\dot{x}^2}{c^2}}}\right) = -\frac{dV}{dx} \equiv F. \tag{10.12}$$

Take note that in the non-relativistic limit ($\dot{x} \ll c$) we have $\sqrt{1 - \frac{(\dot{x})^2}{c^2}} \approx 1$ and the equation is exactly the equation of motion we derived in the last section (Eq. 10.3).

10.2   The Lagrangian of Non-Relativistic Mechanics

It is instructive to have a look at the non-relativistic limit of the Lagrangian we derived in the last section (Eq. 10.9). The "non-relativistic limit" means, we consider a situation where the particle moves slowly compared with the speed of light: $\dot{x} \ll c$. We can then use the Taylor formula for the Lagrangian[4]

[3] We add $-V(x)$ instead of $V(x)$, because then we get in the formula for the total energy of the system, which we can compute using Noether's theorem, the potential energy term $+V(x)$.

[4] See Appendix B.3

$$-mc^2\sqrt{1 - \frac{\dot{x}^2}{c^2}} = -mc^2\left(1 - \frac{1}{2}\frac{\dot{x}^2}{c^2} + \ldots\right). \qquad (10.13)$$

In the limit $\dot{x} \ll c$ we can neglect higher order terms and the Lagrangian reads

$$\mathcal{L} = -mc^2 + \frac{1}{2}m\dot{x}^2 - V(x). \qquad (10.14)$$

We already know that a constant like $-mc^2$ has no influence on the equation of motion and therefore the **Lagrangian for non-relativistic mechanics** reads

$$\mathcal{L} = \frac{1}{2}m\dot{x}^2 - V(x). \qquad (10.15)$$

Without the external potential, i.e. $V(x) = 0$, this is exactly the Lagrangian we used in section 4.5.1 to illustrate the conserved quantities that follow from Noether's theorem. Putting this Lagrangian into the Euler-Lagrange equation (Eq. 4.7) yields

$$\frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) = 0$$

$$\rightarrow \frac{\partial}{\partial x}\left(\frac{1}{2}m\dot{x}^2 - V(x)\right) - \frac{d}{dt}\left(\frac{\partial}{\partial \dot{x}}\left(\frac{1}{2}m\dot{x}^2 - V(x)\right)\right) = 0$$

$$\rightarrow -\frac{\partial}{\partial x}V(x) - \frac{d}{dt}m\dot{x} = 0$$

$$\rightarrow \frac{d}{dt}m\dot{x} = -\frac{\partial}{\partial x}V(x)$$

$$(10.16)$$

This is once more exactly the equation of motion we derived at the beginning of this chapter (Eq. 10.3).

11

Electrodynamics

We already derived in Chapter 7 one of the most important equations of classical electrodynamics: the inhomogeneous Maxwell equations (Eq. 7.22)

$$\partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) = J^\rho \tag{11.1}$$

or in a more compact form using the electromagnetic tensor[1]

$$\partial_\sigma F^{\sigma\rho} = J^\rho. \tag{11.2}$$

We discovered in Section 7.1.6 that J^ρ is a Noether current, i.e. $\partial_\rho J^\rho = 0$. In a macroscopic theory this conserved current is the electric four-current. The tensor $F^{\sigma\rho}$ is antisymmetric $F^{\sigma\rho} = -F^{\rho\sigma}$, which can be seen directly from the definition $F^{\sigma\rho} = \partial^\sigma A^\rho - \partial^\rho A^\sigma$, and has therefore 6 independent components. Three are

$$F^{i0} = \partial^i A^0 - \partial^0 A^i \tag{11.3}$$

with $i = 1, 2, 3$ and the three other are

$$F^{ij} = \partial^i A^j - \partial^j A^i = (\delta^i_l \delta^j_m - \delta^i_m \delta^j_l)\partial^l A^m \underbrace{=}\ \epsilon^{ijk}\epsilon^{klm}\partial^l A^m. \tag{11.4}$$
$$\text{This is a standard relation for the multiplication of two } \epsilon \text{ with one coinciding index}$$

The standard way to label these components is

$$\partial^i A^0 - \partial^0 A^i \equiv E^i \tag{11.5}$$

$$\epsilon^{ijk}\partial^j A^k \equiv -B^i \tag{11.6}$$

and therefore

$$F^{i0} = E^i \tag{11.7}$$

$$F^{ij} = \epsilon^{ijk}\epsilon^{klm}\partial^l A^m = -\epsilon^{ijk}B^k. \tag{11.8}$$

If we now rewrite the inhomogeneous Maxwell equations[2] as

$$\partial_\sigma F^{\rho\sigma} = \partial_0 F^{\rho 0} - \partial_k F^{\rho k} = J^\rho \tag{11.9}$$

[3] $\nabla \times \vec{B}$ is $\epsilon^{ikl}\partial_k B^l$ in vector notation and commonly called cross product.

we have for the three spatial components[3] ($\rho \to i$)

$$\partial_0 F^{i0} - \partial_k F^{ik} = \partial_0 E^i + \epsilon^{ikl}\partial_k B^l = J^i$$

$$\to \partial_t \vec{E} + \nabla \times \vec{B} = \vec{J} \tag{11.10}$$

and for the time-component ($\rho \to 0$)

$$\partial_0 \underbrace{F^{00}}_{=0 \text{ see the definition of } F^{\rho\sigma}} - \partial_k F^{0k} \overbrace{=}^{\text{Because } F^{\mu\nu}=-F^{\nu\mu}} \partial_k F^{k0} \underbrace{=}_{\text{Eq. 11.7}} \partial_k E^k = J^0$$

$$\to \nabla \vec{E} = J^0. \tag{11.11}$$

This is the form of the inhomogeneous Maxwell equations that, for example, engineers use.

11.1   The Homogeneous Maxwell Equations

It follows directly from the definition of the electromagnetic tensor $F_{\mu\nu}$ that if we multiply it with something totally antisymmetric[4]

[4] $\epsilon^{\mu\nu\rho\sigma}$ is the four-dimensional Levi-Civita symbol, which is defined in Appendix B.5.5.

$$\tilde{F}^{\mu\nu} = \epsilon^{\mu\nu\rho\sigma} F^{\rho\sigma} \tag{11.12}$$

the derivative ∂_μ of this new object $\tilde{F}^{\mu\nu}$, which is called the **dual electromagnetic tensor**, vanishes:

$$\partial_\mu \tilde{F}^{\mu\nu} = \partial_\mu \epsilon^{\mu\nu\rho\sigma}(\partial_\sigma A_\rho - \partial_\rho A_\sigma) = 0. \tag{11.13}$$

This follows from the fact that if we contract two symmetric indices with two antisymmetric indices, the result is always zero[5]. This can be seen, focussing for brevity on the first term, as follows

[5] This is explained in Appendix B.5.4.

$$\epsilon^{\mu\nu\rho\sigma}\partial_\mu\partial_\sigma A_\rho = \frac{1}{2}(\epsilon^{\mu\nu\rho\sigma}\partial_\mu\partial_\sigma A_\rho + \epsilon^{\mu\nu\rho\sigma}\partial_\mu\partial_\sigma A_\rho)$$

$$\underbrace{=}_{\text{Renaming dummy indices}} \frac{1}{2}(\epsilon^{\mu\nu\rho\sigma}\partial_\mu\partial_\sigma A_\rho + \epsilon^{\sigma\nu\rho\mu}\partial_\sigma\partial_\mu A_\rho)$$

$$\underbrace{=}_{\text{Because } \epsilon^{\mu\nu\rho\sigma}=-\epsilon^{\sigma\nu\rho\mu} \text{ and } \partial_\mu\partial_\sigma=\partial_\sigma\partial_\mu} \frac{1}{2}(\epsilon^{\mu\nu\rho\sigma}\partial_\mu\partial_\sigma A_\rho - \epsilon^{\mu\nu\rho\sigma}\partial_\mu\partial_\sigma A_\rho) = 0 \quad \checkmark \tag{11.14}$$

[6] Plural because we have one equation for each component $v = 0, 1, 2, 3$.

Equally the second term is zero. The equations[6]

$$\partial_\mu \tilde{F}^{\mu\nu} = 0 \tag{11.15}$$

are known as **homogeneous Maxwell equations** and we can see that they are a direct consequence of the definition of $F^{\mu\nu}$. In order to

rewrite this in terms of B and E, we take a look at the component $\nu = 0$:

$$
\begin{aligned}
0 &= \partial_\mu \tilde{F}^{\mu 0} \\
&= \partial_\mu \epsilon^{\mu 0 \rho \sigma} F^{\rho \sigma} \\
&= \partial_0 \underbrace{\epsilon^{00\rho\sigma}}_{=0} F^{\rho\sigma} + \partial_i \epsilon^{i0\rho\sigma} F^{\rho\sigma} \\
&\underbrace{=}_{\text{Because } \epsilon^{i0\rho\sigma}=0 \text{ for } \rho=0 \text{ or } \sigma=0} \partial_i \epsilon^{i0jk} F^{jk} \\
&= -\partial_i \epsilon^{0ijk} F^{jk} \\
&\underbrace{=}_{\text{See Eq. 11.8}} \partial_i \underbrace{\epsilon^{0ijk}\epsilon^{ljk}}_{=2\delta_{il}} B^l \\
&= 2\partial_i \delta_{il} B^l = 2\partial_i B^i
\end{aligned}
$$

$$
\Rightarrow \partial_i B^i = 0 \quad \text{or in vector notation} \quad \nabla \vec{B} = 0 \qquad (11.16)
$$

Analogously, we can take a look at the components $\nu = i$ and derive

$$
\nabla \times \vec{E} + \partial_t \vec{B} = 0 \qquad (11.17)
$$

This is the conventional form of the homogeneous Maxwell equations that is used in practical applications.

11.2 The Lorentz Force

We can use the connection between quantum and classical mechanics that we discovered in the last chapter (the Ehrenfest theorem), to derive the famous **Lorentz force law**. The starting point is the equation describing a non-relativistic particle, without spin, in an external electromagnetic field, i.e. the Schrödinger equation with coupling to an external electromagnetic field[7] (Eq. 8.24):

$$
i\partial_t \Psi = \underbrace{\left(\frac{1}{2m}(\vec{p} - q\vec{A})^2 + q\Phi \right)}_{\equiv H} \Psi. \qquad (11.18)
$$

If we define the momentum[8] of this system as $\vec{\Pi} = \vec{p} - q\vec{A}$, we can write the Hamiltonian:

$$
H = \frac{1}{2m}\vec{\Pi}^2 + q\Phi. \qquad (11.19)
$$

Having defined the Hamiltonian H we are able to follow the exact same steps described in the last section and arrive at Eq. 10.1. But

[7] This equation can be derived from the Klein Gordon equation with coupling to an external electromagnetic field, which we derived from the Lagrangian describing spin 0 particles which are coupled to a massless spin 1 field, i.e. the photon field. Therefore, the real starting point is once more Lorentz and gauge symmetry, which we used to derive the corresponding Lagrangian. We use here the notation $A_0 \equiv \Phi$.

[8] This is actually the momentum of the system following from invariance under translations using the Noether theorem $\frac{\partial L}{\partial \dot{x}} = \Pi$.

this time we can't neglect the partial time-derivative term, because the operator we are going to look at is time dependent:

$$\frac{d}{dt}\langle \hat{O} \rangle = \frac{1}{i}\langle [O, H] \rangle + \langle \frac{\partial O}{\partial t} \rangle$$

$$\rightarrow \frac{d}{dt}\langle \vec{\Pi} \rangle = \frac{1}{i}\langle [\vec{\Pi}, H] \rangle + \langle \frac{\partial \vec{\Pi}}{\partial t} \rangle .$$

We can see that $\frac{\partial \vec{\Pi}}{\partial t} \neq 0$ because A can change with time. If we now put the explicit form of H (Eq. 11.19) into the equation, we get

$$\rightarrow \frac{d}{dt}\langle \vec{\Pi} \rangle = \frac{1}{i}\langle [\vec{\Pi}, \frac{1}{2m}\vec{\Pi}^2 + q\Phi] \rangle + \langle \frac{\partial \vec{\Pi}}{\partial t} \rangle$$

$$\rightarrow \frac{d}{dt}\langle \vec{\Pi} \rangle = \frac{1}{i}\langle [\vec{\Pi}, \frac{1}{2m}\vec{\Pi}^2] \rangle + \underbrace{\frac{1}{i}\langle [\vec{\Pi}, q\Phi] \rangle}_{=\langle q\nabla\Phi \rangle} + \langle \frac{\partial \vec{\Pi}}{\partial t} \rangle .$$

In the last step we use that $[\vec{\Pi}, q\Phi]$ can be computed analogous to $[\hat{p}, V]$, which we considered in the last chapter, because $[A, \Phi] = 0$. The next task is to compute $[(\vec{\Pi})^2, \vec{\Pi}]$, which is non-trivial, because the components Π_i do not commute. Instead, we have

$$[\Pi_i, \Pi_j] = -\frac{q}{i}\left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} \right) = -\frac{q}{i}\epsilon_{ijk}\underbrace{B_k}_{=\epsilon_{klm}\frac{\partial}{\partial x_l}A_m} \tag{11.20}$$

with the usual definition of the magnetic field $B = \nabla \times A$ written in index form. If we now define the speed[9] of our particle as $\vec{v} \equiv \frac{\vec{\Pi}}{m}$ we arrive at

$$\frac{1}{2m}[\vec{\Pi}^2, \vec{\Pi}] = \frac{q}{2i}(v \times B - B \times v) = \frac{q}{i}(\vec{v} \times \vec{B})$$

Then we can write the classical equation of motion as

$$\frac{d}{dt}\langle \vec{\Pi} \rangle = \langle q\nabla\Phi \rangle - \langle q(v \times B) \rangle + \underbrace{\langle \frac{\partial \vec{\Pi}}{\partial t} \rangle}_{=-q\frac{\partial \vec{A}}{\partial t}}$$

$$\frac{d}{dt}\langle \vec{\Pi} \rangle = -q\langle (v \times B) \rangle + q\underbrace{\langle \nabla\Phi - \frac{\partial \vec{A}}{\partial t} \rangle}_{=\langle E \rangle \text{ see Eq. 11.5}} .$$

We finally get

$$\frac{d}{dt}\langle \vec{\Pi} \rangle \equiv F_{\text{Lorentz}} = -q(\langle (v \times B) \rangle + \langle E \rangle). \tag{11.21}$$

This is the equation of motion that describes the classical trajectory of a particle in an external electromagnetic field.

11.3   Coulomb Potential

We learned in Chapter 7 that our Lagrangians are invariant under internal transformations. We can use this freedom to simplify computations, i.e. we transform the field in question with an internal transformation such that the computation becomes especially simple. This is allowed because the physics we describe with the field and the transformed field are the same, as long as we stick to the gauge transformations that leave the Lagrangian invariant. An often used choice is gauging the photon field A^μ such that[10] $\partial_\mu A^\mu = 0$. This is called the Lorenz gauge. Using this gauge simplifies the inhomogeneous Maxwell equations to

$$\partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) = \partial_\sigma \partial^\sigma A^\rho = J^\rho \tag{11.22}$$

We now have a look at the physical situation where a fixed, static charge is located in a spherically symmetric region around the origin of our coordinate system. We want to describe physics in the outside region, which means in the region without a source in it, i.e. $J^\rho = 0$. Therefore the Maxwell equations are in this region

$$\partial_\sigma \partial^\sigma A^\rho = \partial_0 \partial_0 A^\rho - \partial_i \partial_i A^\rho = 0. \tag{11.23}$$

We now use that we are considering a static ($\partial_0 A^\rho = 0$), spherically symmetric system, by rewriting the equation using spherical coordinates[11]. Then we can neglect all terms[12] but the term involving the ∂_r derivative. This yields

$$\rightarrow \frac{\partial^2}{\partial r^2}(rA^\mu) = 0. \tag{11.24}$$

The general solution of this equation is

$$A^\mu = \epsilon^\mu \frac{C}{r} + \epsilon^\mu D \tag{11.25}$$

with some constant four-vector ϵ^μ and constants C and D. The field A^μ must vanish at infinity and therefore $D = 0$. For the zeroth component[13] of A^μ this is the famous Coulomb potential

$$A^0 = \Phi = \frac{C}{r} = \frac{Ze}{r}, \tag{11.26}$$

where Z is an integer and e the electric charge of an electron. The reason for writing the constant C like this is that the electric charge is quantized in terms of multiples of the electron charge. In the standard model there is no satisfying explanation for this curious fact of nature.

[10] It can be shown explicitly that there is such a gauge choice. Further details can be found in the standard textbooks about electrodynamics.

[11] Instead of using x, y, z to determine the position of some objects, it's possible to use two angels θ, ϕ and the distance from the origin r. Then we have $\partial_i \partial_i A^\rho = \frac{\partial^2}{\partial r^2}(rA^\rho) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta \frac{\partial A^\rho}{\partial\theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2 A^\rho}{\partial\phi^2}$. This is especially useful when considering spherically symmetric systems, because these do only depend on r.

[12] For a spherically symmetric field, we have $\partial_\theta A = \partial_\phi A = 0$.

[13] In a lengthy computation using symmetry considerations it can be shown that all other components vanish.

Further Reading Tips

[14] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics: Volume 2*. Addison-Wesley, 1st edition, 2 1977. ISBN 9780201021172

[15] David J. Griffiths. *Introduction to Electrodynamics*. Addison-Wesley, 4th edition, 10 2012. ISBN 9780321856562

- **Richard P. Feynman - The Feynman Lectures on Physics Volume 2**[14] is a great book to start learning electrodynamics.

- **David J. Griffiths - Introduction to Electrodynamics**[15] is another great book to learn more about the concepts of electrodynamics.

12

Gravity

Unfortunately, the best theory of gravity we have does not fit into the picture outlined in the rest of this book. This is one of the biggest problems of modern physics and the following paragraphs try to give you a first impression.

The modern theory of gravity is **Einstein's general relativity**. The fundamental idea is that gravity is a result of the curvature of spacetime. Mass and energy change the curvature of spacetime and in turn the changed curvature influences the movement of mass and energy. This interplay between energy and curvature is described by the famous Einstein equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{12.1}$$

On the left-hand side is the Einstein tensor $G_{\mu\nu}$, which describes the curvature and on the right-hand side is the energy-momentum tensor[1] $T_{\mu\nu}$. G is the gravitation constant.

From the idea gravity = curvature of spacetime, the derivation of the Einstein equation is, from a modern point of view, relatively straightforward[2]. Firstly, one of the most important laws of physics is the conservation of energy and momentum, which as we saw, follows directly when considering a homogeneous spacetime. In a homogeneous spacetime the laws of physics are invariant under translations in space and time and using Noether's theorem, we can derive the conservation of momentum and energy. Therefore, one of the most basic assumptions of physics is that spacetime is homogeneous and therefore energy and momentum are conserved. In mathematical terms this conservation law is expressed as (Eq. 4.36)

$$\partial^{\mu} T_{\mu\nu} = 0. \tag{12.2}$$

[1] Recall that the energy-momentum tensor is the quantity which is directly related to translational symmetry (Eq. 4.36).

[2] Einstein needed 100 years ago ca. 6 years for the derivation of the correct equation. Today, with the power of hindsight we are much faster.

[3] Up to this point we only considered the Minkowski metric
$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \text{ and }$$

the Euclidean metric $\delta^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
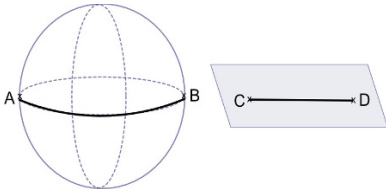
Fig. 12.1: Distance between two points in a curved and a flat space

[4] $\partial^\mu G_{\mu\nu} = 0$

[5] This means two indices $\mu\nu$, which is a requirement, because $T_{\mu\nu}$ on the right-hand side has two indices, too.

Next we need something to describe curvature mathematically. This is what makes general relativity computationally very demanding. Nevertheless, we already know the most important object: the metric. Recall that metrics are the mathematical objects that enable us to compute the distance between two points[3]. In a curved space the distance between two points is different than in a flat space as illustrated in Fig. 12.1. Therefore metrics will play a very important role when thinking about curvature in mathematical terms.

Having talked about this, we are ready to "derive" the Einstein equation, because it turns out that there is exactly one mathematical object that we can put on the left-hand side: the Einstein tensor $G_{\mu\nu}$. The Einstein tensor is the only divergence-free[4] function of the metric $g_{\mu\nu}$ and at most its first and second partial derivative. Therefore, the Einstein tensor may be very complicated, but it's the only object we are allowed to write on the left-hand side describing curvature. This follows, because we can conclude from

$$T_{\mu\nu} = C G_{\mu\nu} \quad \text{that} \quad \partial^\mu T_{\mu\nu} = 0 \to \partial^\mu G_{\mu\nu} = 0 \tag{12.3}$$

must hold, too. The Einstein tensor is a second rank tensor[5] and has exactly this property.

The Einstein tensor is defined as a sum of the Ricci Tensor $R_{\mu\nu}$ and the trace of the Ricci tensor, called Ricci scalar $R = R^\nu_\nu$

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \tag{12.4}$$

where the Ricci Tensor $R_{\mu\nu}$ is defined in terms of the Christoffel symbols $\Gamma^\mu_{\nu\rho}$

$$R_{\alpha\beta} = \partial_\rho \Gamma^\rho_{\beta\alpha} - \partial_\beta \Gamma^\rho_{\rho\alpha} + \Gamma^\rho_{\rho\lambda} \Gamma^\lambda_{\beta\alpha} - \Gamma^\rho_{\beta\lambda} \Gamma^\lambda_{\rho\alpha} \tag{12.5}$$

and the Christoffel Symbols are defined in terms of the metric

$$\Gamma_{\alpha\beta\rho} = \frac{1}{2} \left(\frac{\partial g_{\alpha\beta}}{\partial x^\rho} + \frac{\partial g_{\alpha\rho}}{\partial x^\beta} - \frac{\partial g_{\beta\rho}}{\partial x^\alpha} \right) = \frac{1}{2} \left(\partial_\rho g_{\alpha\beta} + \partial_\beta g_{\alpha\rho} - \partial_\alpha g_{\beta\rho} \right). \tag{12.6}$$

This can be quite intimidating and shows why computations in general relativity very often need massive computational efforts.

Next we need to know how things react to such a curved spacetime. What's the path of an object from A to B in curved spacetime? The first guess is the correct one: An object follows the shortest path between two points in curved spacetime. We can start with a given distribution of energy and mass, which means some $T_{\mu\nu}$, compute the metric or Christoffel symbols with the Einstein equation and then get

the trajectory through the **geodesic equation**

$$\frac{d^2x^\lambda}{dt^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} = 0. \tag{12.7}$$

The geodesic is the locally shortest[6] curve between two points on a manifold.

It is interesting to note that Einstein thought about the Christoffel Symbols as the gravitational field

> If the $\Gamma^\mu_{\nu\rho}$ vanish, then the point moves uniformly in a straight line. These quantities therefore condition the deviation of the motion from uniformity. They are the components of the gravitational field.

> **- Albert Einstein**[7]

We can understand what Einstein means by looking at Eq. 12.7. For $\Gamma^\mu_{\nu\rho} = 0$ the geodesic equation reduces to

$$\frac{d^2x^\lambda}{dt^2} = 0. \tag{12.8}$$

The solutions of this equation describe a straight line.

Another interesting aspect of a curved spacetime is that the notion of differentiation changes. Remember how the derivative is defined in flat space using the difference quotient

$$f'(a) = \lim_{h\to 0}\frac{f(a+h)-f(a)}{h}. \tag{12.9}$$

This definition requires that we compare the function in question at two different points. In a curved space this comparison is not as trivial as in the flat space. Take a look at Fig. 12.2. If we want to compare two vectors on a sphere, how can we make sure that the vectors are really different and the difference is not just an effect of the curved space? The answer of differential geometry is parallel transport. We have to move one vector to the location of the other one, to be able to compare them[8].

The derivative becomes in a curved space the **covariant derivative**[9]

$$D_b v^a \equiv \partial_b v^a + \Gamma^a{}_{bc} v^c. \tag{12.10}$$

Therefore, if we want any equation we derived so far to be valid in curved spacetime, we need to change

$$\partial_b \to D_b = \partial_b + \Gamma^a{}_{bc}. \tag{12.11}$$

[6] This is a bit oversimplified, but the correct definition needs some terms from differential geometry we haven't introduced here.

[7] Albert Einstein. The foundation of the general theory of relativity. 1916
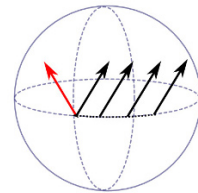
Fig. 12.2: In order to be able to compare the red arrow with the black arrow, we transport the black arrow to the location of the red arrow.

[8] In fact in differential geometry one has only local coordinate systems. The defining feature of a manifold is that it looks locally flat=Euclidean. This was discussed in Section 3.11. Therefore, the coefficients we use to describe the objects in questions are only valid in a small region of the manifold and thus we can only compare coefficients of the same coordinate system.

[9] In this context the Christoffel symbols $\Gamma^a{}_{bc}$ are often called connection coefficients, because of their property to connect the points we want to compare.

Does this look familiar? Take a look again at Eq. 7.18. We learned in an earlier chapter that a locally $U(1)$ invariant Lagrangian for spin 0 or spin $\frac{1}{2}$ fields required a specific coupling with a spin 1 field. This specific coupling can be summarized by the prescription

$$\partial_\mu \to D_\mu \equiv \partial_\mu + ieA_\mu. \tag{12.12}$$

Remember that this wasn't just a mathematical gimmick. This prescription gives us the correct theory of electromagnetism. The same is true for weak

$$\partial_\mu \to D_\mu \equiv \partial_\mu - ig \underbrace{W_\mu}_{=W_\mu^i \sigma^i} \tag{12.13}$$

and strong interactions (see Eq. 7.165)

$$\partial_\mu \to D_\mu = \partial_\mu - ig' \underbrace{G_\mu}_{=T^C G_\mu^C}. \tag{12.14}$$

Although things look quite similar here, there is for many reasons no formulation of gravity that is compatible with the quantum description of all other forces. All other forces are described in a quantum theory and one can only make probability predictions. In contrast, general relativity is a classical theory, because particles follow defined trajectories and there is no need for probability predictions. To make things worse, at the current time no experiment can shine any light on the interplay between those forces. The effects of gravity on elementary particles is too weak to be measured. Because of this, the standard model, which ignores gravity entirely and only takes the weak, the strong and the electromagnetic interactions into account, works very well. The effects of general relativity only become measurable with very heavy objects. For such quantum effects play no role, because massive objects consist of many, many elementary particles and all quantum effects get averaged out. We discovered in Chapter 10 that the equation of motion for the average value is just the classical one and no quantum effects are measurable. One can make a long list of things that make constructing a quantum theory of gravity so difficult, but Einstein formulated the difference between gravity and all other forces very concisely:

> ...according to the general theory of relativity, gravitation occupies an exceptional position with regard to other forces, particularly the electromagnetic forces, since the ten functions representing the gravitational field at the same time define the metrical properties of the space measured.

> **- Albert Einstein**[10]

[10] Albert Einstein and Francis A. Davis. *The Principle of Relativity*. Dover Publications, reprint edition, 6 1952. ISBN 9780486600819

We are able do describe quantum particles in a curved space, by changing the derivative to the covariant derivative. But this is of course no dynamical theory of gravity. We could make the right-hand side of the Einstein equation quantum, if we make the usual identification with the corresponding generator, but what about the left-hand side? The Einstein tensor in terms of the Christoffel Symbols is

$$G_{\alpha\beta} = (\delta_\alpha^\gamma \delta_\beta^\zeta - \frac{1}{2} g_{\alpha\beta} g^{\gamma\zeta})(\partial_\epsilon \Gamma^\epsilon_{\gamma\zeta} - \partial_\zeta \Gamma^\epsilon_{\gamma\epsilon} + \Gamma^\epsilon_{\epsilon\sigma} \Gamma^\sigma_{\gamma\zeta} - \Gamma^\epsilon_{\zeta\sigma} \Gamma^\sigma_{\epsilon\gamma}). \quad (12.15)$$

Thus maybe we can think of the Einstein equation as the field equation[11] for $\Gamma^\epsilon_{\zeta\sigma}$, and the terms generated by the prescription

$$\partial_b \rightarrow D_b = \partial_b + \Gamma^a{}_{bc}$$

as the corresponding coupling between the gravitational field $\Gamma^\epsilon_{\zeta\sigma}$ and the other fields?

Regardless of if you prefer to think of the metric or the Christoffel symbols as the gravitational field, the two or three vector indices indicate that we may need to investigate the $(1,1)$ or even higher representations, which we would call consequently spin $2, 3, \ldots$ representation of the Poincaré group. In fact most physicists believe that the boson responsible for gravitational attraction, the graviton, has spin 2.

Until the present day, there is no working[12] theory of quantum gravity and for further information have a look at the books mentioned in the next section.

Further Reading Tips

For more information about the standard theory of gravity, Einstein's general relativity, see

- **Ta-Pei Cheng - Relativity, Gravitation and Cosmology**[13] is a great, rather low-level introduction to general relativity with many very enlightening explanations. Perfect to get a quick overview.

- **A. Zee - Einstein Gravity in a Nutshell**[14], is the best book to learn about general relativity. It really starts at the beginning, avoids unnecessary, confusing mathematical tools and does a great job explaining the origin and usage of general relativity.

- **Charles W. Misner, Kip S. Thorne, John Archibald Wheeler - Gravitation**[15] is a really, really big book, but often offers in depth explanations for points that remain unclear in most other books.

[11] Analogous to the Maxwell equation for the electromagnetic field.

[12] Many attempts result in an infinite number of infinity terms, which is quite bad for probability predictions.

[13] Ta-Pei Cheng. *Relativity, Gravitation and Cosmology: A Basic Introduction*. Oxford University Press, 2nd edition, 1 2010. ISBN 9780199573646

[14] Anthony Zee. *Einstein Gravity in a Nutshell*. Princeton University Press, 1st edition, 5 2013. ISBN 9780691145587

[15] Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, 1st edition, 9 1973. ISBN 9780716703440

For more information about attempts to quantize gravity have a look at

[16] Lee Smolin. *Three Roads to Quantum Gravity*. Basic Books, 3 edition, 8 2017. ISBN 9780465094547

[17] John C. Baez and Javier P. Muniain. *Gauge Fields, Knots, and Gravity*. World Scientific Pub Co Inc, 1st edition, 9 1994. ISBN 9789810220341

- **Lee Smolin -Three Roads to Quantum Gravity**[16], which is brilliant popular science book that summarizes the various attempts to quantize gravity.

- **John C. Baez, Javier P. Muniain - Gauge Fields, Knots, and Gravity**[17] which is a magnificent book. The focus lies on introducing the mathematical tools needed to understand attempts to quantize gravity in a way that physicists understand.

13

Closing Words

In my humble opinion we are a long way from a theory that is able to explain all the things that we would like. Even the beautiful theory we developed in the major part of this book still has many loose ends that need clarification. In addition, we still have no clue how to derive the correct quantum theory of gravity as discussed in the last chapter.

Besides that there is experimental evidence, mostly from cosmology and astroparticle physics (Dark Matter and Dark Energy), which indicates that the present theories are not the end of the story.

I personally think there is still much to come and maybe a completely new framework is needed to overcome the present obstacles. Anyway, the future developments will be very interesting and I hope you will continue following the story and maybe contribute something yourself.

Part V
Appendices

A

Vector calculus

It is often useful in physics to describe the position of some object using three numbers $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$. This is what we call a vector $\vec{v}$ and denote by a little arrow above the letter. The three numbers are the components of the vector along the three coordinate axes. The first number tells us how far the vector in question goes in the x-direction, the second how far in the y-direction and the third how far in the z-direction. For example, $\vec{w} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}$ is a vector that points exclusively in the y-direction.

Vectors can be added

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \quad \vec{w} = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \quad \rightarrow \quad \vec{v} + \vec{w} = \begin{pmatrix} v_x + w_x \\ v_y + w_y \\ v_z + w_y \end{pmatrix} \tag{A.1}$$

and multiplied

$$\vec{v} \cdot \vec{w} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \cdot \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} = v_x w_x + v_y w_y + v_z w_z. \tag{A.2}$$

The result of this multiplication is not a vector, but a number (= a scalar), hence the name: **scalar product**. The scalar product of a vector with itself is directly related to its length:

$$\text{length}(\vec{v}) = \sqrt{\vec{v} \cdot \vec{v}}. \tag{A.3}$$

Take note that we can't simply write three quantities below each other between two brackets and expect it to be a vector. For example,
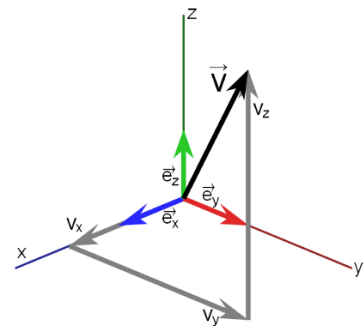
let's say we put the temperature T, the pressure P and the humidity H of a room between two brackets:

$$\begin{pmatrix} T \\ P \\ H \end{pmatrix}. \tag{A.4}$$

Nothing prevents us from doing so, but the result would be rather pointless and definitely not a vector, because there is no linear connection between these quantities that could lead to the mixing of these quantities. In contrast, the three position coordinates transform into each other, for example if we look at the vector from a different perspective[1]. Therefore writing the coordinates below each other between two big brackets is useful. Another example would be the momentum of some object. Again, the components mix if we look at the object from a different perspective and therefore writing it like the position vector is useful.

For the moment let's say a vector is a quantity that transforms exactly like the position vector $\vec{v}$. This means, if under some transformation we have $\vec{v} \rightarrow \vec{v}' = M\vec{v}$ any quantity that transforms like $\vec{w} \rightarrow \vec{w}' = M\vec{w}$ is a vector. Examples are the momentum or acceleration of some object.

We will encounter this idea quite often in physics. If we write quantities below each other between two brackets, they aren't necessarily vectors, but the quantities can transform into each other through some linear operation. This is often expressed by multiplication with a matrix.

A.1   Basis Vectors

We can make the idea of components along the coordinate axes more general by introducing basis vectors. Basis vectors are linearly independent[2] vectors of length one. In three dimensions we need three basis vectors and we can write every vector in terms of these basis vectors. An obvious choice is:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \tag{A.5}$$

[1] This will be made explicit in a moment.

[2] A set of vectors $\{\vec{a}, \vec{b}, \vec{c}\}$ is called linearly independent if the equation $c_1\vec{a} + c_2\vec{b} + c_3\vec{c} = 0$ is only true for $c_1 = c_2 = c_3 = 0$. This means that no vector can be written as a linear combination of the other vectors, because if we have $c_1\vec{a} + c_2\vec{b} + c_3\vec{c} = 0$ for numbers different than zero, we can write $c_1\vec{a} + c_2\vec{b} = -c_3\vec{c}$.

and an arbitrary three-dimensional vector $\vec{v}$ can be expressed in terms of these basis vectors

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3 = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad \text{(A.6)}$$

The numbers v_1, v_2, v_3 are called the components of $\vec{v}$. Take note that these components depend on the basis vectors.

The vector $\vec{w}$ we introduced above[3] can therefore be written as $\vec{w} = 0\vec{e}_1 + 4\vec{e}_2 + 0\vec{e}_3$. An equally good choice for the basis vectors would be

[3] $\vec{w} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}$

$$\tilde{e}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{e}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \tilde{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad \text{(A.7)}$$

In this basis the vector $\vec{w}$ "looks" quite different:

$$\vec{w} = 2\sqrt{2}\tilde{e}_1 - 2\sqrt{2}\tilde{e}_2 + 0\tilde{e}_3 = 2\sqrt{2}\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - 2\sqrt{2}\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}.$$
$$\text{(A.8)}$$

Therefore we can write $\vec{w}$ in terms of components with respect to this new basis as

$$\tilde{w} = \begin{pmatrix} 2\sqrt{2} \\ -2\sqrt{2} \\ 0 \end{pmatrix}.$$

This is not a different vector just a different description! To be precise, $\tilde{w}$ is the description of the vector $\vec{w}$ in a coordinate system that is rotated relative to the coordinate system we used in the first place.

A.2 Change of Coordinate Systems

The connection between different coordinate systems can be made precise through the use of matrices. Two different coordinate systems can mean that we have two different observers that look at our experiment from different perspectives or this can simply mean that **one** observer decides to use a different set of basis vectors. How are those descriptions related? To avoid complications, let's assume that the origin of the two coordinate systems coincide and both coordinate systems have the same z-axes. Therefore only the x and y coordinates
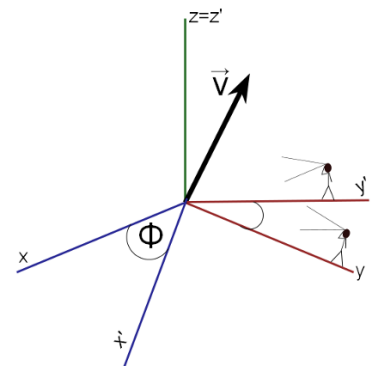
are different. Let's assume further that the position of something important in the experiment is described by the vector $\vec{v}$.

If the first observer sees the vector $\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$, we can compute how the same vector looks like in the coordinate system of the second observer $\vec{v} = \begin{pmatrix} v_{x'} \\ v_{y'} \\ v_{z'} \end{pmatrix}$ by using the usual trigonometric functions $\sin(\phi), \cos(\phi)$ and $\tan(\phi) = \frac{\sin(\phi)}{\cos(\phi)}$, as illustrated in Fig. A.1.
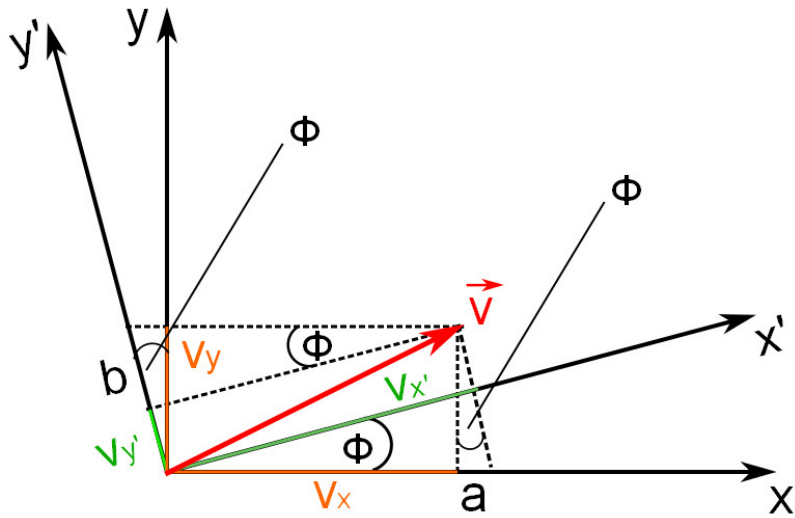
The relationship between v_x and $v_{x'}$ can be computed using

$$\cos(\phi) = \frac{v_{x'}}{v_x + a} \rightarrow v_{x'} = (v_x + a)\cos(\phi)$$

and

$$\tan(\phi) = \frac{a}{v_y} \rightarrow a = v_y \tan(\phi).$$

This yields

$$v_{x'} = (v_x + v_y \tan(\phi))\cos(\phi) = \left(v_x + v_y \frac{\sin(\phi)}{\cos(\phi)}\right)\cos(\phi)$$

$$= v_x \cos(\phi) + v_y \sin(\phi).$$

Analogously we can use

$$\cos(\phi) = \frac{v_y}{v_{y'} + b} \rightarrow v_{y'} = v_y \frac{1}{\cos(\phi)} - b$$

and

$$\tan(\phi) = \frac{b}{v_{x'}} \rightarrow b = v_{x'} \tan(\phi),$$

which yields using $\sin^2(\phi) + \cos^2(\phi) = 1$

$$v_{y'} = v_y \frac{1}{\cos(\phi)} - v_{x'} \tan(\phi) = v_y \frac{1}{\cos(\phi)} - (v_x \cos(\phi) + v_y \sin(\phi)) \frac{\sin(\phi)}{\cos(\phi)}$$

$$= v_y \frac{\sin^2(\phi) + \cos^2(\phi)}{\cos(\phi)} - v_x \sin(\phi) - v_y \frac{\sin^2(\phi)}{\cos(\phi)} = v_y \cos(\phi) - v_x \sin(\phi)$$

Therefore $v_{y'} = -v_x \sin(\phi) + v_y \cos(\phi)$.

We can write this using a rotation matrix:

$$\begin{pmatrix} v_{x'} \\ v_{y'} \\ v_{z'} \end{pmatrix} = R_z(\phi)\vec{v} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix}$$

$$= \begin{pmatrix} \cos(\phi)v_x + \sin(\phi)v_y \\ -\sin(\phi)v_x + \cos(\phi)v_y \\ v_z \end{pmatrix}. \tag{A.9}$$

We multiply each row of the matrix with the unrotated vector to compute the rotated vector. As already noted above, the component along the z-axis v_3 is the same for both observers. The matrix $R_z(\phi)$ describes a rotation by the angle ϕ about the z-axis.

A.3 Matrix Multiplication

Computations like this are tremendously simplified through the use of matrices. The rule for Matrix multiplication is always row times column. We can see the scalar product introduced above as a special case of this, if we interpret a vector as a matrix with one column and three rows (a 3×1 matrix). The scalar product of two vectors is then

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w} = \begin{pmatrix} v_x & v_y & v_z \end{pmatrix} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} = v_x w_x + v_y w_y + v_z w_z, \tag{A.10}$$

where the T denotes transposing, which means that every columns becomes a row and every row a column. Therefore, $\vec{v}^T$ is a matrix
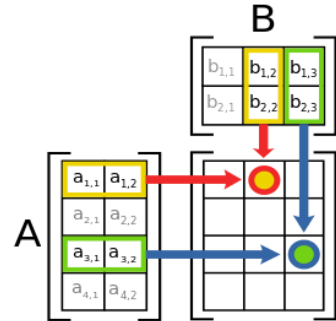
Fig. A.2: Schematic matrix multiplication. The important thing to keep in mind is **row times column**. The first index denotes the row number, the second the column number. In the example, the red element of the product matrix is $c_{1,2} = a_{1,1}b_{1,2} + a_{1,2}b_{2,2}$ and the blue element is $c_{3,3} = a_{3,1}b_{1,3} + a_{3,2}b_{2,3}$. In general $c_{i,j} = a_{i,k}b_{k,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \ldots$. Figure by Olivier Perrin (Bilou Wikimedia Commons) released under a CC BY-SA 3.0 licence: http://creativecommons.org/licenses/by-sa/3.0/deed.en . URL: http://commons.wikimedia.org/wiki/File:Matrix_multiplication_diagram_2.svg , Accessed: 28.1.2015

with 1 row and 3 columns. Written in this way the scalar product is once more a matrix product with row times columns.

Analogously, we get the matrix product of two matrices from the multiplication of each row of the matrix to the left with a column of the matrix to the right. This is explained in Fig. A.2. An explicit example for the multiplication of two matrices is

$$M = \begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix} \quad N = \begin{pmatrix} 0 & 1 \\ 4 & 8 \end{pmatrix}$$

$$MN = \begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 4 & 8 \end{pmatrix} = \begin{pmatrix} 2\cdot0+3\cdot4 & 2\cdot1+3\cdot8 \\ 1\cdot0+0\cdot4 & 1\cdot1+0\cdot8 \end{pmatrix} = \begin{pmatrix} 12 & 26 \\ 0 & 1 \end{pmatrix}$$
(A.11)

The rule to keep in mind is **row times column**. Take note that the multiplication of two matrices is not commutative, which means in general $MN \neq NM$.

A.4   Scalars

An important thing to notice is that the scalar product of two vectors has the same value for all observers. This can be seen as the definition of a scalar: A scalar is the same for all observers. This does not simply mean that every number is a scalar, because each component of a vector is a number, but as we have seen above a different number for different observers. In contrast the scalar product of two vectors must be the same for all observers. This follows from the fact that the scalar product of a vector with itself is directly related to the length of the vector. Changing the perspective or the location we choose to look at our experiment may not change the length of anything. The length of a vector is called an invariant for rotations, because it stays the same no matter how we rotate our system.

A.5   Right-handed and Left-handed Coordinate Systems

When we talked above about two observers, we implicitly assumed they agree in terms of the definition of their coordinate system. In fact, there are two possible choices, which are again related by matrix multiplication, but not by rotations. One observer may choose what we call a right-handed coordinate system and another observer what we call a left-handed coordinate system.

There is no way to rotate a left-handed into a right-handed coordinate system. Instead, such coordinate systems are related through a

Fig. A.3: Right-handed and left-handed coordinate system. Figure by Primalshell (Wikimedia Commons) released under a CC-BY-SA-3.0 licence: http://creativecommons.org/licenses/by-sa/3.0/deed.en. URL: http://commons.wikimedia.org/wiki/File:3D_Cartesian_Coodinate_Handedness.jpg , Accessed: 1.12.2014

reflection in a mirror. This means the descriptions in a right-handed and a left-handed coordinate system are related by a transformation of the form

$$
\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \rightarrow \begin{pmatrix} -v_1 \\ -v_2 \\ -v_3 \end{pmatrix} , \tag{A.12}
$$

which means we flip the sign of all spatial coordinates. The conventional name for this kind of transformation is **parity transformation**. We can describe a parity transformation by

$$
\vec{v} \rightarrow \vec{v}' = P\vec{v} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} -v_1 \\ -v_2 \\ -v_3 \end{pmatrix} . \tag{A.13}
$$

B

Calculus

B.1 Product Rule

The product rule

$$\frac{d\big(f(x)g(x)\big)}{dx} = \Big(\frac{df(x)}{dx}\Big)g(x) + f(x)\Big(\frac{dg(x)}{dx}\Big) \equiv f'g + fg' \quad \text{(B.1)}$$

follows directly from the definition of derivatives

$$
\begin{aligned}
\frac{d}{dx}[f(x)g(x)] &= \lim_{h\to 0}\frac{f(x+h)g(x+h) - f(x)g(x)}{h}\\
&= \lim_{h\to 0}\frac{[f(x+h)g(x+h) - f(x+h)g(x)] + [f(x+h)g(x) - f(x)g(x)]}{h}\\
&= \lim_{h\to 0} f(x+h)\frac{g(x+h) - g(x)}{h} + g(x)\frac{f(x+h) - f(x)}{h}\\
&= f(x)g'(x) + g(x)f'(x)
\end{aligned}
$$

B.2 Integration by Parts

A (likely apocryphal) story goes: when Peter Lax was awarded the National Medal of Science, the other recipients (presumably non-mathematicians) asked him what he did to deserve the Medal. Lax responded: " I integrated by parts."

- Willie Wong[1]

[1] Told on www.math.stackexchange.com

An important rule for integrals follows directly from the product rule. Integrating the product rule[2]

[2] See Eq. B.1 and we then use for the first term the fundamental theorem of calculus $\int_a^b dx\, h'(x) = h(b) - h(a)$.

$$\underbrace{\int_a^b dx \frac{d\Big(f(x)g(x)\Big)}{dx}}_{=f(x)g(x)\big|_a^b} = \int_a^b dx \left(\frac{df(x)}{dx}\right)g(x) + \int_a^b dx f(x)\left(\frac{dg(x)}{dx}\right) \quad \text{(B.2)}$$

and rearranging the terms yields

$$\int_a^b dx \left(\frac{df(x)}{dx}\right)g(x) = f(x)g(x)\Big|_a^b - \int_a^b dx f(x)\left(\frac{dg(x)}{dx}\right). \qquad \text{(B.3)}$$

This rule is particularly useful in physics when working with fields, because if we integrate over all space, i.e. $a = -\infty, b = \infty$, the boundary term vanishes $f(x)g(x)\Big|_{a=-\infty}^{b=\infty} = 0$, because all fields must vanish at infinity[3].

[3] This follows, because otherwise the total field energy would be infinity and such field configurations are non-physical.

B.3 The Taylor Series

The Taylor series is a formula that enables us to write any infinitely differentiable function in terms of a power series

$$f(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a) + \ldots \qquad \text{(B.4)}$$

- On the one hand we can use it if we want to know how we can write some function in terms of a series. This can be used for example to show that $e^{ix} = \cos(x) + i\sin(x)$.

- On the other hand we can use the Taylor series to get approximations for a function about a point. This is useful when we can neglect for some reasons higher order terms and don't need to consider infinitely many terms. If we want to evaluate a function $f(x)$ in some neighbouring point of a point y, say $y + \Delta y$, we can write[4]

[4] Don't let yourself get confused by the names of our variables here. In the formula above we want to evaluate the function at x by doing computations at a. Here we want to know something about f at $y + \Delta y$, by using information at y. To make the connection precise: $x = y + \Delta y$ and $a = y$.

$$f(y + \Delta y) = f(y) + (y + \Delta y - y)f'(y) + \ldots = f(y) + \Delta y f'(y) + \ldots . \tag{B.5}$$

This means we get an approximation for the function value at $y + \Delta y$, by evaluating the function at y. In the extreme case of an infinitesimal neighbourhood $\Delta y \to \epsilon$, the change of the function can be written by one (the linear) term of the Taylor series.

$$\Delta f = f(y + \epsilon) - f(y) = f(y) + \epsilon f'(y) + \ldots - f(y) = \epsilon f'(y) + \underbrace{\ldots}_{\approx 0 \text{ for } \epsilon^2 \approx 0}$$
$$\text{(B.6)}$$

This formula is one of the most useful mathematical tools and we can derive it using the fundamental theorem of calculus and integration by parts. The fundamental theorem tells us

$$\int_a^x dt\, f'(t) = f(x) - f(a) \quad \rightarrow \quad f(x) = f(a) + \int_a^x dt\, f'(t). \quad \text{(B.7)}$$

We can rewrite the second term by integrating by parts, because we have of course an implicit 1 in front of $f'(t) = 1f'(t)$, which we can use as a second function in Eq. B.3: $g'(t) = 1$. The rule for integration by parts tells us that we can rewrite an integral $\int_a^b v'u = vu\big|_a^b + \int vu'$ by integrating one term and differentiating the other. Here we integrate $g'(t) = 1$ and differentiate $f'(t)$, i.e. in the formula $g' = v'$ and $f' = u$. This yields

$$f(x) = f(a) + \int_a^x dt\, f'(t) = f(a) + g(t)f'(t)\Big|_a^x - \int_a^x dt\, g(t)f''(t).$$
$$\text{(B.8)}$$

Now we need to know what $g(t)$ is. At this point the only information we have is $g'(t) = 1$, but there are infinitely many functions with this derivative: For any constant c the function $g = t + c$ satisfies $g'(t) = 1$. Our formula becomes particularly useful[5] for $g = t - x$, i.e. we use minus the upper integration boundary $-x$ as our constant c. Then we have for the second term in the equation above

$$g(t)f'(t)\Big|_a^x = (t-x)f'(t)\Big|_a^x = \underbrace{(x-x)}_{=0}f'(x) - (a-x)f'(a) = (x-a)f'(a)$$
$$\text{(B.9)}$$

and the formula now reads

$$f(x) = f(a) + (x-a)f'(a) + \int_a^x dt\, (x-t)f''(t). \quad \text{(B.10)}$$

We can then evaluate the last term once more using integration by parts, now with[6] $v' = (x-t)$ and $u = f''(t)$:

[6] Take note that integrating $v' = (x-t)$ yields a minus sign: $\rightarrow v = -\frac{1}{2}(x-t)^2 + d$, because our variable here is t and with some constant d we choose to be zero.

$$\rightarrow \int_a^x dt\, \underbrace{(x-t)}_{=v'}\underbrace{f''(t)}_{=u} = \underbrace{-\frac{1}{2}(x-t)^2}_{=v}\underbrace{f''(t)}_{=u}\Big|_a^x - \int_a^x dt\, \underbrace{\left(-\frac{1}{2}(x-t)^2\right)}_{=v}\underbrace{f'''(t)}_{=u'}$$

where the boundary term is again simple

$$-\frac{1}{2}(x-t)^2 f''(t)\Big|_a^x = -\frac{1}{2}\underbrace{(x-x)^2}_{=0}f''(x) + \frac{1}{2}(x-a)^2 f''(a).$$

This gives us the formula

$$f(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a) + \int_a^x dt\, \frac{1}{2}(x-t)^2 f'''(t).$$
$$\text{(B.11)}$$

We could go on and integrate the last term by parts, but the pattern should be visible by now. The Taylor series can be written in a more compact form using the mathematical sign for a sum $\sum$:

$$
\begin{aligned}
f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)(x-a)^n}{n!} \\
&= \frac{f^{(0)}(a)(x-a)^0}{0!} + \frac{f^{(1)}(a)(x-a)^1}{1!} + \frac{f^{(2)}(a)(x-a)^2}{2!} \\
&\quad + \frac{f^{(3)}(a)(x-a)^3}{3!} + \dots ,
\end{aligned}
\tag{B.12}
$$

where $f^{(n)}$ denotes the n-th derivative of f, e.g. $f^{(2)} = f''$ and $n!$ is the factorial of n, i.e. $n! = 1 \cdot 2 \cdot 3 \dots n$. For example for $n = 5$ we have $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$, $2! = 2 \cdot 1 = 2$ and by definition $0! = 1$. Series are the topic of the next section.

B.4 Series

In the last section we stumbled upon a very important formula that includes an infinite sum. In this section some basic tricks for sum manipulation and some very important series will be introduced.

B.4.1 Important Series

In the last section we learned that we can write every infinitely differentiable function as a series. Let's start with maybe the most important function: The exponential function e^x. The Taylor series for the exponential function can be written right away

$$
e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},
\tag{B.13}
$$

by using the defining feature of the exponential function that the derivative is the exponential function itself: $(e^x)' = e^x$, evaluating the Taylor series about $a = 0$ and using $e^0 = 1$. This yields the Taylor series (Eq. B.12) for the exponential function:

$$
e^x = \sum_{n=0}^{\infty} \frac{e^0(x-0)^n}{n!} = \sum_{n=0}^{\infty} \frac{x^n}{n!}
\tag{B.14}
$$

This series can be seen as a definition of e^x.

Two other important, infinitely differentiable functions are $\sin(x)$ and $\cos(x)$. We can compute the Taylor series for these functions, by using $(\sin(x))' = \cos(x)$, $(\cos(x))' = -\sin(x)$, $\cos(0) = 1$ and

$\sin(0) = 0$.

$$\sin(x) = \sum_{n=0}^{\infty} \frac{\sin^{(n)}(0)(x-0)^n}{n!}.$$

Because of $\sin(0) = 0$ every term with even n vanishes, which we can use if we split the sum. Observe that

$$\sum_{n=0}^{\infty} n = \sum_{n=0}^{\infty} (2n+1) + \sum_{n=0}^{\infty} (2n)$$

$$1 + 2 + 3 + 4 + 5 + 6 \ldots = 1 + 3 + 5 + \ldots \quad + 2 + 4 + 6 + \ldots \quad \text{(B.15)}$$

Splitting the sum for $\sin(x)$ yields

$$\sin(x) = \sum_{n=0}^{\infty} \frac{\sin^{(2n+1)}(0)(x-0)^{2n+1}}{(2n+1)!} + \underbrace{\sum_{n=0}^{\infty} \frac{\sin^{(2n)}(0)(x-0)^{2n}}{(2n)!}}_{=0}$$

$$= \sum_{n=0}^{\infty} \frac{\sin^{(2n+1)}(0)(x-0)^{2n+1}}{(2n+1)!}. \qquad \text{(B.16)}$$

Every even derivative of $\sin(x)$, i.e. $\sin^{(2n)}$ is again $\sin(x)$ (with possibly a minus sign in front of it) and therefore the second term vanishes because of $\sin(0) = 0$. Every uneven derivative of $\sin(x)$ is $\cos(x)$, with possibly a minus sign in front of it. We have

$$\sin(x)^{(1)} = \cos(x)$$
$$\sin(x)^{(2)} = \cos'(x) = -\sin(x)$$
$$\sin(x)^{(3)} = -\sin'(x) = -\cos(x)$$
$$\sin(x)^{(4)} = -\cos'(x) = \sin(x)$$
$$\sin(x)^{(5)} = \sin'(x) = \cos(x). \qquad \text{(B.17)}$$

The pattern is therefore $\sin^{(2n+1)}(x) = (-1)^n \cos(x)$, as you can check by putting some integer values for n into the formula[7]. We can therefore rewrite Eq. B.16 as

[7] $\sin^{(1)}(x) = \sin^{(2\cdot0+1)}(x) = (-1)^{\cdot0}\cos(x) = \cos(x)$, $\sin^{(3)}(x) = \sin^{(2\cdot1+1)}(x) = (-1)^1\cos(x) = -\cos(x)$

$$\sin(x) = \sum_{n=0}^{\infty} \frac{\sin^{(2n+1)}(0)(x-0)^{2n+1}}{(2n+1)!}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n \cos(0)(x-0)^{2n+1}}{(2n+1)!}$$

$$\underset{\cos(0)=1}{=} \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n+1}}{(2n+1)!}. \qquad \text{(B.18)}$$

This is the Taylor series for $\sin(x)$, which again can be seen as a definition of $\sin(x)$. Analogous we can derive

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n}}{(2n)!}, \qquad (B.19)$$

because this time uneven derivatives are proportional to $\sin(0) = 0$.

B.4.2 Splitting Sums

In the last section we used a trick that is quite useful in many computations. There we used the example

$$\sum_{n=0}^{\infty} n = \sum_{n=0}^{\infty} (2n+1) + \sum_{n=0}^{\infty} (2n)$$

$$1 + 2 + 3 + 4 + 5 + 6 \ldots = 1 + 3 + 5 + \ldots \quad + 2 + 4 + 6 + \ldots, \quad (B.20)$$

to motivate how we can split any sum in terms of even and uneven integers. $2n$ is always an even integer, whereas $2n+1$ is always an uneven integer. We already saw in the last section that this can be useful, but let's look at another example. What happens if we split the exponential series with complex argument ix?

$$e^{ix} = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!}$$

$$= \sum_{n=0}^{\infty} \frac{(ix)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(ix)^{2n+1}}{(2n+1)!}. \qquad (B.21)$$

This can be rewritten using that $(ix)^{2n} = i^{2n} x^{2n}$ and $i^{2n} = (i^2)^n = (-1)^n$. In addition we have of course $i^{2n+1} = i \cdot i^{2n} = i(-1)^n$. Then we have

$$e^{ix} = \sum_{n=0}^{\infty} \frac{(ix)^n}{n!}$$

$$= \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n+1}}{(2n+1)!}$$

$$= \underbrace{\sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n}}{(2n)!}}_{=\cos(x) \text{ see Eq. B.19}} + i \underbrace{\sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n+1}}{(2n+1)!}}_{=\sin(x) \text{ see Eq. B.18}}$$

$$= \cos(x) + i \sin(x). \qquad (B.22)$$

This is famously known as **Euler's formula**.

B.4.3 Einstein's Sum Convention

Sums are very common in physics and writing the big sum sign $\sum$ all the time can be quite cumbersome. For this reason a clever man introduced a convention, called Einstein sum convention. According to this convention every time an index appears twice in some term, like n in the sums above, an implicit sum is understood. This means

$$a_n b_n \equiv \sum_n a_n b_n. \tag{B.23}$$

Other examples are

$$a_n b_n c_m \equiv \sum_n a_n b_n c_m. \tag{B.24}$$

$$a_m b_n c_m \equiv \sum_m a_m b_n c_m. \tag{B.25}$$

but

$$a_n b_m \neq \sum_n a_n b_m, \tag{B.26}$$

because in general $m \neq n$. An index without a partner is called a **free index**, an index with a partner a **dummy index**, for reasons that will be explained in the next section.

For example in the sum $a_n b_n c_m \equiv \sum_n a_n b_n c_m$, the index n is a dummy index, but m is a free index. Equivalently, in $a_m b_n c_m \equiv \sum_m a_m b_n c_m$, the index m is a dummy index and n is free.

B.5 Index Notation

B.5.1 Dummy Indices

It is important to take note that the name of indices with a partner plays absolutely no role. Renaming $n \rightarrow k$, changes absolutely nothing[8], as long as n is contracted

$$a_n b_n c_m = a_k b_k c_m \equiv \sum_n a_n b_n c_m \equiv \sum_k a_k b_k c_m. \tag{B.27}$$

On the other hand free indices can not be renamed freely. For example, $m \rightarrow q$ can make quite a difference because there must be some term on the other side of the equation with the same free index. This means when we look at a term like $a_n b_n c_m$ isolated, we must always take into account that there might be other terms with the same free index m that must be renamed, too. Let's look at an example

$$F_i = \epsilon_{ijk} a_j b_k. \tag{B.28}$$

[8] Of course we can't change an index into another **type** of index. For example, we can change $i \rightarrow j$ but not $i \rightarrow \mu$, because Greek indices like μ are always summed from 0 to 3 and Roman indices, like i from 1 to 3.

A new thing that appears here is that some object, here ϵ_{ijk}, is allowed to carry more than one index, but don't let that bother you, because we will come back to this in a moment. Therefore, if we look at $\epsilon_{ijk}a_j b_k$ we can change the names of j and k as we like, because these indices are contracted. For example $j \to u$, $k \to z$, which yields $\epsilon_{iuz}a_u b_z$ is really the same. On the other hand i is not a dummy index and we can't rename it $i \to m$: $\epsilon_{muz}a_u b_z$, because then our equation would read

$$F_i = \epsilon_{muz}a_u b_z. \tag{B.29}$$

This may seem pedantic at this point, because it is clear that we need to rename i at F_i, too in order to get something sensible, but more often than not will we look at isolated terms and it is important to know what we are allowed to do without changing anything.

B.5.2   Objects with more than One Index

Now, let's talk about objects with more than one index. Objects with two indices are simply matrices. The first index tells us which row and the second index which column we should pick our value from. For example

$$M_{ij} \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}. \tag{B.30}$$

This means for example that M_{12} is the object in the first row in the second column.

We can use this to write matrix multiplication using indices. The product of two matrices is

$$MN \equiv (MN)_{ij} = M_{ik}N_{kj}. \tag{B.31}$$

On the left hand side we have the element in row i in column j of the product matrix (MN), which we get from multiplying the i-th row of M with the j-th column of N. The index k appears twice and therefore an implicit sum is assumed. One can give names to objects with three or more indices (tensors). For the purpose of this book two are enough and we will discuss only one exception, which is the topic of one of the next sections.

B.5.3   Symmetric and Antisymmetric Indices

A matrix is said to be symmetric if $M_{ij} = M_{ji}$. This means in our two dimensional example $M_{12} = M_{21}$ and an example for a symmetric

matrix is

$$\begin{pmatrix} 9 & 3 \\ 3 & 17 \end{pmatrix}. \tag{B.32}$$

A matrix is called **totally antisymmetric** if $M_{ij} = -M_{ji}$ for **all** i, j holds. An example would be.

$$\begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix}. \tag{B.33}$$

Take note that the diagonal elements must vanish here, because $M_{11} = -M_{11}$, which is only satisfied for $M_{11} = 0$ and analogously for M_{22}.

B.5.4 Antisymmetric $\times$ Symmetric Sums

An important observation is that every time we have a sum over something symmetric in its indices multiplied with something anti-symmetric in the same indices, the result is zero:

$$\sum_{ij} a_{ij} b_{ij} = 0 \tag{B.34}$$

if $a_{ij} = -a_{ji}$ and $b_{ij} = b_{ji}$ holds for all i, j. We can see this by writing

$$\sum_{ij} a_{ij} b_{ij} = \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} + \sum_{ij} a_{ij} b_{ij} \right) \tag{B.35}$$

As explained earlier we are free to rename our dummy indices $i \to j$ and $j \to i$, which we use in the second term

$$\to \sum_{ij} a_{ij} b_{ij} = \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} + \sum_{ij} a_{ji} b_{ji} \right) \tag{B.36}$$

Then we use the symmetry of b_{ij} and antisymmetry of a_{ij}, to switch the indices in the second term, which yields[9]

$$\to \sum_{ij} a_{ij} b_{ij} = \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} + \sum_{ij} \underbrace{a_{ji}}_{=-a_{ij}} \underbrace{b_{ji}}_{=b_{ij}} \right)$$

$$= \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} - \sum_{ij} a_{ij} b_{ij} \right) = 0. \tag{B.37}$$

[9] If this looks like a cheap trick to you, you should compute some explicit examples to see that this is really true.

B.5.5 Two Important Symbols

One of the most important matrices is of course the **unit matrix**. In two dimensions we have

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{B.38}$$

In index notation the unit matrix is called the Kronecker symbol, denoted δ_{ij}, which is then defined for arbitrary dimensions by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{B.39}$$

The Kronecker symbol is symmetric because $\delta_{ij} = \delta_{ji}$.

Equally important is the **Levi-Civita symbol** ϵ_{ijk}, which is defined in two dimensions as follows:

$$\epsilon_{ij} = \begin{cases} 1 & \text{if } (i,j) = \{(1,2)\}, \\ 0 & \text{if } i = j, \\ -1 & \text{if } (i,j) = \{(2,1)\}. \end{cases} \tag{B.40}$$

In matrix form, we can write it as

$$\epsilon_{ij} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{B.41}$$

In three dimensions the Levi-Civita symbol is

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } (i,j,k) = \{(1,2,3),(2,3,1),(3,1,2)\}, \\ 0 & \text{if } i = j \text{ or } j = k \text{ or } k = i, \\ -1 & \text{if } (i,j,k) = \{(1,3,2),(3,2,1),(2,1,3)\}. \end{cases} \tag{B.42}$$

and in four dimensions

$$\epsilon_{ijkl} = \begin{cases} 1 & \text{if } (i,j,k,l) \text{ is an even permutation of}\{(1,2,3,4)\}, \\ -1 & \text{if } (i,j,k,l) \text{ is an uneven permutation of}\{(1,2,3,4)\}, \\ 0 & \text{otherwise.} \end{cases} \tag{B.43}$$

For example $(1,2,4,3)$ is an uneven (because we make one change) and $(2,1,4,3)$ is an even permutation (because we make two changes) of $(1,2,3,4)$.

The Levi-Civita symbol is **totally anti-symmetric** because if we change two indices, we always get, by definition, a minus sign: $\epsilon_{ijk} = -\epsilon_{jik}$, $\epsilon_{ijk} = -\epsilon_{ikj}$ etc. or in two dimensions $\epsilon_{ij} = -\epsilon_{ji}$.

C

Linear Algebra

Many computations can be simplified by using matrices and tricks from the linear algebra toolbox. Therefore, let's look at some basic transformations.

C.1 Basic Transformations

The **complex conjugate of a matrix** is defined as

$$M_{ij}^{\star} = \begin{pmatrix} M_{11}^{\star} & M_{12}^{\star} \\ M_{21}^{\star} & M_{22}^{\star} \end{pmatrix}, \tag{C.1}$$

which means we simply take the complex conjugate of each element[1].

The **transpose of a matrix** is defined by $M_{ij}^{T} = M_{ji}$, in matrix form

$$M_{ij} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \quad \rightarrow \quad M_{ij}^{T} = \begin{pmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{pmatrix}, \tag{C.2}$$

which means we swap columns and rows of the matrix. An important consequence of this definition and the definition of the product of two matrices is that we have $(MN)^{T} \neq M^{T}N^{T}$. Instead $(MN)^{T} = N^{T}M^{T}$, which means by transposing we switch the position of two matrices in a product. We can see this directly in index notation

$$MN \equiv (MN)_{ij} = M_{ik}N_{kj}$$

$$(MN)^{T} \equiv ((MN)_{ij})^{T} = (MN)_{ji} = (M_{ik}N_{kj})^{T}$$

$$(M_{ik}N_{kj})^{T} = M_{ik}^{T}N_{kj}^{T} = M_{ki}N_{jk} = N_{jk}M_{ki} \equiv N^{T}M^{T}, \tag{C.3}$$

[1] Recall that the complex conjugate of a complex number $z = a + ib$, where a is the real part and b the imaginary part, is simply $z^{\star} = a - ib$.

where in the last step we use the general rule that in matrix notation we always multiply rows of the left matrix with columns of the right matrix. To write this in matrix notation, we change the position of the two terms to $N_{jk}M_{ki}$, which is rows of the left matrix times columns of the right matrix, as it should be and we can write in matrix notation $N^T M^T$.

Take note that in index notation we can always change the position of the objects in question freely, because for example M_{ki} and N_{jk} are just individual elements of the matrices, i.e. ordinary numbers.

C.2 Matrix Exponential Function

We already derived how the exponential function looks as a series, and therefore we can define what we mean when we put a **matrix into the exponential function**:

e^M, with an arbitrary matrix M, is defined by the series

$$e^M = \sum_{n=0}^{\infty} \frac{M^n}{n!}. \tag{C.4}$$

It is important to take note that in general $e^M e^N \neq e^{M+N}$. The identity $e^M e^N = e^{M+N}$ is only correct if $MN = NM$.

C.3 Determinants

The determinant of a matrix is a rather unintuitive, but immensely useful notion. For example, if the determinant of some matrix is non-zero, we automatically know that the matrix is invertible[2]. Unfortunately proving this lies beyond the scope of this text and the interested reader is referred to the standard texts about linear algebra.

The determinant of a 3×3 matrix can be defined using the Levi-Civita symbol

$$\det(A) = \sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{k=1}^{3} \epsilon_{ijk} A_{1i} A_{2j} A_{3k} \tag{C.5}$$

and analogously for n-dimensions

$$\det(A) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \dots \sum_{i_n=1}^{n} \epsilon_{i_1 i_2 \dots i_n} A_{1i_1} A_{2i_2} \dots A_{ni_n}. \tag{C.6}$$

It is instructive to look at an explicit example in two dimensions:

[2] A matrix M is invertible, if we can find an inverse matrix, denoted by M^{-1}, with $M^{-1}M = 1$.

$$\det(A) = \det \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} = (3 \cdot 2) - (5 \cdot 1) = 1.$$

For a general three dimensional matrix, we have

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = a_1(b_2 c_3 - b_3 c_2) - a_2(b_1 c_3 - b_3 c_1) + a_3(b_1 c_2 - b_2 c_1).$$

$$(C.7)$$

C.4 Eigenvalues and Eigenvectors

Two very important notions from linear algebra that are used all over physics are eigenvalues and eigenvectors. The eigenvectors $\vec{v}$ and eigenvalues λ are defined for each matrix M by the equation

$$M\vec{v} = \lambda\vec{v}. \tag{C.8}$$

The important thing is that we have on both sides of the equation the same vector $\vec{v}$. In words this equation means that the vector $\vec{v}$ remains, up to a constant λ, unchanged if multiplied with the matrix M. To each eigenvector we have a corresponding eigenvalue. There are quite sophisticated computational schemes for finding the eigenvectors and eigenvalues of a matrix and the details can be found in any book about linear algebra.

To get a feeling for the importance of these notions think about rotations. We can describe rotations by matrices and the eigenvector of a rotation matrix defines the rotational axis.

C.5 Diagonalization

Eigenvectors and eigenvalues can be used to bring matrices into diagonal form, which can be quite useful for computations and physical interpretations. It can be shown that any diagonalizable matrix M can be rewritten in the form[3]

$$M = N^{-1}DN, \tag{C.9}$$

where the matrix N consists of the eigenvectors as its column and D is diagonal with the eigenvalues of M on its diagonal:

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = N^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} N = \left(\vec{v}_1, \vec{v}_2 \right)^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \left(\vec{v}_1, \vec{v}_2 \right).$$

$$(C.10)$$

[3] In general, a transformation of the form $M' = N^{-1}MN$ refers to a basis change. M' is the matrix M in another coordinate system. Therefore, the result of this section is that we can find a basis where M is particularly simple, i.e. diagonal.

D

Additional Mathematical Notions

D.1   Fourier Transform

The idea of the Fourier transform is similar to the idea that we can express any vector $\vec{v}$ in terms of basis vectors[1] ($\vec{e}_1, \vec{e}_2, \vec{e}_3$). In ordinary Euclidean space the most common choice is

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad , \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad , \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \qquad (\text{D.1})$$

and an arbitrary three-dimensional vector $\vec{v}$ can be expressed in terms of these basis vectors

$$\vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = v_1\vec{e}_1 + v_2\vec{e}_2 + v_3\vec{e}_3 = v_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + v_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + v_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (\text{D.2})$$

The idea of the Fourier transform is that we can do the same thing with functions[2]. For periodic functions such a basis is given by $\sin(kx)$ and $\cos(kx)$. This means we can write every periodic function $f(x)$ as

$$f(x) = \sum_{k=0}^{\infty} \left(a_k \cos(kx) + b_k \sin(kx) \right) \qquad (\text{D.3})$$

with constant coefficients a_k and b_k.

An arbitrary (not necessarily periodic) function can be written in terms of the basis e^{ikx} and e^{-ikx}, but this time with an integral instead of a sum[3]

$$f(x) = \int_0^\infty dk \left(a_k e^{ikx} + b_k e^{-ikx} \right),$$ (D.4)

which we can also write as

$$f(x) = \int_{-\infty}^\infty dk f_k e^{-ikx}.$$ (D.5)

The expansion coefficients f_k are often denoted $\tilde{f}(k)$, which is then called the Fourier transform of $f(x)$.

D.2 Delta Distribution

In some sense, the delta distribution is to integrals what the Kronecker delta[4] is to sums. We can use the Kronecker delta δ_{ij} to pick one specific term of any sum. For example, consider

$$\sum_{i=1}^3 a_i b_j = a_1 b_j + a_2 b_j + a_3 b_j$$ (D.6)

and let's say we want to pick the second term of the sum. We can do this using the Kronecker delta δ_{2i}, because then

$$\sum_{i=1}^3 \delta_{2i} a_i b_j = \underbrace{\delta_{21}}_{=0} a_1 b_j + \underbrace{\delta_{22}}_{=1} a_2 b_j + \underbrace{\delta_{23}}_{=0} a_3 b_j = a_2 b_j.$$ (D.7)

Or more general

$$\sum_{i=1}^3 \delta_{ik} a_i b_j = a_k b_j.$$ (D.8)

The delta distribution $\delta(x - y)$ is defined as an object that has exactly this property:

$$\int dx f(x) \delta(x - y) = f(y).$$ (D.9)

Completely analogous to the Kronecker delta, the delta distribution picks one term from the integral[5]. In addition, we can use this analogy to motivate from the equality

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij}$$ (D.10)

the equality

$$\frac{\partial f(x_i)}{\partial f(x_j)} = \delta(x_i - x_j).$$ (D.11)

This is of course by no means a proof, but this equality can be shown in a rigorous way, too. There is a lot more one can say about this object, but for the purpose of this book it is enough to understand what

the delta distribution does. In fact, this is how the delta distribution
was introduced in the first place by Dirac.

Bibliography

John C. Baez and Javier P. Muniain. *Gauge Fields, Knots, and Gravity*. World Scientific Pub Co Inc, 1st edition, 9 1994. ISBN 9789810220341.

Alessandro Bettini. *Introduction to Elementary Particle Physics*. Cambridge University Press, 2nd edition, 4 2014. ISBN 9781107050402.

Ta-Pei Cheng. *Relativity, Gravitation and Cosmology: A Basic Introduction*. Oxford University Press, 2nd edition, 1 2010. ISBN 9780199573646.

Predrag Cvitanović. *Group Theory: Birdtracks, Lie's, and Exceptional Groups*. Princeton University Press, 7 2008. ISBN 9780691118369.

Paul A. M. Dirac and Physics. *Lectures on Quantum Mechanics*. Dover Publications, 1st edition, 3 2001. ISBN 9780486417134.

Albert Einstein. The foundation of the general theory of relativity. 1916.

Albert Einstein and Francis A. Davis. *The Principle of Relativity*. Dover Publications, reprint edition, 6 1952. ISBN 9780486600819.

Richard P. Feynman and Albert R. Hibbs. *Quantum Mechanics and Path Integrals: Emended Edition*. Dover Publications, emended editon edition, 7 2010. ISBN 9780486477220.

Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics, Volume 3*. Addison Wesley, 1st edition, 1 1971. ISBN 9780201021189.

Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics: Volume 2*. Addison-Wesley, 1st edition, 2 1977. ISBN 9780201021172.

Daniel Fleisch. *A Student's Guide to Vectors and Tensors*. Cambridge University Press, 1st edition, 11 2011. ISBN 9780521171908.

Jon Fripp, Deborah Fripp, and Michael Fripp. *Speaking of Science*. Newnes, 1st edition, 4 2000. ISBN 9781878707512.

William Fulton and Joe Harris. *Representation Theory: A First Course*. Springer, 1st corrected edition, 8 1999. ISBN 9780387974958.

Howard Georgi. *Lie Algebras In Particle Physics: from Isospin To Unified Theories (Frontiers in Physics)*. Westview Press, 2 edition, 10 1999. ISBN 9780738202334.

Herbert Goldstein, Charles P. Poole Jr., and John L. Safko. *Classical Mechanics*. Addison-Wesley, 3rd edition, 6 2001. ISBN 9780201657029.

David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Prentice Hall, 2nd edition, 4 2004. ISBN 9780131118928.

David J. Griffiths. *Introduction to Electrodynamics*. Addison-Wesley, 4th edition, 10 2012. ISBN 9780321856562.

Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1st edition, 1 1984. ISBN 9780471887416.

Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148.

Robert D. Klauber. *Student Friendly Quantum Field Theory*. Sandtrove Press, 2nd edition, 12 2013. ISBN 9780984513956.

Cornelius Lanczos. *The Variational Principles of Mechanics*. Dover Publications, 4th edition, 3 1986. ISBN 9780486650678.

C. S. Lewis. *Reflections on the Psalms*. HarperOne, reprint edition, 2 2017. ISBN 9780062565488.

Michele Maggiore. *A Modern Introduction to Quantum Field Theory*. Oxford University Press, 1st edition, 2 2005. ISBN 9780198520740.

Franz Mandl and Graham Shaw. *Quantum Field Theory*. Wiley, 2nd edition, 5 2010. ISBN 9780471496847.

Rollo May. *The Courage to Create*. W. W. Norton and Company, reprint edition, 3 1994. ISBN 9780393311068.

Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. W. H. Freeman, 1st edition, 9 1973. ISBN 9780716703440.

David Morin. *Introduction to Classical Mechanics: With Problems and Solutions*. Cambridge University Press, 1 edition, 2 2008. ISBN 9780521876223.

Matthew Robinson. *Symmetry and the Standard Model*. Springer, 1st edition, August 2011. ISBN 978-1-4419-8267-4.

Robert S. Root-Bernstein and Michele M. Root-Bernstein. *Sparks of Genius*. Mariner Books, 1st edition, 8 2001. ISBN 9780618127450.

Lewis H. Ryder. *Quantum Field Theory*. Cambridge University Press, 2nd edition, 6 1996. ISBN 9780521478144.

J. J. Sakurai. *Modern Quantum Mechanics*. Addison Wesley, 1st edition, 9 1993. ISBN 9780201539295.

Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 1 edition, 12 2013. ISBN 9781107034730.

Lee Smolin. *Three Roads to Quantum Gravity*. Basic Books, 3 edition, 8 2017. ISBN 9780465094547.

Michael Spivak. *A Comprehensive Introduction to Differential Geometry, Vol. 1, 3rd Edition*. Publish or Perish, 3rd edition, 1 1999. ISBN 9780914098706.

Andrew M. Steane. An introduction to spinors. *ArXiv e-prints*, December 2013.

John Stillwell. *Naive Lie Theory*. Springer, 1st edition, August 2008. ISBN 978-0387782140.

Edwin F. Taylor and John Archibald Wheeler. *Spacetime Physics*. W. H. Freeman, 2nd edition, 3 1992. ISBN 9780716723271.

Frank Wilczek. Riemann-einstein structure from volume and gauge symmetry. *Phys. Rev. Lett.*, 80:4851–4854, Jun 1998. DOI: 10.1103/PhysRevLett.80.4851.

Harry Woolf, editor. *Some Strangeness in the Proportion*. Addison-Wesley, 1st edition, 2 1981. ISBN 9780201099249.

Anthony Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 1st edition, 3 2003. ISBN 9780691010199.

Anthony Zee. *Einstein Gravity in a Nutshell*. Princeton University Press, 1st edition, 5 2013. ISBN 9780691145587.

Index

action, 96
adjoint representation, 170
angular momentum, 104
annihilation operator, 214
anticommutator, 218

B field, 239
Baker-Campbell-Hausdorff formula, 41
basis generators, 42
basis spinors, 217
Bohmian mechanics, 197
boost, 67
boost x-axis matrix, 68
boost y-axis matrix, 68
boost z-axis matrix, 68
boson, 174
bra, 189
broken symmetry, 153

Cartan generators, 53
Casimir elements, 53
Casimir operator SU(2), 54
charge conjugation, 86
classical electrodynamics, 239
closure, 41
commutator, 41
commutator quantum field theory, 120
commutator quantum mechanics, 118
complex Klein-Gordon field, 124
conjugate momentum, 111, 119, 211
continuous symmetry, 26
Coulomb potential, 243
coupling constants, 123
covariance, 21, 122
covariant derivative, 138
creation operator, 214

dagger, 34, 92
delta distribution, 278
determinant, 274
diagonalization, 275
Dirac equation, 127
Dirac equation solution, 216
dotted index, 75
doublet, 131, 144
Dyson series, 224

E field, 239
Ehrenfest theorem, 234
eigenvalue, 275
eigenvector, 275
Einstein, 11
Einstein summation convention, 17
Einstein's Sum Convention, 269
electric charge, 142
electric charge density, 142
electric field, 239
electric four-current, 142, 239
electrodynamics, 239
elementary particles, 90
energy-momentum relation, 178
energy-momentum tensor, 107
Euclidean space, 20
Euler's formula, 268
Euler-Lagrange equation, 99
    field theory, 101
    particle theory, 101
expectation value, 181, 233
extrema of functionals, 97

Fermat's principle, 96
fermion, 174
fermion mass, 162
field energy, 108
field momentum, 108
field-strength tensor, 239