

Undergraduate Lecture Notes in Physics

Jakob Schwichtenberg

Physics from Symmetry

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

Professor Emeritus, University of Colorado, Boulder, CO, USA

William Brantley

Professor, Furman University, Greenville, SC, USA

Michael Fowler

Professor, University of Virginia, Charlottesville, VA, USA

Morten Hjorth-Jensen

Professor, University of Oslo, Oslo, Norway

Michael Inglis

Professor, SUNY Suffolk County Community College, Long Island, NY, USA

Heinz Klose

Professor Emeritus, Humboldt University Berlin, Germany

Helmy Sherif

Professor, University of Alberta, Edmonton, AB, Canada

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

Jakob Schwichtenberg

Physics from Symmetry

 Springer

Jakob Schwichtenberg
Karlsruhe
Germany

ISSN 2192-4791 ISSN 2192-4805 (electronic)
Undergraduate Lecture Notes in Physics
ISBN 978-3-319-19200-0 ISBN 978-3-319-19201-7 (eBook)
DOI 10.1007/978-3-319-19201-7

Library of Congress Control Number: 2015941118

Springer Cham Heidelberg New York Dordrecht London
© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media
(www.springer.com)

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

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

- **Albert Einstein**¹

¹ As quoted in Jon Fripp, Deborah Fripp, and Michael Fripp. *Speaking of Science*. Newnes, 1st edition, 4 2000. ISBN 9781878707512

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 readers 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 book² 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 reading³.

² Starting with Chap. A. In addition, the corresponding appendix chapters are mentioned when a new mathematical concept is used in the text.

³ 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

I want to thank everyone who helped me create this book. I am especially grateful to Fritz Waitz, whose comments, ideas and corrections have made this book so much better. I am also very indebted to Arne Becker and Daniel Hilpert for their invaluable suggestions, comments and careful proofreading. Thanks to Robert Sadlier for his help with the English language and to Jakob Karalus for his comments.

I want to thank Marcel Köpke for for many insightful discussions and Silvia Schwichtenberg and Christian Nawroth for their support.

Finally, my greatest debt is to my parents who always supported me and taught me to value education above all else.

If you find an error in the text I would appreciate a short email to errors@jakobschwichtenberg.com . All known errors are listed at <http://physicsfromsymmetry.com/errata> .

Contents

Part I Foundations

1	Introduction	3
1.1	What we Cannot Derive	3
1.2	Book Overview	5
1.3	Elementary Particles and Fundamental Forces	7
2	Special Relativity	11
2.1	The Invariant of Special Relativity	12
2.2	Proper Time	14
2.3	Upper Speed Limit	16
2.4	The Minkowski Notation	17
2.5	Lorentz Transformations	19
2.6	Invariance, Symmetry and Covariance	20

Part II Symmetry Tools

3	Lie Group Theory	25
3.1	Groups	26
3.2	Rotations in two Dimensions	29
3.2.1	Rotations with Unit Complex Numbers	31
3.3	Rotations in three Dimensions	33
3.3.1	Quaternions	34
3.4	Lie Algebras	38
3.4.1	The Generators and Lie Algebra of $SO(3)$	41
3.4.2	The Abstract Definition of a Lie Algebra	44
3.4.3	The Generators and Lie Algebra of $SU(2)$	45
3.4.4	The Abstract Definition of a Lie Group	47
3.5	Representation Theory	49
3.6	$SU(2)$	53
3.6.1	The Finite-dimensional Irreducible Representations of $SU(2)$	53

3.6.2	The Casimir Operator of $SU(2)$	56
3.6.3	The Representation of $SU(2)$ in one Dimension	57
3.6.4	The Representation of $SU(2)$ in two Dimensions	57
3.6.5	The Representation of $SU(2)$ in three Dimensions	58
3.7	The Lorentz Group $O(1,3)$	58
3.7.1	One Representation of the Lorentz Group	61
3.7.2	Generators of the Other Components of the Lorentz Group	64
3.7.3	The Lie Algebra of the Proper Orthochronous Lorentz Group	66
3.7.4	The $(0,0)$ Representation	67
3.7.5	The $(\frac{1}{2},0)$ Representation	68
3.7.6	The $(0,\frac{1}{2})$ Representation	69
3.7.7	Van der Waerden Notation	70
3.7.8	The $(\frac{1}{2},\frac{1}{2})$ Representation	75
3.7.9	Spinors and Parity	78
3.7.10	Spinors and Charge Conjugation	81
3.7.11	Infinite-Dimensional Representations	82
3.8	The Poincare Group	84
3.9	Elementary Particles	85
3.10	Appendix: Rotations in a Complex Vector Space	87
3.11	Appendix: Manifolds	87
4	The Framework	91
4.1	Lagrangian Formalism	91
4.1.1	Fermat's Principle	92
4.1.2	Variational Calculus - the Basic Idea	92
4.2	Restrictions	93
4.3	Particle Theories vs. Field Theories	94
4.4	Euler-Lagrange Equation	95
4.5	Noether's Theorem	97
4.5.1	Noether's Theorem for Particle Theories	97
4.5.2	Noether's Theorem for Field Theories - Spacetime Symmetries	101
4.5.3	Rotations and Boosts	104
4.5.4	Spin	105
4.5.5	Noether's Theorem for Field Theories - Internal Symmetries	106
4.6	Appendix: Conserved Quantity from Boost Invariance for Particle Theories	108
4.7	Appendix: Conserved Quantity from Boost Invariance for Field Theories	109

Part III The Equations of Nature

5	Measuring Nature	113
5.1	The Operators of Quantum Mechanics	113
5.1.1	Spin and Angular Momentum	114
5.2	The Operators of Quantum Field Theory	115
6	Free Theory	117
6.1	Lorentz Covariance and Invariance	117
6.2	Klein-Gordon Equation	118
6.2.1	Complex Klein-Gordon Field	119
6.3	Dirac Equation	120
6.4	Proca Equation	123
7	Interaction Theory	127
7.1	$U(1)$ Interactions	129
7.1.1	Internal Symmetry of Free Spin $\frac{1}{2}$ Fields	129
7.1.2	Internal Symmetry of Free Spin 1 Fields	131
7.1.3	Putting the Puzzle Pieces Together	132
7.1.4	Inhomogeneous Maxwell Equations and Minimal Coupling	134
7.1.5	Charge Conjugation, Again	135
7.1.6	Noether's Theorem for Internal $U(1)$ Symmetry	136
7.1.7	Interaction of Massive Spin 0 Fields	138
7.1.8	Interaction of Massive Spin 1 Fields	138
7.2	$SU(2)$ Interactions	139
7.3	Mass Terms and Unification of $SU(2)$ and $U(1)$	145
7.4	Parity Violation	152
7.5	Lepton Mass Terms	156
7.6	Quark Mass Terms	160
7.7	Isospin	161
7.7.1	Labelling States	162
7.8	$SU(3)$ Interactions	164
7.8.1	Color	166
7.8.2	Quark Description	167
7.9	The Interplay Between Fermions and Bosons	169

Part IV Applications

8	Quantum Mechanics	173
8.1	Particle Theory Identifications	174
8.2	Relativistic Energy-Momentum Relation	174
8.3	The Quantum Formalism	175
8.3.1	Expectation Value	177

8.4	The Schrödinger Equation	178
8.4.1	Schrödinger Equation with External Field	180
8.5	From Wave Equations to Particle Motion	180
8.5.1	Example: Free Particle	181
8.5.2	Example: Particle in a Box	181
8.5.3	Dirac Notation	185
8.5.4	Example: Particle in a Box, Again	187
8.5.5	Spin	187
8.6	Heisenberg's Uncertainty Principle	191
8.7	Comments on Interpretations	192
8.8	Appendix: Interpretation of the Dirac Spinor Components	194
8.9	Appendix: Solving the Dirac Equation	199
8.10	Appendix: Dirac Spinors in Different Bases	200
8.10.1	Solutions of the Dirac Equation in the Mass Basis	202
9	Quantum Field Theory	205
9.1	Field Theory Identifications	206
9.2	Free Spin 0 Field Theory	207
9.3	Free Spin $\frac{1}{2}$ Theory	212
9.4	Free Spin 1 Theory	215
9.5	Interacting Field Theory	215
9.5.1	Scatter Amplitudes	216
9.5.2	Time Evolution of States	216
9.5.3	Dyson Series	220
9.5.4	Evaluating the Series	221
9.6	Appendix: Most General Solution of the Klein-Gordon Equation	225
10	Classical Mechanics	227
10.1	Relativistic Mechanics	229
10.2	The Lagrangian of Non-Relativistic Mechanics	230
11	Electrodynamics	233
11.1	The Homogeneous Maxwell Equations	234
11.2	The Lorentz Force	235
11.3	Coulomb Potential	237
12	Gravity	239
13	Closing Words	245
 Part V Appendices		
A	Vector calculus	249

A.1	Basis Vectors	250
A.2	Change of Coordinate Systems	251
A.3	Matrix Multiplication	253
A.4	Scalars	254
A.5	Right-handed and Left-handed Coordinate Systems	254
B	Calculus	257
B.1	Product Rule	257
B.2	Integration by Parts	257
B.3	The Taylor Series	258
B.4	Series	260
B.4.1	Important Series	260
B.4.2	Splitting Sums	262
B.4.3	Einstein's Sum Convention	262
B.5	Index Notation	263
B.5.1	Dummy Indices	263
B.5.2	Objects with more than One Index	264
B.5.3	Symmetric and Antisymmetric Indices	264
B.5.4	Antisymmetric \times Symmetric Sums	265
B.5.5	Two Important Symbols	265
C	Linear Algebra	267
C.1	Basic Transformations	267
C.2	Matrix Exponential Function	268
C.3	Determinants	268
C.4	Eigenvalues and Eigenvectors	269
C.5	Diagonalization	269
D	Additional Mathematical Notions	271
D.1	Fourier Transform	271
D.2	Delta Distribution	272
	Bibliography	273
	Index	277

Part I

Foundations

"The truth always turns out to be simpler than you thought."

Richard P. Feynman

as quoted by

K. C. Cole. *Sympathetic Vibrations*.

Bantam, reprint edition, 10 1985.

ISBN 9780553342345

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 theories¹, 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 experiments² with high accuracy that there is no fourth generation.
- 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 first fundamental representations of the double cover of the Poincare 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 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.

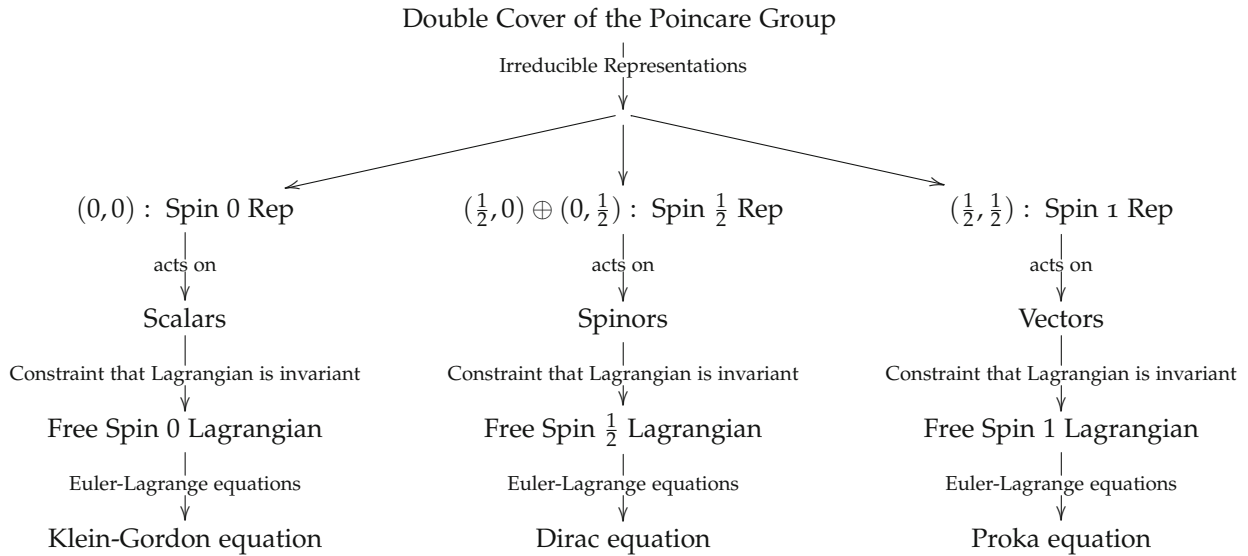
In addition, there are two things that can't be derived from symmetry, but which must be taken into account in order to get a sensible theory:

- 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.
- For similar reasons we can show that if particles with half-integer spin would behave exactly as particles with integer spin there wouldn't be any stable matter in this universe. Therefore, *something* must be different and we are left with only one possible, sensible choice³ which turns out to be correct. This leads to the notion of Fermi-Dirac statistics for particles with half-integer spin and Bose-Einstein statistics for particles with integer spin. Particles with half-integer spin are often called Fermions and there can never be two of them in exactly the same state. In contrast, for particles with integer spin, often called Bosons, this is possible.

³ We use the anticommutator instead of the commutator as the starting point for quantum field theory. This prevents our theory from being unbounded from below.

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



This book uses **natural units**, which means setting the Planck constant $\hbar = 1$ and the speed of light $c = 1$. This is conventional in fundamental theories, because it avoids a lot of unnecessary writing. For applications the constants need to be added again to return to standard SI units.

The 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 **Poincare group**. To be able to utilize them, the mathematical theory that enables us to work with symmetries is introduced. This branch of mathematics is called group theory. We will derive the irreducible representations of the Poincare group⁴, 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. Spin is on the one hand a label for different kinds of particles/fields and on the other hand can be seen as something like internal angular momentum.

Afterwards, the **Lagrangian formalism** is introduced, which makes working with symmetries in a physical context very convenient. The central object is the **Lagrangian**, which we will be able to

⁴ To be technically correct: We will derive the representations of the **double-cover** of the Poincare group instead of the Poincare 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 Sec. 3.3.1 in detail.

derive from symmetry considerations for different physical systems. In addition, the **Euler-Lagrange equations** are derived, which enable us to derive the equations of motion from a given Lagrangian. Using the irreducible representations of the Poincare 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 Poincare 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 the spin $\frac{1}{2}$ and spin 1 field. 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 **spontaneous symmetry breaking** and the **Higgs mechanism**. These enable us to describe particles with mass⁵.

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

$$[\hat{x}_i, \hat{p}_i] = i\delta_{ij} \quad (1.1)$$

and quantum field theory

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

We continue by taking the non-relativistic⁶ 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 motion⁷ and Eq. 1.2. Afterwards, we take interactions into account, by taking a closer look

⁵ Before spontaneous symmetry breaking, terms describing mass in the Lagrangian spoil the symmetry and are therefore forbidden.

⁶ 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.

⁷ The Klein-Gordon, Dirac, Proca and Maxwell equations.

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 behaviour 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 for macroscopic objects (classical mechanics) or elementary particles with low energy (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 behaviour of these particles:

- Fermions are responsible for matter
- Bosons for the forces of nature

This means, for example, that atoms consist of fermions⁸, but the electromagnetic-force is mediated by bosons, called photons. One of the most dramatic consequences of this is that there is stable matter. If there could be infinitely many fermions in the same state, there would be no stable matter at all, as we will discuss in Chap. 6.

There are four presently known fundamental forces

⁸ Atoms consist of electrons, protons and neutrons, which are all fermions. But take note that protons and neutrons are not fundamental and consist of quarks, which are fermions, too.

- The electromagnetic force, which is mediated by massless **photons**.
- The weak force, which is mediated by massive W^+ , W^- and **Z-bosons**.
- The strong force, which is mediated by massless **gluons**.
- Gravity, which is (maybe) mediated by **gravitons**.

Some of the corresponding bosons are massless and some are not, which 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 interact via the electromagnetic force.
- For the weak force, the charge is called¹⁰ **isospin**. 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**, which are for example electrons and neutrinos. The difference is that quarks interact via the strong force, which means carry color and leptons do not. There are three quark and lepton **generations**, which consist each of two particles:

	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	$-e$	$-\frac{1}{2}$	-

⁹ All charges have a beautiful common origin that will be discussed in Chap. 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.

In general, the 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, as we will derive in Sec. 4.5.4. Bosons carry integer spin, whereas fermions carry half-integer spin. The fundamental fermions we listed above have spin $\frac{1}{2}$. The fundamental bosons have spin 1. In addition, 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 sign¹¹. 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 photon¹² 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 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 frames¹. Albert Einstein was the first who 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 very interesting, very 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 same² 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

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

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

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.

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 indicate that it is correct.

2.1 The Invariant of Special Relativity

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.

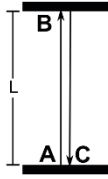


Fig. 2.1: Illustration of the thought experiment

Imagine, we have a spectator, standing at the origin of his coordinate 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** is³

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

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

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 $\Delta t = \frac{\Delta s}{v}$

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, relative to the first spectator⁴. For simplicity let's assume that the origin of this second spectator 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

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

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

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

We assume movement along the x-axis, therefore

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

and equally of course

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

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 .

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

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

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

We therefore conclude, using Eq. 2.6

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

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

$$\begin{aligned} c\Delta t' &= 2\sqrt{\left(\frac{1}{2}\Delta x'\right)^2 + L^2} \\ \rightarrow (c\Delta t')^2 &= 4\left(\left(\frac{1}{2}\Delta x'\right)^2 + L^2\right) \end{aligned}$$

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.

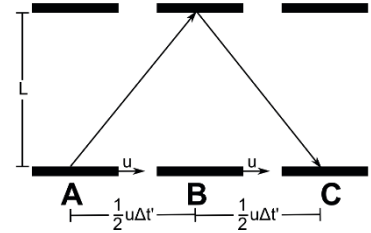


Fig. 2.2: Illustration of the thought experiment for a moving spectator. The second spectator moves to the left and therefore the first spectator (and the experiment) moves relative to him to the right.

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

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

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

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.

So finally, we arrive⁵ at

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

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

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

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

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

In addition, we learned in this Sec. 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 in the next section.

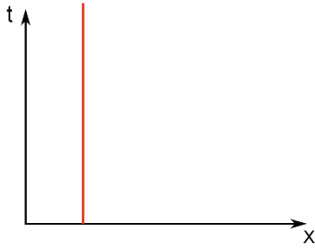


Fig. 2.3: World line of an object at rest. The position of the object stays the same as time goes on.

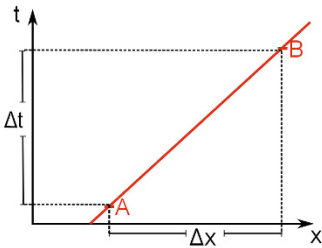


Fig. 2.4: World line of a moving object with two events A and B. The distance travelled between A and B is Δx and the time that passed the events is Δt .

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 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

$$(\Delta s)^2 = (c\Delta t)^2 - (\Delta x)^2 \quad (2.14)$$

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

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

This is one of the most famous phenomena of special relativity and commonly called **time-dilation**. Time-intervals are observer dependent, as is spatial distance. The clocks tick differently for each different observer and therefore both 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

$$(\Delta s)^2 = (c\Delta t')^2, \quad (2.17)$$

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

$$(\Delta s)^2 = (c\Delta \tau)^2, \quad (2.18)$$

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.

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

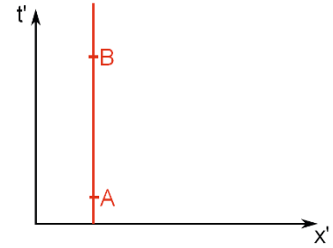


Fig. 2.5: World line of the same moving object, as observed from someone moving with the same constant speed as the object. The distance travelled between A and B is for this observer $\Delta x' = 0$.

proper time is sensible. In mathematical terms this requires we make the transition to infinitesimal intervals $\Delta \rightarrow d$:

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

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 that Δs^2 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 time⁶, 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

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

$$\begin{aligned} \Delta s_{\min}^2 = 0 &= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \\ &\rightarrow (c\Delta t)^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \\ &\rightarrow c^2 = \frac{(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2}{(\Delta t)^2}. \end{aligned} \quad (2.20)$$

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

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

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.

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

$$\rightarrow c^2 = v^2. \quad (2.22)$$

This means nothing can move faster than speed 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 Minkowski⁷

⁷ In a speech at the 80th Assembly of German Natural Scientists and Physicians (21 September 1908)

We can rewrite the invariant of special relativity

$$ds^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2 \quad (2.23)$$

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

$$\begin{aligned} ds^2 &= \eta^{\mu\nu} dx_\mu dx_\nu = \eta^{00}(x_0)^2 + \eta^{11}(dx_1)^2 + \eta^{22}(dx_2)^2 + \eta^{33}(dx_3)^2 \\ &= dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2. \end{aligned} \quad (2.24)$$

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 indices⁸, like μ, ν or σ , are always summed from 0 to 3: $x_\mu y_\mu = \sum_{i=0}^3 x_i y_i$.
- 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
- 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 is⁹ $\eta^{\mu\nu} = \text{diag}(1, -1, -1, -1)$)

⁸ In contrast, Roman indices like i, j, k are always summed: $x_i x_i \equiv \sum_{i=1}^3 x_i x_i$ from 1 to 3. Much later in the book we will use capital Roman letters like A, B, C that are summed from 1 to 8.

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

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

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

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

$$\begin{aligned} (ds)^2 &= dx_\mu \eta^{\mu\nu} dx_\nu = \begin{pmatrix} dx_0 & dx_1 & dx_2 & dx_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} dx_0 \\ dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} \\ &= dx_0^2 - dx_1^2 - dx_2^2 - dx_3^2 \end{aligned} \quad (2.26)$$

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 Euclidean¹⁰ space the squared (shortest) distance between two points is given by¹¹

$$\begin{aligned} (ds)^2 &= dx_i \delta_{ij} dx_j = \begin{pmatrix} dx_1 & dx_2 & dx_3 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} dx_1 \\ dx_2 \\ dx_3 \end{pmatrix} \\ &= (ds)^2 = (dx_1)^2 + (dx_2)^2 + (dx_3)^2 \end{aligned} \quad (2.27)$$

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.

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_{ij} . 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 $\eta^{\mu\nu}$. 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 itself¹²

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

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

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

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

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

There is another, notational convention to make computations more streamlined. We define a four-vector with upper index as¹³

$$x^\mu \equiv \eta^{\mu\nu} x_\nu \quad (2.29)$$

or equally

$$y^\nu \equiv \eta^{\mu\nu} y_\mu \underbrace{=}_{\text{The Minkowski metric is symmetric } \eta^{\mu\nu} = \eta^{\nu\mu}} \eta^{\nu\mu} y_\mu \quad (2.30)$$

Therefore, we can write the scalar product as¹⁴

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

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.

¹³ Four-vectors with a lower index are often called covariant and four-vectors with an upper index contravariant.

¹⁴ The name of the index makes no difference. For more information about this have a look at appendix B.5.1.

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:

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

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:

$$dx_\mu \rightarrow dx'_\mu = \Lambda_\mu^\sigma dx_\sigma. \quad (2.33)$$

Then we can write the invariance condition as

$$\begin{aligned} (ds)^2 &= (ds')^2 \\ \rightarrow dx \cdot dx &\stackrel{!}{=} dx' \cdot dx' \rightarrow dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} dx'_\mu dx'_\nu \eta^{\mu\nu} \underbrace{=}_{\text{Eq. 2.33}} \Lambda_\mu^\sigma dx_\sigma \Lambda_\nu^\gamma dx_\gamma \eta^{\mu\nu} \\ &\quad \underbrace{\rightarrow}_{\text{Renaming dummy indices}} dx_\mu dx_\nu \eta^{\mu\nu} \stackrel{!}{=} \Lambda_\sigma^\mu dx_\mu \Lambda_\gamma^\nu dx_\nu \eta^{\sigma\gamma} \\ &\quad \underbrace{\rightarrow}_{\text{Because the equation holds for arbitrary } dx_\mu} \eta^{\mu\nu} \stackrel{!}{=} \Lambda_\sigma^\mu \Lambda_\gamma^\nu \eta^{\sigma\gamma} \end{aligned} \quad (2.34)$$

¹⁵ If you wonder about the transpose here have a look at appendix C.1.

Or written in matrix notation¹⁵

$$\eta = \Lambda^T \eta \Lambda \quad (2.35)$$

This is the condition that transformations Λ_μ^ν 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 transformations¹⁶ O that leave the scalar product of Euclidean space invariant¹⁷

¹⁶ The name O will become clear in a moment.

¹⁷ 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.

¹⁸ 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.

¹⁹ Recall that the length of a vector is given by the scalar product of a vector with itself.

²⁰ This is explained in appendix A.5.

²¹ 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.

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

Therefore¹⁸ $O^T 1 O \stackrel{!}{=} 1$ 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 product¹⁹. Additionally we must include that rotations do not change the orientation²⁰ 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 inversions²¹

We define the **Lorentz transformations** as those transformations that leave the scalar product of Minkowski spacetime invariant, which means nothing more than respecting the conditions 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: $x_\mu y^\mu = x_\mu y_\nu \eta^{\mu\nu}$. 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 $F = F(A, B, C, \dots)$ that depends on different quantities A, B, C, \dots . If we transform $A, B, C, \dots \rightarrow A', B', C', \dots$ and we have

$$F(A', B', C', \dots) = F(A, B, C, \dots) \quad (2.37)$$

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 translations 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.

²² Edwin F. Taylor and John Archibald Wheeler. *Spacetime Physics*. W. H. Freeman, 2nd edition, 3 1992. ISBN 9780716723271

²³ Daniel Fleisch. *A Student's Guide to Vectors and Tensors*. Cambridge University Press, 1st edition, 11 2011. ISBN 9780521171908

²⁴ Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

²⁵ Anthony Zee. *Einstein Gravity in a Nutshell*. Princeton University Press, 1st edition, 5 2013. ISBN 9780691145587

Further Reading Tips

- **E. Taylor and J. Wheeler - Spacetime Physics: Introduction to Special Relativity**²² is a very good book to start with.
- **D. Fleisch - A Student's Guide to Vectors and Tensors**²³ 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**²⁴ is another good source for the mathematics needed in special relativity.
- **A. Zee - Einstein Gravity in a nutshell**²⁵ is a book about general relativity, but has many great explanations regarding special relativity, too.

Part II

Symmetry Tools

"Numbers measure size, groups measure symmetry."

Mark A. Armstrong
in Groups and Symmetry.
Springer, 2nd edition, 2 1997.
ISBN 9780387966755

3

Lie Group Theory

Chapter Overview

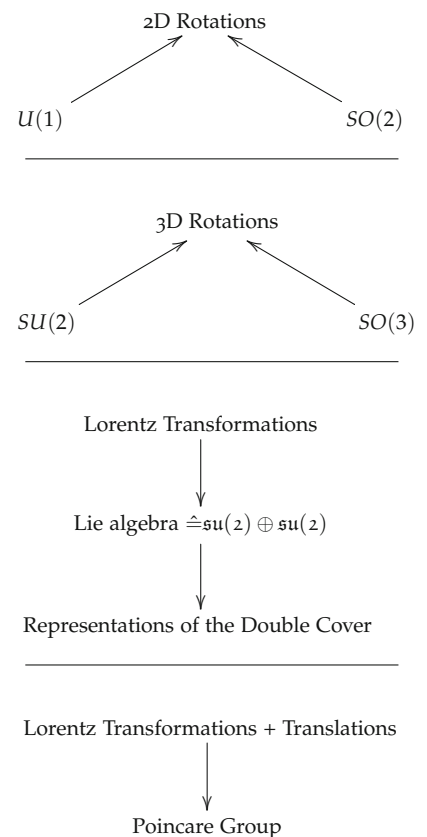
The final goal of this chapter is the derivation of the **fundamental representations of the double cover of the Poincare group**, which is assumed to be the fundamental symmetry group of spacetime. These fundamental representations are the tools needed to describe all elementary particles, each representation 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 into Lie theory we introduce two ways for describing rotations in two dimensions:

- the 2×2 rotation matrix and
- the unit complex numbers.

Then we will try to find a similar second description of rotations in three dimensions, which leads us to a very important group, called¹ **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 double covers the Poincare group. We will always 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

This diagram explains the structure of this chapter. You should come back here whenever you feel lost. There is no need to spend much time here at a first encounter.



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.

fundamental about the **Lorentz group**, which is an important part of the Poincare 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 is then called the Poincare group. The Poincare 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 Poincare group, which we will use 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 type of Group Theory is **Lie Theory**, which deals with continuous symmetries, as we encounter them often in nature.

Symmetry is defined as invariance under a set 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:

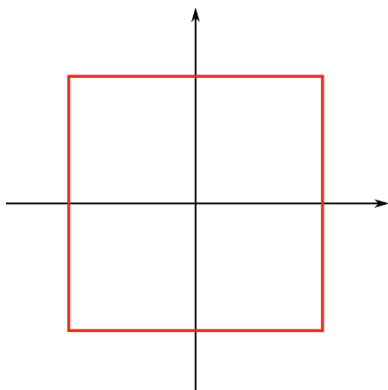


Fig. 3.1: Illustration of the square

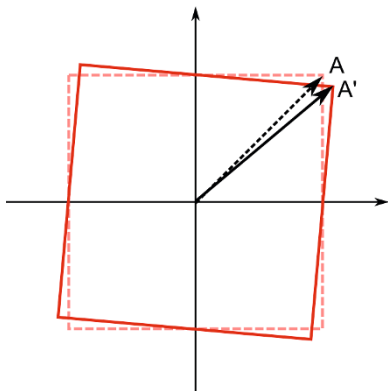


Fig. 3.2: Illustration of the square, rotated by 5°

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 if 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 instance, 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.

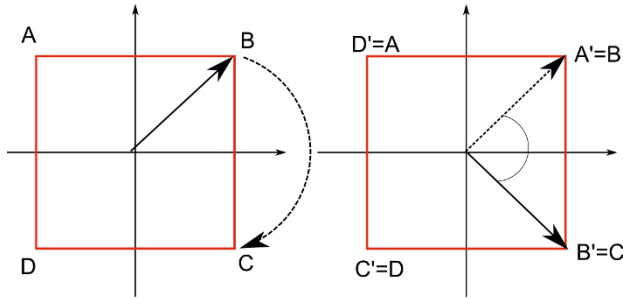


Fig. 3.3: Illustration of the square rotated by 90°

Another 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.

In mathematics, there are clearly objects other than just geometric shapes, and one can find symmetries for different kinds of objects, too. For instance, considering vectors, one can look at the set of transformations that leave the length of any vector unchanged. For this reason, the definition of symmetry I gave at the beginning was very general: Symmetry means invariance under a transformation. Happily, there is **one** mathematical theory, called group theory, that lets us work with **all** kinds of symmetries²

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

- 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° .

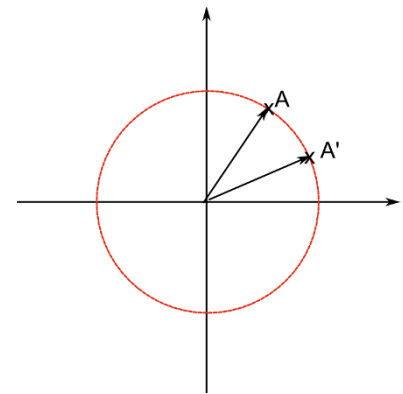


Fig. 3.4: Illustration of the rotation of the unit circle. For arbitrary rotations about the origin, all rotated points lie again in the initial set.

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.
- The combination of transformations must be associative³. 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:

$$R(110^\circ)R(40^\circ)R(90^\circ) = R(110^\circ)(R(40^\circ)R(90^\circ)) = R(110^\circ)R(130^\circ) \quad (3.1)$$

and

$$R(110^\circ)R(40^\circ)R(90^\circ) = (R(110^\circ)R(40^\circ))R(90^\circ) = R(150^\circ)R(90^\circ) \quad (3.2)$$

This property is called **associativity**.

- 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 matrices⁴ and the rule for combining the group elements (the corresponding rotation matrices) would be ordinary matrix multiplication. Nevertheless, there are often different ways of describing the same thing⁵ and group theory enables us to study this very systematically. The branch of group theory that deals with different descriptions of the same transformations is called **representation theory**, which is the topic of Sec. 3.5.

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

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

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.

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 is a set G , together with a binary operation \circ defined on G if the group (G, \circ) satisfies the following axioms⁶

- Closure: For all $g_1, g_2 \in G$, $g_1 \circ g_2 \in G$
- Identity: There exists an identity element $e \in G$ such that for all $g \in G$, $g \circ e = g = e \circ g$
- Inverses: 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 symmetry group. For Minkowski spacetime, the object that is left invariant is the Minkowski metric⁷ and the corresponding symmetry group is called Poincare 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 into Group theory, we start with an easy, but important, example. What are transformations in two dimensions that leave the length of any vector unchanged? After thinking about it for a while, we come up with⁸ 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 matrices⁹, which are of the form

⁶ 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 useful results of group theory to learn more about the transformations in question.

⁷ Recall, this is the tool which we use to compute distances and lengths in Minkowski space.

⁸ 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.

⁹ For an explicit derivation of these matrices have a look a look at appendix A.2.

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

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

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

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 $a \rightarrow a'$, the equation

$$a' \cdot a' \stackrel{!}{=} a \cdot a \quad (3.5)$$

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

$$a \cdot a = a^T a \rightarrow a'^T a' = (Oa)^T Oa = a^T O^T Oa \stackrel{!}{=} a^T a = a \cdot a, \quad (3.6)$$

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

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

where I denotes the unit matrix¹⁰. You can check that the well-known rotation and reflection matrices we cited above fulfil exactly this condition¹¹. This condition for two dimensional matrices defines the $O(2)$ group, the group of all¹² orthogonal 2×2 matrices. We can find a subgroup of this group that includes only rotations, by taking notice of the fact that it follows from the condition in Eq. 3.7 that

$$\begin{aligned} \det(O^T O) &\stackrel{!}{=} \det(I) = 1 \\ \rightarrow \det(O^T O) &= \det(O^T) \det(O) \stackrel{!}{=} \det(I) = 1 \\ \rightarrow (\det(O))^2 &\stackrel{!}{=} 1 \rightarrow \det(O) \stackrel{!}{=} \pm 1 \end{aligned} \quad (3.8)$$

The transformations of the group with $\det(O) = 1$ are rotations¹³ and the two conditions

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

$$\det O = 1 \quad (3.10)$$

$$^{10} I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

¹¹ With the matrix from Eq. 3.3 we have $R_\theta^T R =$

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} = \begin{pmatrix} \cos^2(\theta) + \sin^2(\theta) & 0 \\ 0 & \sin^2(\theta) + \cos^2(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \checkmark$$

¹² 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.

¹³ 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.

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-handed¹⁴ coordinate system must stay right-handed. In the language of linear algebra this means that the determinant of our matrices must be +1.

3.2.1 Rotations with Unit Complex Numbers

There is a quite 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 condition¹⁵ $|z|^2 = z^*z = 1$).

The unit complex numbers form a group, called¹⁶ $U(1)$ under ordinary complex number multiplication, as you can check by looking at the group axioms. To make the connection with the group definitions for $O(3)$ and $SO(3)$ introduced above, we write the condition as¹⁷

$$U^*U = 1. \quad (3.11)$$

Another way to write a unit complex number is¹⁸

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

because then

$$R_\theta^* R_\theta = e^{-i\theta} e^{i\theta} = (\cos(\theta) - i \sin(\theta))(\cos(\theta) + i \sin(\theta)) = 1 \quad (3.13)$$

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

$$z \rightarrow z' = e^{i90^\circ} z = \underbrace{(\cos(90^\circ))}_{=0} + \underbrace{i \sin(90^\circ)}_{=1} (3 + 5i) = i(3 + 5i) = 3i - 5 \quad (3.14)$$

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 ϕ . 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.

We can make the connection to the previous description by representing complex numbers by real 2×2 matrices. We define

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

¹⁴ 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.

¹⁵ The $*$ symbol denotes complex conjugation: $z = a + ib \rightarrow z^* = a - ib$

¹⁶ The U stands for unitary, which means the condition $U^*U = 1$

¹⁷ For more general information about the definition of groups involving a complex product, have a look at the appendix in Sec. 3.10.

¹⁸ 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 : $\text{Re}(z) = a$ and b the imaginary part: $\text{Im}(z) = b$. In Euler's formula $\cos(\theta)$ is the real part, and $\sin(\theta)$ the imaginary part of R_θ .

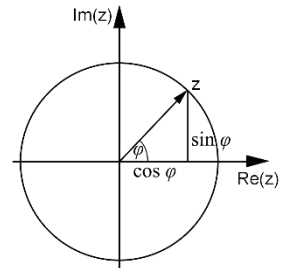


Fig. 3.5: The unit complex numbers lie on the unit circle in the complex plane.

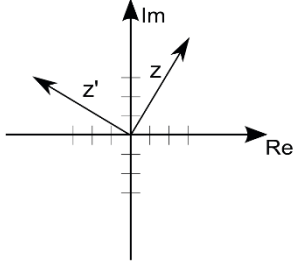


Fig. 3.6: Rotation of a complex number, by multiplication with a unit complex number

You can check that these matrices fulfil

$$1^2 = 1, \quad i^2 = -1, \quad 1i = i1 = i. \quad (3.16)$$

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

$$\begin{aligned} R_\theta &= \cos(\theta) + i \sin(\theta) = \cos(\theta) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin(\theta) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \end{aligned} \quad (3.17)$$

By making the map

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

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

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

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

$$\begin{aligned} z' &= \begin{pmatrix} a' & -b' \\ b' & a' \end{pmatrix} = R_\theta z = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta)a + \sin(\theta)b & -\cos(\theta)b + \sin(\theta)a \\ -\sin(\theta)a + \cos(\theta)b & \sin(\theta)b + \cos(\theta)a \end{pmatrix}. \end{aligned} \quad (3.19)$$

By comparing we get

$$\rightarrow a' = \cos(\theta)a + \sin(\theta)b \quad (3.20)$$

$$\rightarrow b' = -\sin(\theta)a + \cos(\theta)b, \quad (3.21)$$

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

$$\begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \cos(\theta)a + \sin(\theta)b \\ -\sin(\theta)a + \cos(\theta)b \end{pmatrix} = \begin{pmatrix} a' \\ b' \end{pmatrix}. \quad (3.22)$$

We see that both representations do exactly the same thing and mathematically speaking we have an isomorphism¹⁹ between $SO(2)$ and

¹⁹ An isomorphism is a one-to-one map Π that preserves the product structure $\Pi(g_1)\Pi(g_2) = \Pi(g_1g_2) \quad \forall g_1, g_2 \in G$.

$U(1)$. This is a very important discovery and we will elaborate on such lines of thoughts in the following chapters.

Next we want to describe rotations in three dimensions and find similarly two descriptions for rotations in three dimensions²⁰.

3.3 Rotations in three Dimensions

The standard methods to rotate vectors in three dimensions uses 3×3 rotation matrices:

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

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

Analogous to the definition for $SO(2)$, these matrices form a basis²¹ for a group, called $SO(3)$. This means an arbitrary element of $SO(3)$ can be written as a linear combination of the above matrices.

If we want to rotate the vector

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

around the z-axis²², we multiply it with the corresponding rotation matrix

$$R_z(\theta)\vec{v} = \begin{pmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ -\sin(\theta) \\ 0 \end{pmatrix}. \quad (3.24)$$

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 quaternions²³, have exactly 3 degrees of freedom.

²⁰ 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.

²¹ This notion is explained in appendix A.1.

²² A general, rotated vector is derived explicitly in appendix A.2.

²³ Remember that we used the constraint to unit complex numbers in the two dimensional case, too.

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

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1. \quad (3.25)$$

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

$$q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}. \quad (3.26)$$

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

$$\mathbf{ijk} = -1 \quad (3.27)$$

suffices to compute all needed relations, for example $\mathbf{ij}=\mathbf{k}$ follows from multiplying Eq. 3.27 with \mathbf{k} :

$$\mathbf{ij} \underbrace{\mathbf{kk}}_{=-1} = -\mathbf{k} \rightarrow \mathbf{ij}=\mathbf{k}. \quad (3.28)$$

The set of unit quaternions $q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ are those satisfying the condition²⁴

$$q^\dagger q \stackrel{!}{=} 1$$

$$\rightarrow (a\mathbf{1} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k})(a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}) = a^2 + b^2 + c^2 + d^2 \stackrel{!}{=} 1. \quad (3.29)$$

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 represent each of the three complex units with a matrix. There are different possible ways of doing this, but one choice that does the job is as complex 2×2 matrices:

$$\begin{aligned} \mathbf{1} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \\ \mathbf{j} &= \begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} \quad , \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \end{aligned} \quad (3.30)$$

You can check that these matrices fulfil the defining conditions in Eq. 3.25 and Eq. 3.27. A generic quaternion can then be written, using

²⁴ The symbol \dagger , here is called "dagger" and denotes transposition plus complex conjugation: $a^\dagger = (a^*)^T$. 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.

these matrices, as

$$q = a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} = \begin{pmatrix} a + di & -b - ci \\ b - ci & a - di \end{pmatrix}. \quad (3.31)$$

Furthermore, we have

$$\det(q) = a^2 + b^2 + c^2 + d^2. \quad (3.32)$$

Comparing this with Eq. 3.29 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

$$U^\dagger U = 1 \quad \text{and} \quad \det(U) = 1. \quad (3.33)$$

In addition, the matrices in Eq. 3.30 are linearly independent²⁵ and therefore form a basis for the group called $SU(2)$. Take note that the way we define $SU(2)$ here, is analogous to how we defined $SO(2)$. The S denotes special, which means $\det(U) = 1$ and U stands for unitary, which means the property²⁶ $U^\dagger U = 1$. Every unit quaternion can be identified with an element of $SU(2)$.

Now, how is $SU(2)$ related to rotations? Unfortunately, the map between $SU(2)$ and²⁷ $SO(3)$ is not as simple as the one between $U(1)$ and $SO(2)$.

In 2-dimensions the 2 parameters of a complex number $z = a + ib$ could be easily identified with the two spatial axes, i.e. $v = x + iy$. The restriction to unit complex numbers automatically makes sure that the resulting matrix preserves the length of any vector²⁸ $(Rz)^* Rz = z^* R^* Rz = z^* z$. The quaternions have 4 parameters, so an identification with the 3 coordinates of a usual three-dimensional vector is not obvious. If we define

$$v \equiv x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \quad (3.34)$$

we can compute, using the matrix representation of the quaternions

$$\det(v) = x^2 + y^2 + z^2. \quad (3.35)$$

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. The restriction to **unit** quaternions means that we must restrict to matrices with **unit** determinant. Everything seems straight forward, but now comes 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

²⁵ This notion is explained in appendix A.1.

²⁶ For some more information about this, have a look at the appendix Sec. 3.10 at the end of this chapter.

²⁷ Recall that $SO(3)$ is the set of the usual rotation matrices acting on 3 dimensional vectors.

²⁸ Recall that here R is a unit complex number, because complex numbers can be rotated by multiplication with unit complex numbers. Therefore we have $R^* R = 1$, which is the defining condition for unit complex numbers.

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. Instead the transformation that does the job is given by

$$v' = q^{-1}vq. \quad (3.36)$$

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 will define u as a unit vector in $\mathbb{R}\mathbf{i} + \mathbb{R}\mathbf{j} + \mathbb{R}\mathbf{k}$ and denote a unit quaternion with

$$t = \cos(\theta) + \sin(\theta)u. \quad (3.37)$$

Using Eq. 3.34 a generic vector can be written

$$\vec{v} = (v_x, v_y, v_z)^T = v_x\mathbf{i} + v_y\mathbf{j} + v_z\mathbf{k} \underset{\text{Eq. 3.31}}{=} \begin{pmatrix} iv_z & -v_x - iv_y \\ v_x - iv_y & -iv_z \end{pmatrix}. \quad (3.38)$$

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. 3.30)

$$\vec{v} = (1, 0, 0)^T \rightarrow v = 1\mathbf{i} + 0\mathbf{j} + 0\mathbf{k} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (3.39)$$

$$R_z(\theta) = \cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{k} = \begin{pmatrix} \cos(\theta) + i\sin(\theta) & 0 \\ 0 & \cos(\theta) - i\sin(\theta) \end{pmatrix}, \quad (3.40)$$

²⁹ For a derivation have a look at appendix B.4.2.

which we can rewrite using Euler's formula²⁹

$$e^{ix} = \cos(x) + i\sin(x)$$

$$\Rightarrow R_z(\theta) = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}. \quad (3.41)$$

Inverting the quaternion rotation matrix yields

$$R_z(\theta)^{-1} = \begin{pmatrix} \cos(\theta) - i\sin(\theta) & 0 \\ 0 & \cos(\theta) + i\sin(\theta) \end{pmatrix} = \begin{pmatrix} e^{-i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix}. \quad (3.42)$$

The rotated vector is then using Eq. 3.36

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

$$= \begin{pmatrix} 0 & -e^{-i2\theta} \\ e^{2i\theta} & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\cos(2\theta) + i\sin(2\theta) \\ \cos(2\theta) + i\sin(2\theta) & 0 \end{pmatrix} \quad (3.43)$$

On the other hand, an arbitrary vector can be written in this quaternion notation (Eq. 3.38)

$$v' = \begin{pmatrix} iv'_z & -v'_x - iv'_y \\ v'_x - iv'_y & -iv'_z \end{pmatrix} \quad (3.44)$$

which we now compare with Eq. 3.43. This yields

$$v'_x = \cos(2\theta) \quad , \quad v'_y = -\sin(2\theta) \quad , \quad v'_z = 0. \quad (3.45)$$

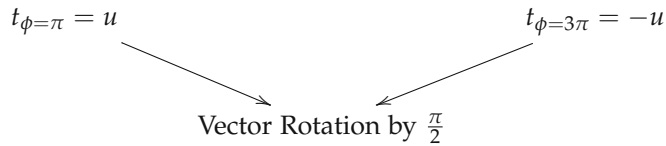
Therefore, written again in the conventional vector notation

$$\rightarrow \vec{v}' = (\cos(2\theta), -\sin(2\theta), 0)^T. \quad (3.46)$$

The identifications do indeed induce rotations³⁰, 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 and rewrite Eq. 3.37, which yields

$$t = \cos\left(\frac{\phi}{2}\right) + \sin\left(\frac{\phi}{2}\right)u. \quad (3.47)$$

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 example³¹



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 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 fundamental³².

In order to reveal the group that covers a given group, we need to introduce the most important tool of Lie theory: Lie algebras, which is the topic of the next section.

Take note that the fact we had one quaternion parameter too much, may be interpreted as a hint towards relativity. One may argue that a more natural identification would have been, as in the

³⁰ See the example in Eq. 3.24 where we rotated the vector, using the conventional rotation matrix

³¹ Because a rotation by π is the same as a rotation by $3\pi = 2\pi + \pi$ for ordinary vectors, because $2\pi = 360^\circ$ 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 .

³² To spoil the surprise: We will use the double cover of the Lorentz group, instead of the Lorentz group itself, because otherwise we miss something important: **Spin**. Spin is some kind of internal momentum and one of the most important particle labels. This is discussed in detail in Sec. 4.5.4 and Sec. 8.5.5.

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 working 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 by³³ 6 parameters. There is no 7-dimensional generalisation of complex numbers, which together with the constraint to unit objects would have 6 free parameters, but we see that **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.

³³ Using ordinary matrices, we need in four dimensions 4×4 matrices. The two conditions $O^T O = 1$ and $\det(O) = 1$ reduce the 16 components of an arbitrary 4×4 matrix to six independent components.

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 means there are elements of the group which are arbitrary close to the identity transformation, which changes nothing at all. In contrast, a discrete group has just a finite number of elements and therefore no element close to the identity. Consider again the symmetries of a square. A rotation by $0,000001^\circ$, 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, the symmetry transformations of a circle include a rotation by $0,000001^\circ$. 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

$$g(\epsilon) = I + \epsilon X \quad (3.48)$$

where the 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

$$h(\theta) = (I + \epsilon X)(I + \epsilon X)(I + \epsilon X) \dots = (I + \epsilon X)^k, \quad (3.49)$$

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

$$g(\theta) = I + \frac{\theta}{N} X. \quad (3.50)$$

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

$$h(\theta) = \lim_{N \rightarrow \infty} (I + \frac{\theta}{N} X)^N, \quad (3.51)$$

which is in the limit just the exponential function³⁴

$$h(\theta) = \lim_{N \rightarrow \infty} (I + \frac{\theta}{N} X)^N = e^{\theta X} \quad (3.52)$$

In some sense the object X generates the finite transformation h , which is why it's called the **generator**. This will be made more precise in a moment, but first let's look at this from another perspective:

If we are considering a continuous group of transformations that are given by matrices, we can make a Taylor expansion³⁵ of an element of the group about the identity. The Taylor series is given by

$$h(\theta) = I + \frac{dh}{d\theta} \big|_{\theta=0} \theta + \frac{d^2 h}{d\theta^2} \big|_{\theta=0} \theta^2 + \dots = \sum_n \frac{d^n h}{d\theta^n} \big|_{\theta=0} \theta^n. \quad (3.53)$$

This series expansion can be written in a more compact form, using the series expansion of the exponential function³⁶

$$h(\theta) = e^{\frac{dh}{d\theta} \big|_{\theta=0} \theta} \equiv \sum_n \frac{d^n h}{d\theta^n} \big|_{\theta=0} \theta^n. \quad (3.54)$$

and we can see now, how to make the connection to the previous description. By comparison with the result above:

$$X = \frac{dh}{d\theta} \big|_{\theta=0}. \quad (3.55)$$

³⁴ 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.

³⁵ If you've never heard of the Taylor expansion, or Taylor series before, you are encouraged to have a look at appendix B.3.

³⁶ This is derived in appendix B.4.1.

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**.

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 terms³⁷

³⁷ The Lie algebra which belongs to a group G is conventionally denoted by the 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 $t \in \mathbb{R}$.

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 matrices³⁸, but the multiplication of two matrices of the Lie algebra doesn't need to be an element of the Lie algebra. Instead there is another combination rule for the Lie algebra that is directly connected to the combination rule of the corresponding Lie group.

³⁸ A famous theorem of Lie group theory, called Ado's Theorem, states that every Lie algebra is isomorphic to a matrix Lie algebra.

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 formula³⁹

$$e^X \circ e^Y = e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+\dots} \quad (3.56)$$

³⁹ 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

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)

$$\underbrace{g}_{\in G} \circ \underbrace{h}_{\in G} = e^X \circ e^Y = \underbrace{e^{X+Y+\frac{1}{2}[X,Y]+\frac{1}{12}[X,[X,Y]]-\frac{1}{12}[Y,[X,Y]]+\dots}}_{\in G} \quad (3.57)$$

⁴⁰ Recall, that the Lie algebra which belongs to a group G is conventionally denoted by the Fraktur letter \mathfrak{g} .

with the generators⁴⁰ $X, Y \in \mathfrak{g}$. 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!⁴¹

⁴¹ A very illuminating proof of this fact can be found in John Stillwell. *Naive Lie Theory*. Springer, 1st edition, August 2008a. ISBN 978-0387782140

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 $[\cdot, \cdot]$. One says, the Lie algebra is **closed** under the Lie bracket, just as the group is closed under the corresponding composition rule \circ , e.q. matrix multiplication. Closure means that the composition of two elements lies again in the same set⁴².

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, which corresponds to many Lie groups by the more abstract definition. We will make this more concrete after introducing the modern definition of a Lie group.

Now we want to take a look at an example of how one 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 $SO(3)$ group are (Eq. 3.10)

$$O^T O \stackrel{!}{=} I \quad \text{and} \quad \det(O) \stackrel{!}{=} 1. \quad (3.58)$$

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

$$O = e^{\Phi J}. \quad (3.59)$$

Putting this into the first defining condition yields

$$O^T O = e^{\Phi J^T} e^{\Phi J} \stackrel{!}{=} 1 \rightarrow J^T + J \stackrel{!}{=} 0. \quad (3.60)$$

Using the second condition in Eq. 3.58 and the identity⁴³ $\det(e^A) = e^{\text{tr}(A)}$ for the matrix exponential function, we see

$$\begin{aligned} \det(O) \stackrel{!}{=} 1 & \rightarrow \det(e^{\Phi J}) = e^{\Phi \text{tr}(J)} \stackrel{!}{=} 1 \\ & \rightarrow \text{tr}(J) \stackrel{!}{=} 0 \end{aligned} \quad (3.61)$$

⁴² 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 $X \circ Y \notin \mathfrak{g}$

⁴³ $\text{tr}(A)$ 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 $\text{tr}(A) = A_{11} + A_{22}$.

⁴⁴ This is explained in appendix A.1.

Three linearly independent⁴⁴ matrices fulfilling the conditions Eq. 3.60 and Eq. 3.61 are

$$J_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad J_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \quad J_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.62)$$

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 symbol⁴⁵

⁴⁵ The Levi-Civita symbol is explained in appendix B.5.5.

$$(J_i)_{jk} = -\epsilon_{ijk}, \quad (3.63)$$

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

$$(J_1)_{jk} = -\epsilon_{1jk} \leftrightarrow \begin{pmatrix} (J_1)_{11} & (J_1)_{12} & (J_1)_{13} \\ (J_1)_{21} & (J_1)_{22} & (J_1)_{23} \\ (J_1)_{31} & (J_1)_{32} & (J_1)_{33} \end{pmatrix} = - \begin{pmatrix} \epsilon_{111} & \epsilon_{112} & \epsilon_{113} \\ \epsilon_{121} & \epsilon_{122} & \epsilon_{123} \\ \epsilon_{131} & \epsilon_{132} & \epsilon_{133} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (3.64)$$

⁴⁶ This is exactly the two-dimensional Levi-Civita symbol $(j_1)_{ij} = \epsilon_{ijk}$ in matrix form (see appendix B.5.5), which is the generator of rotations in two dimensions (of $SO(2)$).

Let's see what finite transformation matrix we get from the first of these basis generators. We can focus on the lower right 2×2 matrix⁴⁶ j_1 and ignore the zeroes for a moment:

$$J_1 = \begin{pmatrix} 0 & & \\ & \underbrace{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}}_{\equiv j_1} & \\ & & \end{pmatrix}. \quad (3.65)$$

We can immediately compute

$$(j_1)^2 = -1, \quad (3.66)$$

therefore

$$(j_1)^3 = \underbrace{(j_1)^2}_{=-1} j_1 = -j_1, \quad (j_1)^4 = +1, \quad (j_1)^5 = +j_1. \quad (3.67)$$

In general, we have

$$(j_1)^{2n} = (-1)^n I \quad \text{and} \quad (j_1)^{2n+1} = (-1)^n j_1, \quad (3.68)$$

⁴⁷ This is derived in appendix B.4.1. The trick used here is explained in more detail in appendix B.4.2 and the series expansions of sine and cosine are derived in appendix B.4.1, too.

which we can use when we evaluate the exponential function as series expansion⁴⁷

$$\begin{aligned}
R_{1 \text{ fin}} &= e^{\Phi j_1} = \sum_{n=0}^{\infty} \frac{\Phi^n j_1^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{\Phi^{2n}}{(2n)!} \underbrace{(j_1)^{2n}}_{(-1)^n I} + \sum_{n=0}^{\infty} \frac{\Phi^{2n+1}}{(2n+1)!} \underbrace{(j_1)^{2n+1}}_{(-1)^n j_1} \\
&= \underbrace{\left(\sum_{n=0}^{\infty} \frac{\Phi^{2n}}{(2n)!} (-1)^n \right)}_{=\cos(\phi)} I + \underbrace{\left(\sum_{n=0}^{\infty} \frac{\Phi^{2n+1}}{(2n+1)!} (-1)^n \right)}_{=\sin(\phi)} j_1 \\
&= \cos(\phi) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \sin(\phi) \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{pmatrix}
\end{aligned} \tag{3.69}$$

Therefore the complete, finite transformation matrix is, using $e^0 = 1$ for the upper-left component

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}, \tag{3.70}$$

which we can recognize as one of the well-known rotation matrices in 3-dimensions that were quoted at the beginning of this chapter (Eq. 3.23). Following the same steps, we can derive the matrices for rotations around the other axes.

We now have the generators of the group in explicit matrix form (Eq. 3.62) and we can compute the corresponding Lie bracket⁴⁸ by brute force. This yields⁴⁹

$$[J_i, J_j] = \epsilon_{ijk} J_k, \tag{3.71}$$

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", that is instead of $e^{\phi J}$, we write $e^{i\phi J}$ and our generators are then

$$J_1 = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \quad J_2 = i \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad J_3 = i \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.72}$$

and the Lie algebra⁵⁰ reads

$$[J_i, J_j] = i\epsilon_{ijk} J_k. \tag{3.73}$$

We do this in physics to get Hermitian generators, which means⁵¹

⁴⁸ As explained above, the natural product of the Lie algebra is the Lie bracket. Here we compute how the basis generators behave, when put into the Lie bracket. All other generators can be constructed by linear combination of these basis generators. Therefore, if we know the result of the Lie bracket of the basis generators, we know automatically the result for all other generators. This behaviour of the basis generators in the Lie bracket, will become incredibly important in the next section. Everything that is important about a Lie algebra, is encoded in the Lie bracket relation of the basis generators.

⁴⁹ For example, we have

$$\begin{aligned}
[J_1, J_2] &= J_1 J_2 - J_2 J_1 \\
&= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} = \\
&= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \\
&= \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \underbrace{\epsilon_{12k}}_{=0 \text{ except for } k=3} J_k \\
&= \epsilon_{123} J_3 = J_3
\end{aligned}$$

⁵⁰ We will call the Lie bracket relation of the basis generators **the** Lie algebra, because everything important is encoded here.

⁵¹ For example now we have $J_1^\dagger =$

$$\begin{aligned}
&i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \text{ and therefore} \\
J_1^\dagger &= (J_1^*)^T = i \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = J
\end{aligned}$$

$J^\dagger = (J^*)^T = J$, because Hermitian matrices have real eigenvalues and this becomes important in quantum mechanics when the eigenvalues of the generators become the values we can expect to measure in experiments, which will be discussed in Sec. 8.3. Otherwise, that is without the "i", the generators are anti-Hermitian $J^\dagger = (J^*)^T = -J$ and the corresponding eigenvalues are complex.

We can derive the basis generators in another way, by starting with the well known rotation matrices and using from Eq. 3.55 that $X = \frac{dh}{d\theta}|_{\theta=0}$. For the first rotation matrix, as quoted in Eq. 3.23 and derived in Eq. 3.70, this yields

$$\begin{aligned} J_1 &= \frac{dR_1}{d\theta}|_{\theta=0} = \frac{d}{d\theta} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix} \Big|_{\theta=0} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\sin(\theta) & -\cos(\theta) \\ 0 & \cos(\theta) & \sin(\theta) \end{pmatrix} \Big|_{\theta=0} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \end{aligned} \quad (3.74)$$

which is exactly the first generator in Eq. 3.62. 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 compute only afterwards the explicit matrix form for the Lorentz transformations. If you already have explicit transformation matrices, you can always use Eq. 3.55 to derive the corresponding generators.

Before we move on, we will have a look at the modern definition of a Lie algebra.

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 $e^X \in G$. Later we learned that an important part of a group, the rule for the combination of group elements, is encoded in the Lie algebra in form of the Lie bracket. As we did for groups, we distil the defining features of this idea into precise mathematical axioms:

A Lie algebra is a vector space \mathfrak{g} equipped with a binary operation $[\cdot, \cdot]: \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$. 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 matrices fulfils all these conditions and of course this standard commutator was used historically to get to these axioms. Nevertheless, there are quite different binary operations that fulfil these axioms, for example, the famous Poisson bracket of classical mechanics.

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 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 combination, fulfil the same Lie bracket relation as the basis generators of $SO(3)$ (Eq. 3.73). 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 cover⁵² of $SO(3)$.

Remember that $SU(2)$ is the group of unitary 2×2 matrices with unit determinant⁵³:

$$U^\dagger U = U U^\dagger = 1 \quad (3.75)$$

$$\det(U) = 1. \quad (3.76)$$

The first thing we 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, \dots yields⁵⁴

$$U^\dagger U = (e^{iJ_i})^\dagger e^{iJ_i} \stackrel{!}{=} 1 \quad (3.77)$$

$$\det(U) = \det(e^{iJ_i}) \stackrel{!}{=} 1 \quad (3.78)$$

The first condition tells us, using the Baker-Champell-Hausdorf Theorem (Eq. 3.56) and $[J_i, J_i] = 0$

⁵² Recall that this means that the map from $SU(2)$ to $SO(3)$ identifies two elements of $SU(2)$ with the same element of $SO(3)$.

⁵³ This is what the "S" stands for: Special = unit determinant.

⁵⁴ As discussed above, we now work with an extra "i" in the exponent, in order to get Hermitian matrices, which guarantees that we get real numbers as predictions for experiments in quantum mechanics.

$$\begin{aligned}
(e^{iJ_i})^\dagger e^{iJ_i} &= e^{-iJ_i^\dagger} e^{iJ_i} \stackrel{!}{=} 1 \\
&\rightarrow e^{-iJ_i^\dagger + iJ_i + \frac{1}{2}[J_i^\dagger, J_i] + \dots} \stackrel{!}{=} 1 \\
\lim_{e^0=1} J_i^\dagger &\stackrel{!}{=} J_i.
\end{aligned} \tag{3.79}$$

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 from the second condition:

$$\det(e^{iJ_i}) = e^{i\text{tr}(J_i)} = 1 \lim_{e^0=1} \text{tr}(J_i) \stackrel{!}{=} 0. \tag{3.80}$$

We conclude the generators of $SU(2)$ must be Hermitian traceless matrices. A basis for Hermitian traceless 2×2 matrices is given by 3 matrices⁵⁵:

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.81}$$

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, which yields

$$[\sigma_i, \sigma_j] = 2i\epsilon_{ijk}\sigma_k, \tag{3.82}$$

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

$$[J_i, J_j] = i\epsilon_{ijk}J_k \tag{3.83}$$

Take note that this is exactly the same Lie bracket relation we derived for $SO(3)$ (Eq. 3.73)! 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

⁵⁵ A complex 2×2 matrix has 4 complex entries and therefore 8 degrees of freedom. Because of the two conditions only three degrees of freedom remain.

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 chapters.

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 by $z^*z = 1$, which reads if we write $z = a + ib$:

$$z^*z = (a + ib)^*(a + ib) = (a - ib)(a + ib) = a^2 + b^2 = 1. \quad (3.84)$$

This is exactly the defining condition of the unit circle⁵⁶. The set of unit-complex numbers is the unit circle in the complex plane. Furthermore, we found that there is a one-to-one map⁵⁷ 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. 3.29)

$$a^2 + b^2 + c^2 + d^2 \stackrel{!}{=} 1, \quad (3.85)$$

which is the same condition that defines⁵⁸ the three sphere S^3 ! Therefore this map provides 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 a the three sphere S^3 .

These observations motivate the modern definition of a Lie group⁵⁹:

A Lie group is a group, which is also a differentiable manifold⁶⁰. 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 .

⁵⁶ 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 $x_1^2 + x_2^2 = 1$.

⁵⁷ To be precise: An isomorphism. To say two things are isomorphic is the mathematical way of saying that they are "the same thing" and two things are called isomorphic if there exists an isomorphism between them.

⁵⁸ 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 sphere.

⁵⁹ The technical details that follow aren't important for what we want to do in this book. The important message to take away is: Lie group = manifold.

⁶⁰ 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 Sec. 3.11 at the end of this chapter.

By the abstract definition of a Lie algebra we say that $SO(3)$ and $SU(2)$ have the same Lie algebra (Eq. 3.83). Now it's time to talk about the remark at the end of Sec. 3.4:

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

We can now understand a bit better, why this one group is distinguished. 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 point⁶¹.

To emphasize this important point:⁶²

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

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 this Lie algebra, i.e. Eq. 3.83. We can get all other groups belonging to this Lie algebra through maps from $SU(2)$.

We can now see 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, $SO(3)$ is one half of the unit sphere.

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.

I want to take the view in this book that in order 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 symmetry.

⁶¹ 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.

⁶² 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

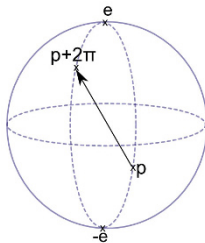


Fig. 3.7: Two-dimensional slice of the three Sphere S^3 (which is a three dimensional surface and therefore not drawable itself). We can see that the top half of the sphere is $SO(3)$, because to get from $SU(2)$ to $SO(3)$ we identify two points, for example, p and $p + 2\pi$, with each other.

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 Poincare group, instead of just the usual representation one uses to transform four-vectors. To describe nature at the most fundamental level, one must use the covering group, instead of any of the other groups one can map to from the covering group.

We are able to derive the representations⁶³ 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 hides⁶⁴ something from us, because it is not the most fundamental group belonging to this symmetry. The covering group of the Poincare group⁶⁵ is the fundamental group and therefore we will use it to describe nature.

To summarize⁶⁶

- $S^1 \triangleq U(1) \xleftrightarrow[\text{one-to-one}]{} SO(2)$
- $S^3 \triangleq SU(2) \xleftrightarrow[\text{two-to-one}]{} SO(3) \triangleq \text{half of } S^3$
 $\Rightarrow SU(2)$ is the distinguished group belonging to the Lie algebra $[J_i, J_j] = i\epsilon_{ijk}J_k$ (Eq. 3.83), 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 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.

⁶³ This notion will be made precise in the next section.

⁶⁴ For those who already know some quantum mechanics: The standard symmetry group hides spin from us!

⁶⁵ For brevity, we will avoid writing "double cover of" or "covering group of" most of the time. We will use one representation of the Poincare group to derive the corresponding Lie algebra. Then we will use this Lie algebra to derive the representations of the one distinguished group that belongs to this Lie algebra. In other words: The representations of the double cover of the Poincare group.

⁶⁶ Maybe you wonder why S^2 , the surface of the sphere in three dimensions, is missing. S^2 is not a Lie group and this is closely related to the fact that there are no three-dimensional complex numbers. Recall that we had to move from two-dimensional complex numbers with just i to the four-dimensional quaternions with i, j, k .

⁶⁷ This will make much more sense in a moment.

⁶⁸ The mathematical term for a map with these special properties is **homomorphism**. The definition of an isomorphism is then a homomorphism, which is in addition one-to-one.

⁶⁹ In the context of this book this will always mean that we map each group element to a matrix. Each group element is then given by a matrix that acts by usual matrix multiplication on the elements of some vector space.

⁷⁰ This concept can be formulated more generally if one accepts arbitrary (not 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.

⁷¹ R^3 denotes three dimensional Euclidean space, where elements are ordinary 3 component vectors, as we use them for example in appendix A.1.

⁷² We will learn later in this chapter how to derive these. At this point just take notice that it is possible.

An important idea is that **one** group can act on **many** different kinds of objects⁶⁷. This idea motivates the definition of a representation: A representation is a map⁶⁸ between any group element g of a group G and a linear transformation⁶⁹ $R(g)$ of some vector-space V

$$g \xrightarrow[R]{} R(g) \quad (3.86)$$

in such a way that the group properties are preserved:

- $R(e) = I$ (The identity element of the group transforms nothing at all)
- $R(g^{-1}) = (R(g))^{-1}$ (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 representation⁷⁰ 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 space⁷¹ R^3 . The rotation matrices are linear transformations on R^3 . But take note 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 C^1 , which is especially easy, as we will see later, or two: C^2 , which we will discuss in detail in the following sections. The objects living in 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 C^2 are just the usual matrices we already know for $SU(2)$. In addition, we can examine how $SU(2)$ acts on C^3 . There is a well defined framework for constructing such representations and as a result, $SU(2)$ acts on complex vectors of dimension three as 3×3 matrices. For example, a basis for the $SU(2)$ generators on C^3 is given by⁷²

$$J_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (3.87)$$

As usual, we can then compute $SU(2)$ matrices in this representation 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 call⁷³ this group S^3 instead of $SU(2)$, because usually $SU(2)$ is defined as the set of complex 2×2 (!) matrices satisfying (Eq. 3.33)

$$U^\dagger U = 1 \quad \text{and} \quad \det(U) = 1 \quad (3.88)$$

and now we write $SU(2)$ as 3×3 matrices. Therefore 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 by a representation. For example, for $SU(2)$ we started with 2×2 matrices. This enables us to study the group properties concretely, as we did in the preceding chapters. After this initial study it's often more helpful to regard the group as an abstract group⁷⁴, 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 **similarity transformation**. Given a matrix D and an invertible⁷⁵ matrix S then a transformation of the form

$$R \rightarrow R' = S^{-1}RS \quad (3.89)$$

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

$$R(g_1)R(g_2) = R(g_1g_2) \quad (3.90)$$

⁷³ 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.

⁷⁴ For $SU(2)$ this means using S^3 .

⁷⁵ 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} .

If we now look at the similarity transformation of the representation

$$S^{-1}R(g_1)\underbrace{SS^{-1}}_{=1}R(g_2)S = S^{-1}R(g_1)R(g_2)S = S^{-1}R(g_1g_2)S \quad (3.91)$$

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.

The next notion we want to introduce is **invariant subspace**. If we have a representation R of a group G on a vector space V we call $V' \subseteq V$ an invariant subspace if for⁷⁶ $v \in V'$ 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 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

$$R'(g)v = R(g)v \quad (3.92)$$

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 of an **irreducible representation**. An irreducible representation is a representation of a group G on a vector space V that has no invariant subspaces besides 0 and V itself. 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: A irreducible representation cannot be rewritten, using a similarity transformation, in block diagonal form. In contrast to a reducible representation, which can be rewritten in block-diagonal form by similarity transformations. This notion is important because nature uses irreducible representations⁷⁷ to describe elementary particles. We will see later that the behaviour of elementary particles under transformations is described by irreducible representations of the corresponding symmetry group.

⁷⁷ What else?

⁷⁸ 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.

There are many possible representations⁷⁸ for each group, 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 build from generators of the Lie algebra and its defining feature is that it

commutes with every generator X of the group

$$[C, X] = 0. \quad (3.93)$$

What does this mean? A famous Lemma, called Schur's Lemma⁷⁹, tells us that if we have an irreducible representation $R : \mathfrak{g} \rightarrow GL(V)$, any linear operator $T : V \rightarrow V$ 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.⁸⁰ We can then start to investigate the irreducible representations, 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 in the definition of a representation above? The definition states that a representation is a map from the abstract group to the space of linear operators on a vector space. Now, from linear algebra we know that the eigenvectors of a linear operator always form a basis for the vector space. We can use this to inspect the vector space. For any Lie group, one or more of the generators of a Lie group can be diagonalized⁸¹ using similarity transformations and we will use these diagonalized generators to get a basis of our vector space.

We will now start deriving 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 $SU(2)$ algebra. The Lorentz group is part of the Poincare group and we will talk about these groups in this order.

3.6 $SU(2)$

We used in Sec. 3.4.3 specific matrices (=a specific representation) to identify how the generators of $SU(2)$ behave, when put into the Lie bracket⁸². We can use this knowledge to find further representations. We will arrive again at the representation we started with, which means the set of unitary $2 \times$ matrices with unit determinant and are then able to see that it is just one special case. Before we are going to 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 $SU(2)$

To learn something about what finite-dimensional, irreducible representations of $SU(2)$ are possible, we define new operators from the

⁷⁹ A basic result of group theory, which you can look up in any book about group theory.

⁸⁰ This will become much clearer as soon as we look at an example.

⁸¹ The set of diagonal generators is called Cartan subalgebra, and the corresponding generators Cartan generators. These generators play a big role in quantum field theory, 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 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 eigenvalues of this Cartan generator.

⁸² Recall that this is what we use to define the Lie algebra of a group in abstract terms. The final result was Eq. 3.83.

⁸³ We can always diagonalize **one** of the generators. Following the convention we choose J_3 as diagonal and therefore yielding the basis vectors for our vector space. Furthermore, it is conventional to introduce the new operators J_{\pm} in the way we do here.

ones we used in Sec. 3.4.3, by linear combination⁸³

$$J_+ = \frac{1}{\sqrt{2}}(J_1 + iJ_2) \quad (3.94)$$

$$J_- = \frac{1}{\sqrt{2}}(J_1 - iJ_2) \quad (3.95)$$

These new operators obey the following commutation relations, as you can check by using the commutator relations in Eq. 3.83

$$[J_3, J_{\pm}] = \pm J_{\pm} \quad (3.96)$$

$$[J_+, J_-] = J_3. \quad (3.97)$$

If we now investigate how these operators act on an eigenvector v of J_3 with eigenvalue⁸⁴ b we discover something remarkable:

$$\begin{aligned} J_3(J_{\pm}v) &= J_3(J_{\pm}v) + \underbrace{J_{\pm}J_3v - J_3J_{\pm}v}_{=0} \\ &= \underbrace{J_{\pm}J_3v}_{=J_{\pm}bv} + \underbrace{J_3J_{\pm}v - J_{\pm}J_3v}_{=[J_3, J_{\pm}]v} \\ &\stackrel{\text{Eq. 3.96}}{=} (b \pm 1)J_{\pm}v \end{aligned} \quad (3.98)$$

We conclude that $J_{\pm}v$ is again an eigenvector, let's call him w , of J_3 with eigenvalue $(b \pm 1)$:

$$J_3w = (b \pm 1)w \quad \text{with} \quad w = J_{\pm}v. \quad (3.99)$$

The operators J_- and J_+ are called raising and lowering or **ladder operators**. We can construct more and more eigenvectors of J_3 using the operators 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 there must be an eigenvector with a maximum eigenvalue v_{max} . After a finite number N of applications of J_+ we reach the maximum eigenvector v_{max}

$$v_{max} = J_+^N v \quad (3.100)$$

We have

$$J_+v_{max} = 0, \quad (3.101)$$

⁸⁴ This means $J_3v = bv$ as explained in appendix C.4.

because v_{max} is, by definition, the eigenvector with the highest eigenvalue. We call the maximum eigenvalue $j := b + N$. The same must be true for the other direction: There must be an eigenvector with minimum eigenvalue v_{min} for which the following relation holds

$$J_- v_{min} = 0 \quad (3.102)$$

Let us say we reach the minimum after operating M times with J_- on v_{max}

$$v_{min} = J_-^M v_{max}. \quad (3.103)$$

Therefore, v_{min} has eigenvalue $j-M$. To go further we need to know how exactly J_{\pm} acts on eigenvectors. The computation above shows that $J_- v_k$ is, in general, a scalar multiplied by an eigenvector with eigenvalue $k-1$:

$$J_- v_k = \alpha_k v_{k-1}. \quad (3.104)$$

If we inspect in detail how J_- acts on v_{max} we get⁸⁵ the general rule for the scalar factor

$$\alpha_{j-k} = \frac{1}{\sqrt{2}} \sqrt{(2j-k)(k+1)} \quad (3.105)$$

Take note that this scalar factor becomes zero for $k = 2j$ and therefore, we have reached the end of the ladder after $2j$ steps if we start at the top. Therefore v_{min} has eigenvalue $j - 2j = -j$. We conclude that we have in general $2j+1$ eigenstates with eigenvalues

$$\{-j, -j+1, \dots, j-1, j\} \quad (3.106)$$

This is only possible if j is an integer or a half-integer⁸⁶. Now we know that our vector space V has $2j+1$ dimensions⁸⁷, because we have $2j+1$ linearly independent eigenvectors. Those eigenvectors of J_3 span the complete vector space V because J_1 and J_2 can be expressed in terms of J_+ and J_- and therefore take any linear combination $\sum_i a_i v_i$ into a possibly different linear combination $\sum_i b_i v_i$, with scalar factors a_i, b_i . Therefore, the span of the eigenvectors of J_3 is a non-zero invariant subspace of V and because we are looking for irreducible representations they span the complete vector space V .

We can use the construction above to define representations of $SU(2)$ on a vector space V_j with $2j+1$ dimensions and basis given by the eigenvectors v_k of J_3 . Furthermore, it's possible to show that every irreducible representation of $SU(2)$ must be equivalent to one of these⁸⁸.

⁸⁵ See, for example, page 90 in Matthew Robinson. *Symmetry and the Standard Model*. Springer, 1st edition, August 2011. ISBN 978-1-4419-8267-4

⁸⁶ Try it with other fractions if you don't believe this!

⁸⁷ See, for example, page 189 in Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

⁸⁸ See page 190 in: Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

3.6.2 The Casimir Operator of $SU(2)$

As described in Sec. 3.5, we can naturally label representations by using the Casimir operators⁸⁹ of the group. $SU(2)$ has exactly one Casimir operator:

$$J^2 := (J_1)^2 + (J_2)^2 + (J_3)^2 \quad (3.107)$$

that fulfils the defining condition:

$$[J^2, J_i] = 0. \quad (3.108)$$

We can re-express J^2 in terms of J_{\pm} by using the definition of J_{\pm} in Eq. 3.95 and Eq. 3.94:

$$\begin{aligned} J^2 &= J_+ J_- + J_- J_+ + (J_3)^2 \\ &= \frac{1}{2} (J_1 + iJ_2)(J_1 - iJ_2) + \frac{1}{2} (J_1 - iJ_2)(J_1 + iJ_2) + (J_3)^2 \\ &= \frac{1}{2} \left((J_1)^2 - iJ_1 J_2 + iJ_2 J_1 + (J_2)^2 \right) + \frac{1}{2} \left((J_1)^2 + iJ_1 J_2 - iJ_2 J_1 + (J_2)^2 \right) \\ &\quad + (J_3)^2 \\ &= (J_1)^2 + (J_2)^2 + (J_3)^2 \quad \checkmark \end{aligned} \quad (3.109)$$

If we now use⁹⁰

$$J_+ v_k = \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} v_{k+1} \quad (3.110)$$

and

$$J_- v_k = \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} v_{k-1} \quad (3.111)$$

we can compute the fixed scalar value for each representation:

$$\begin{aligned} J^2 v_k &= \left(\frac{1}{2} (J_+ J_- + J_- J_+) + (J_3)^2 \right) v_k \\ &= J_+ J_- v_k + J_- J_+ v_k + k^2 v_k \\ &= J_+ \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} v_{k-1} + J_- \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} v_{k+1} + k^2 v_k \\ &= \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} J_+ v_{k-1} + \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} J_- v_{k+1} + k^2 v_k \\ &= \frac{1}{\sqrt{2}} \sqrt{(j+k)(j-k+1)} \frac{1}{\sqrt{2}} \sqrt{(j+(k-1)+1)(j-(k-1))} v_k \\ &\quad + \frac{1}{\sqrt{2}} \sqrt{(j+k+1)(j-k)} \frac{1}{\sqrt{2}} \sqrt{(j+(k+1))(j-(k+1)+1)} v_k + k^2 v_k \\ &= \frac{1}{2} (j+k)(j-k+1) + \frac{1}{2} (j-k)(j+k+1) v_k + k^2 v_k \\ &= (j^2 + j) v_k = j(j+1) v_k \end{aligned} \quad (3.112)$$

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

⁸⁹ Recall that Casimir operators are defined as operators C , built from the generators of the group that commute with every generator X of the group: $[C, X] = 0$.

⁹⁰ These are just the normalization constants. If we act with J_{\pm} onto a normalized state, the resulting state will in general not be normalized, too. Nevertheless, in physics we always prefer working with normalized states, for reasons that will become clear in the following chapters. The derivation is a bit tedious, but simply starts with $J_{\pm} v_k = c v_{k\pm 1}$ where c is the normalization constant in question. The complete computation can be found in most books about quantum mechanics in the chapter about angular momentum and angular momentum ladder operators. If this is new to you, do not waste too much time here because the result of this section is not too important for everything that follows.

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

The lowest possible value for j is zero. In this case our representation acts on a $2j + 1 = 2 \cdot 0 + 1 = 1$ dimensional vector space. We can see that this representation is trivial, because the only 1×1 matrices fulfilling the commutation relations of the $SU(2)$ Lie algebra

$[J_l, J_m] = i\epsilon_{lmn}J_n$, are trivially 0. If we exponentiate the generator 0 we always get the transformation $U = e^0 = 1$ which changes nothing at all.

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

We now take a look at the next lowest possible value $j = \frac{1}{2}$. This representation is $2\frac{1}{2} + 1 = 2$ dimensional. The generator J_3 has eigenvalues $\frac{1}{2}$ and $\frac{1}{2} - 1 = -\frac{1}{2}$, as can be seen from Eq. 3.106 and is therefore given by

$$J_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad (3.113)$$

because we choose J_3 to be the diagonal generator⁹¹. The eigenvectors corresponding to the eigenvalues $+\frac{1}{2}, -\frac{1}{2}$ are:

$$v_{\frac{1}{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad v_{-\frac{1}{2}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.114)$$

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

$$J_1 = \frac{1}{\sqrt{2}}(J_- + J_+) \quad (3.115)$$

$$J_2 = \frac{i}{\sqrt{2}}(J_- - J_+), \quad (3.116)$$

which we get directly from inverting the definitions of J_{\pm} in Eq. 3.95 and Eq. 3.94. 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

$$J_1 v_{\frac{1}{2}} = \frac{1}{\sqrt{2}}(J_- + J_+)v_{\frac{1}{2}} = \frac{1}{\sqrt{2}}(J_- v_{\frac{1}{2}} + \underbrace{J_+ v_{\frac{1}{2}}}_{=0}) = \frac{1}{\sqrt{2}}J_- v_{\frac{1}{2}} = \frac{1}{2}v_{-\frac{1}{2}}, \quad (3.117)$$

where we used that $\frac{1}{2}$ is already the maximum value for $v_{\frac{1}{2}}$ and we cannot go higher. The factor $\frac{1}{2}$ is the scalar factor we get from Eq. 3.105. Similarly we get

$$J_1 v_{-\frac{1}{2}} = \frac{1}{\sqrt{2}}(J_- + J_+)v_{-\frac{1}{2}} = \frac{1}{2}v_{\frac{1}{2}} \quad (3.118)$$

⁹¹ 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.

Written in matrix form, where our basis is given by $v_{\frac{1}{2}} = (1, 0)^T$ and $v_{-\frac{1}{2}} = (0, 1)^T$:

$$J_1 = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (3.119)$$

You can check that this matrix has the action on the basis vectors we derived above⁹². In the same way, we find

$$J_2 = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \quad (3.120)$$

These are the same generators $J_i = \frac{1}{2}\sigma_i$, with the Pauli matrices σ_i , we found while investigating Lie algebra of $SU(2)$ at the beginning of this chapter (Eq. 3.81). 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 section⁹³.

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

Following the same procedure⁹⁴ as in two-dimensions, we find:

$$J_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad J_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (3.121)$$

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 deriving 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 Picasso⁹⁵

In this section we will use **one** known representation of the Lorentz group to derive the corresponding Lie algebra, which 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

⁹² We derived in Eq. 3.117:

$J_1 v_{\frac{1}{2}} = \frac{1}{2} v_{-\frac{1}{2}}$. Using the explicit matrix form of J_1 we get
 $J_1 v_{\frac{1}{2}} = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{2} v_{-\frac{1}{2}}$ ✓.

⁹³ 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.

⁹⁴ We start again with the diagonal generator J_3 , which we can write down immediately because we know its eigenvalues $(1, 0, -1)$. Afterwards, the other two generators J_1, J_2 can be derived by their action, where we again use that we can write them in terms of J_{\pm} , on the eigenvectors of J_3 .

⁹⁵ As quoted in Robert S. Root-Bernstein and Michele M. Root-Bernstein. *Sparks of Genius*. Mariner Books, 1st edition, 8 2001. ISBN 9780618127450

that this algebra can be seen to be constructed of two copies of the Lie algebra of $SU(2)$. 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 space⁹⁶

$$x^\mu x_\mu = x^\mu \eta_{\mu\nu} x^\nu = (x^0)^2 - (x^1)^2 - (x^2)^2 - (x^3)^2 \quad (3.122)$$

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

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (3.123)$$

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 $x^\mu \rightarrow x'^\mu = \Lambda^\mu_\nu x^\nu$, we get the product

$$x^\mu \eta_{\mu\nu} x^\nu \rightarrow x'^\sigma \eta_{\sigma\rho} x'^\rho = (x^\mu \Lambda^\sigma_\mu) \eta_{\sigma\rho} (\Lambda^\rho_\nu x^\nu) \stackrel{!}{=} x^\mu \eta_{\mu\nu} x^\nu \quad (3.124)$$

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

$$\Lambda^\sigma_\mu \eta_{\sigma\rho} \Lambda^\rho_\nu \stackrel{!}{=} \eta_{\mu\nu} \quad (3.125)$$

or written in matrix form⁹⁷

$$\Lambda^T \eta \Lambda \stackrel{!}{=} \eta. \quad (3.126)$$

This is how the Lorentz transformations Λ are defined! If we take the determinant of the equation and use

$\det(AB) = \det(A) \det(B)$ we get the defining condition

$$\det(\Lambda) \underbrace{\det(\eta)}_{=-1} \det(\Lambda) = \underbrace{\det(\eta)}_{=-1} \rightarrow \det(\Lambda)^2 \stackrel{!}{=} 1 \quad (3.127)$$

⁹⁶ This was derived in Chap. 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.

⁹⁷ 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 .

$$\rightarrow \det(\Lambda) \stackrel{!}{=} \pm 1 \quad (3.128)$$

⁹⁸ We will see in a minute why this is useful.

Furthermore, we get if we look at⁹⁸ the $\mu = \nu = 0$ component in Eq. 3.125

$$\Lambda_0^\sigma \eta_{\sigma\rho} \Lambda_0^\rho \stackrel{!}{=} \underbrace{\eta_{00}}_{=1} \rightarrow \Lambda_0^\sigma \eta_{\sigma\rho} \Lambda_0^\rho = (\Lambda_0^0)^2 - \sum_i (\Lambda_0^i)^2 \stackrel{!}{=} 1 \quad (3.129)$$

and we conclude

$$\Lambda_0^0 \stackrel{!}{=} \pm \sqrt{1 + \sum_i (\Lambda_0^i)^2}. \quad (3.130)$$

We divide the Lorentz group into four components, depending on the signs in the Eq. 3.128 and Eq. 3.130. The components that preserve the orientation⁹⁹ of the coordinate system are those two with $\det(\Lambda) = +1$. Furthermore, if we want to preserve the direction of time we need to restrict to $\Lambda_0^0 \geq 0$, because

$$x^0 = t \rightarrow x'^0 = t' = \Lambda_0^\nu x^\nu = \Lambda_0^0 t + \Lambda_1^0 x^1 + \Lambda_2^0 x^2 + \Lambda_3^0 x^3, \quad (3.131)$$

where we can see that, if $\Lambda_0^0 \geq 0$, then t' has the same sign as t . This component is called $SO(1,3)^\uparrow$ and we will talk about this subgroup most of the time. The fancy term for this subgroup is proper¹⁰⁰ or-thochronous¹⁰¹ Lorentz group. The four components of the Lorentz group are disconnected in the sense that it is not possible to get a Lorentz transformation of another component just by using the Lorentz transformations of one component. Other components can be obtained from $SO(1,3)^\uparrow$ by using¹⁰²

$$\Lambda_P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (3.132)$$

$$\Lambda_T = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.133)$$

Λ_P is called the parity operator. A parity transformation is simply a reflection in a mirror. Λ_T is the time-reversal operator.

The complete Lorentz group $O(1,3)$ can then be seen as the set:

$$O(1,3) = \{SO(1,3)^\uparrow, \Lambda_P SO(1,3)^\uparrow, \Lambda_T SO(1,3)^\uparrow, \Lambda_P \Lambda_T SO(1,3)^\uparrow\} \quad (3.134)$$

⁹⁹ This means a right-handed coordinate system stays right-handed and a left-handed coordinate system stays left-handed. For the definition of left- and right-handed coordinate systems have a look at appendix A.5.

¹⁰⁰ This term refers to the fact that the subgroup $SO(1,3)^\uparrow$ preserves orientation/parity.

¹⁰¹ This means that this subgroup preserves the direction of time.

¹⁰² At least for one representation, these operators look like this. We will see later that for different representations, these operators look quite different.

Therefore, we can restrict our search for representations of the Lorentz group, to representations of $SO(1,3)^\dagger$, because then we only need to find representations for Λ_P and Λ_T , to get representations of the other components.

3.7.1 One Representation of the Lorentz Group

Let's see how we can use the defining condition of the Lorentz group (Eq. 3.125) 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-vectors¹⁰³, 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 conditions¹⁰⁴, 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 the complete Lie algebra of this group. These generators form a basis for this Lie algebra, which 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, involving only space and leaving time unchanged, fulfil the condition in Eq. 3.125. This follows because the spatial part¹⁰⁵ of the Minkowski metric is proportional to the 3×3 identity matrix¹⁰⁶ and therefore for transformations involving only space, we have from Eq. 3.125 the condition

$$\begin{aligned} -R^T I_{3 \times 3} R &= -R^T R \stackrel{!}{=} -I_{3 \times 3} \\ \rightarrow R^T I_{3 \times 3} R &= R^T R \stackrel{!}{=} I_{3 \times 3}. \end{aligned}$$

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

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

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

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

¹⁰³ 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.

¹⁰⁴ You can see this, by putting a generic 4×4 matrix Λ , in $\Lambda^T \eta \Lambda = \eta$.

¹⁰⁵ 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.

¹⁰⁶ Recall $\eta_{11} = \eta_{22} = \eta_{33} = -1$ and $\eta_{ij} = 0$ for $i \neq j$.

with the rotation matrices $R_{3 \times 3}$ cited in Eq. 3.23 and derived in Sec. 3.4.1. The corresponding generators are therefore analogous to those we derived for three spatial dimension in Sec. 3.4.1:

$$J_i = \begin{pmatrix} 0 & \\ & J_i^{3dim} \end{pmatrix}. \quad (3.135)$$

For example, from Eq. 3.65 we now have

$$J_1 = \begin{pmatrix} 0 & \\ & J_1^{3dim} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \quad (3.136)$$

To investigate transformations involving time **and** space we will start, as always in Lie theory, with an infinitesimal transformation¹⁰⁷

$$\Lambda_\rho^\mu \approx \delta_\rho^\mu + \epsilon K_\rho^\mu. \quad (3.137)$$

We put this into the defining condition (Eq. 3.125)

$$\begin{aligned} \Lambda_\rho^\mu \eta_{\mu\nu} \Lambda_\sigma^\nu &\stackrel{!}{=} \eta_{\rho\sigma} \\ \rightarrow (\delta_\rho^\mu + \epsilon K_\rho^\mu) \eta_{\mu\nu} (\delta_\sigma^\nu + \epsilon K_\sigma^\nu) &\stackrel{!}{=} \eta_{\rho\sigma} \\ \rightarrow \eta_{\rho\sigma} + \epsilon K_\rho^\mu \eta_{\mu\sigma} + \epsilon K_\sigma^\nu \eta_{\rho\nu} + \underbrace{\epsilon^2 K_\rho^\mu \eta_{\mu\nu} K_\sigma^\nu}_{\approx 0 \text{ because } \epsilon \text{ is infinitesimal} \rightarrow \epsilon^2 \approx 0} &= \eta_{\rho\sigma} \\ \rightarrow K_\rho^\mu \eta_{\mu\sigma} + K_\sigma^\nu \eta_{\rho\nu} &= 0 \end{aligned} \quad (3.138)$$

which reads in matrix form¹⁰⁸

$$K^T \eta = -\eta K. \quad (3.139)$$

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 boost the description we have, for example in 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 Chap. 2.1: A boost along the x-axis. Because we know that $y' = y$ and $z' = z$ the generator is of the form

$$K_x = \begin{pmatrix} \underbrace{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}_{\equiv k_x} & \\ & \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{pmatrix} \quad (3.140)$$

¹⁰⁷ 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.

¹⁰⁸ 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 $K_\rho^\mu \equiv K^\mu_\rho \rightarrow (K^T)^\mu_\rho = K_\rho^\mu$. Matrix multiplication always works by multiplying rows with columns. Therefore $K^\nu_\sigma \eta_{\rho\nu} = \eta_{\rho\nu} K^\nu_\sigma$, where the ρ -row of η is multiplied with the σ -column of K . This term then is in matrix notation ηK . Furthermore, $K_\rho^\mu \eta_{\mu\sigma} = K^\mu_\rho \eta_{\mu\sigma} = (K^T)_\rho^\mu \eta_{\mu\sigma}$. 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.

and we only need to solve a 2×2 matrix equation. Equation 3.139 reduces to

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = - \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

which is solved by¹⁰⁹

$$k_x = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

$$\begin{aligned} &^{109} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \text{ and} \\ & - \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = - \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \end{aligned}$$

The complete generator for boosts along the x-axis is therefore

$$K_x = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad (3.141)$$

and equally we can find the generators for boosts along the y- and z-axis

$$K_y = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad K_z = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (3.142)$$

Now, we already know from Lie theory how we get from the generators to finite transformations¹¹⁰

$$\Lambda_x(\phi) = e^{\phi K_x}$$

For brevity let's focus again on the exciting part of the generator K_x , i.e. the upper left 2×2 matrix k_x , which is defined in Eq. 3.140. We can then evaluate the exponential function using its series expansion and that¹¹¹ $k_x^2 = 1$

$$\begin{aligned} \Lambda_x(\phi) &= e^{\phi k_x} = \sum_{n=0}^{\infty} \frac{\phi^n k_x^n}{n!} = \sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \underbrace{k_x^{2n}}_{=1} + \sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \underbrace{k_x^{2n+1}}_{=k_x} \\ &= \left(\sum_{n=0}^{\infty} \frac{\phi^{2n}}{(2n)!} \right) I + \left(\sum_{n=0}^{\infty} \frac{\phi^{2n+1}}{(2n+1)!} \right) k_x = \cosh(\phi) I + \sinh(\phi) k_x \\ &= \begin{pmatrix} \cosh(\phi) & 0 \\ 0 & \cosh(\phi) \end{pmatrix} + \begin{pmatrix} 0 & \sinh(\phi) \\ \sinh(\phi) & 0 \end{pmatrix} = \begin{pmatrix} \cosh(\phi) & \sinh(\phi) \\ \sinh(\phi) & \cosh(\phi) \end{pmatrix} \end{aligned} \quad (3.143)$$

This computation is analogous to the computation in Sec. 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

¹¹⁰ Take note that the generators K_x, K_y and K_z are already Hermitian: $K_i^\dagger = K_i$. Therefore, we do not include an extra i here in the exponent, because then the generators would be anti-Hermitian.

¹¹¹ As you can easily check: $k_x^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, equally $k_x^4 = 1$ etc. for all even exponents and of course $k_x^3 = k_x$, $k_x^5 = k_x$ etc. for all uneven exponents.

hyperbolic sine $\sinh(\phi)$ and hyperbolic cosine $\cosh(\phi)$. The complete 4×4 transformation matrix for a boost along the x-axis is therefore

$$\Lambda_x = \begin{pmatrix} \cosh(\phi) & \sinh(\phi) & 0 & 0 \\ \sinh(\phi) & \cosh(\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (3.144)$$

Analogously, we can derive the transformation matrices for boosts along the other axes:

$$\Lambda_y = \begin{pmatrix} \cosh(\phi) & 0 & \sinh(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ \sinh(\phi) & 0 & \cosh(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.145)$$

$$\Lambda_z = \begin{pmatrix} \cosh(\phi) & 0 & 0 & \sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(\phi) & 0 & 0 & \cosh(\phi) \end{pmatrix}. \quad (3.146)$$

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 components¹¹² 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 have¹¹³

$$(\Lambda_P)_{\alpha'}^{\alpha} (\Lambda_P)_{\beta'}^{\beta} (J_i)^{\alpha' \beta'} \underbrace{\quad}_{\text{switching to matrix notation}} \triangleq \Lambda_P J_i (\Lambda_P)^T = J_i \triangleq (J_i)^{\alpha \beta} \quad (3.147)$$

$$(\Lambda_P)_{\alpha'}^{\alpha} (\Lambda_P)_{\beta'}^{\beta} (K_i)^{\alpha' \beta'} \underbrace{\quad}_{\text{switching to matrix notation}} \triangleq \Lambda_P K_i (\Lambda_P)^T = -K_i \triangleq -(K_i)^{\alpha \beta}, \quad (3.148)$$

as you can check by brute force computation, using the explicit matrices derived in the last section. For example

$$J_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \rightarrow J'_x = \Lambda_P J_x (\Lambda_P)^T = J_x, \quad (3.149)$$

¹¹² Recall that the Lorentz group is in fact $O(1,3) = \{SO(1,3)^\uparrow, \Lambda_P SO(1,3)^\uparrow, \Lambda_T SO(1,3)^\uparrow, \Lambda_P \Lambda_T SO(1,3)^\uparrow\}$ and we derived in the last section the generators of $SO(1,3)^\uparrow$.

¹¹³ We need two matrices Λ_P , one for each index. This is just the ordinary transformation behaviour of operators under changes of the coordinate system.

because

$$\begin{aligned}\Lambda_P J_i (\Lambda_P)^T &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^T \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (3.150)\end{aligned}$$

In contrast,

$$K_x = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \rightarrow K'_x = \Lambda_P K_x (\Lambda_P)^T = -K_x, \quad (3.151)$$

because

$$\begin{aligned}\Lambda_P J_i (\Lambda_P)^T &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^T \\ &= - \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3.152)\end{aligned}$$

In conclusion, we have under parity transformations

$$J_i \xrightarrow{\text{P}} J_i \quad K_i \xrightarrow{\text{P}} -K_i \quad (3.153)$$

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

$$(\Lambda_T)_{\alpha'}^{\alpha} (\Lambda_T)_{\beta'}^{\beta} (J_i)^{\alpha' \beta'} \stackrel{\text{switching to matrix notation}}{\hat{=}} \Lambda_T J_i (\Lambda_T)^T = J_i \hat{=} (J_i)^{\alpha \beta} \quad (3.154)$$

$$(\Lambda_T)_{\alpha'}^{\alpha} (\Lambda_T)_{\beta'}^{\beta} (K_i)^{\alpha' \beta'} \stackrel{\text{switching to matrix notation}}{\hat{=}} \Lambda_T K_i (\Lambda_T)^T = -(K_i)^{\alpha \beta}, \quad (3.155)$$

Or shorter:

$$J_i \xrightarrow{\text{T}} J_i \quad K_i \xrightarrow{\text{T}} -K_i. \quad (3.156)$$

3.7.3 The Lie Algebra of the Proper Orthochronous Lorentz Group

¹¹⁴ See Eq. 3.141 for the boost generators and Eq. 3.62 for the rotation generators

¹¹⁵ The Levi-Civita symbol ϵ_{ijk} , is defined in appendix B.5.5.

Now using the explicit matrix form of the generators¹¹⁴ for $SO(1,3)^\dagger$ we can derive the corresponding Lie algebra by brute force computation¹¹⁵

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.157)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \quad (3.158)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k \quad (3.159)$$

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

$$\Lambda = e^{iJ\theta + iK\Phi} \quad (3.160)$$

Equation 3.158 tells us that the two generator types (J_i and K_i) do not commute with each other. While the rotation generators are closed under commutation¹¹⁶, the boost generators are not¹¹⁷. We can now define new operators from the old ones that are closed under commutation and commute with each other

$$N_i^\pm = \frac{1}{2}(J_i \pm iK_i). \quad (3.161)$$

Working out the commutation relations yields

$$[N_i^+, N_j^+] = i\epsilon_{ijk}N_k^+ \quad (3.162)$$

$$[N_i^-, N_j^-] = i\epsilon_{ijk}N_k^- \quad (3.163)$$

$$[N_i^+, N_j^-] = 0. \quad (3.164)$$

These are precisely the commutation relations for the Lie algebra of $SU(2)$ and we have therefore discovered that the Lie algebra of $SO(1,3)_+^\dagger$ consists of two copies of the Lie algebra of $SU(2)$.

This is great news, because we already know how to construct all irreducible representations of the Lie algebra of $SU(2)$. However the Lorentz group is, like $SO(3)$, not simply-connected¹¹⁸ and Lie theory

¹¹⁶ Closed under commutation means that the commutator $[J_i, J_j] = J_i J_j - J_j J_i$, is again a rotation generator. From Eq. 3.157 we can see that this is the case.

¹¹⁷ Eq. 3.159 tells us that the commutator of two boost generators K_i and K_j isn't another boost generator, but a generator of rotations.

¹¹⁸ We will use this simply as a fact here, because a proof would lead us too far apart.

tells us that there is, for groups that aren't simply connected, no one-to-one correspondence between the irreducible representations of the Lie algebra and representations of the corresponding group¹¹⁹. Instead, **by deriving the irreducible representations of the 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. It is a good thing that we find additional representations, because we need those representations to describe certain elementary particles.

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 the Lorentz group.

Each irreducible representation of the Lie algebra of $SU(2)$ can be labelled by the scalar value j of the Casimir operator of $SU(2)$. Therefore, we now know that we can label the irreducible representations of the covering group¹²⁰ 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 $j_1, j_2 = 0, \frac{1}{2}, 1 \dots$ representations for the two copies of the $SU(2)$, which we derived earlier.

It is conventional to write the Lorentz algebra in a more compact way using $M_{\mu\nu}$, which is defined by

$$J_i = \frac{1}{2} \epsilon_{ijk} M_{jk}. \quad (3.165)$$

$$K_i = M_{0i}. \quad (3.166)$$

With this definition the Lorentz algebra reads

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho} M_{\nu\sigma} - \eta_{\mu\sigma} M_{\nu\rho} - \eta_{\nu\rho} M_{\mu\sigma} + \eta_{\nu\sigma} M_{\mu\rho}). \quad (3.167)$$

Next, we want to take a look at what irreducible representations we can construct from the Lie algebra of the Lorentz group.

3.7.4 The (0,0) Representation

The lowest order representation is as for $SU(2)$ trivial, because the vector space is 1 dimensional for both copies of the Lie algebra of $SU(2)$. Our generators must therefore be 1×1 matrices and the only 1×1 matrices fulfilling the commutation relations are trivially 0:

$$N_i^+ = N_i^- = 0 \rightarrow e^{N_i^+} = e^{N_i^-} = e^0 = 1 \quad (3.168)$$

¹¹⁹ 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.

¹²⁰ The covering group of the Lorentz group is $SL(2, C)$, the set of 2×2 matrices with unit determinant and complex entries. The relationship $SL(2, C) \rightarrow SO(1, 3)$ is similar to the relationship $SU(2) \rightarrow SO(3)$ we discovered earlier in this text.

Therefore, the $(0,0)$ representation of the Lorentz group acts on objects that do **not** change under Lorentz transformations. This representation is called the Lorentz **scalar representation**.

3.7.5 The $(\frac{1}{2}, 0)$ Representation

¹²¹ Recall that the dimension of our vector space is given by $2j + 1$. Therefore we have here $2\frac{1}{2} + 1 = 2$ dimensions.

In this representation we use the¹²¹ 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 dimensional representation for the other N_i^- , i.e. $N_i^- = 0$ as explained in the last section. From the definition of N^- in Eq. 3.161 we conclude

$$N_i^- = \frac{1}{2}(J_i - iK_i) = 0 \quad (3.169)$$

$$\rightarrow J_i = iK_i. \quad (3.170)$$

Furthermore, we can use that we already derived in Sec. 3.6.4 the two dimensional representation of $SU(2)$:

$$N_i^+ = \frac{\sigma_i}{2} \quad (3.171)$$

where σ_i denotes once more the Pauli matrices, which were defined in Eq. 3.81. On the other hand, we have

$$N_i^+ \underset{\text{Eq. 3.161}}{=} \frac{1}{2}(J_i + iK_i) \underset{\text{Eq. 3.170}}{=} \frac{1}{2}(iK_i + iK_i) = iK_i \quad (3.172)$$

Comparing Eq. 3.171 with Eq. 3.172 tells us that

$$iK_i = \frac{\sigma_i}{2} \rightarrow K_i = \frac{\sigma_i}{2i} = \frac{i\sigma_i}{2i^2} = \frac{-i}{2}\sigma_i \quad (3.173)$$

$$\text{Eq. 3.170} \rightarrow J_i = iK_i = \frac{-i^2}{2}\sigma_i = \frac{1}{2}\sigma_i. \quad (3.174)$$

We conclude that a Lorentz rotation in this representation is given by

$$R_\theta = e^{i\vec{\theta}\vec{J}} = e^{i\vec{\theta}\frac{\vec{\sigma}}{2}} \quad (3.175)$$

and a Lorentz boost by

$$B_\theta = e^{i\vec{\phi}\vec{K}} = e^{\vec{\phi}\frac{\vec{\sigma}}{2}}. \quad (3.176)$$

By writing out the exponential function as series expansion we can easily get the representation of the Lorentz group from the representation of the generators. For example, rotations about the x-axis e.g. are given by

$$R_x(\theta) = e^{i\theta J_1} = e^{i\theta \frac{1}{2}\sigma_1} = 1 + \frac{i}{2}\theta\sigma_1 + \frac{1}{2}\left(\frac{i}{2}\theta\sigma_1\right)^2 + \dots \quad (3.177)$$

And if we use the explicit matrix form of σ_1 as defined in Eq. 3.81, together with the fact that $\sigma_1^2 = 1$ we get¹²²

$$\begin{aligned} R_x(\theta) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{i}{2}\theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \frac{1}{2}\left(\frac{\theta}{2}\right)^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \dots \\ &= \begin{pmatrix} \cos(\frac{\theta}{2}) & i \sin(\frac{\theta}{2}) \\ i \sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{pmatrix}. \end{aligned} \quad (3.178)$$

Analogous we can compute the transformation matrix for rotations around other axes or boosts. One important thing to notice is 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**¹²³ objects this representation acts on are called **left-chiral spinors**¹²⁴:

$$\chi_L = \begin{pmatrix} (\chi_L)_1 \\ (\chi_L)_2 \end{pmatrix} \quad (3.179)$$

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 spinor¹²⁵ is after a rotation by 2π not the same, but gets a minus sign. This is a pretty crazy property, because all objects we deal with in everyday life are exactly the same after a rotation by $360^\circ = 2\pi$.

"One could say that a spinor is the most basic sort of mathematical object that can be Lorentz-transformed."

- A. M. Steane¹²⁶

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. 3.161

$$N_i^+ = \frac{1}{2}(J_i + iK_i) = 0 \quad (3.180)$$

¹²² The steps are completely analogous to what we did in Sec. 3.4.1

¹²³ We will learn later that these two components correspond to **spin-up** and **spin-down** states.

¹²⁴ 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.

¹²⁵ 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

¹²⁶ Andrew M. Steane. An introduction to spinors. *ArXiv e-prints*, December 2013

$$\rightarrow J_i = -iK_i. \quad (3.181)$$

Take notice of the minus sign. Using the two-dimensional representation of $SU(2)$ for N^+ , which was derived in Sec. 3.6.4, yields

$$N_i^- = \frac{1}{2}\sigma_i = \frac{1}{2}(J_i - iK_i) \underset{\text{Eq. 3.181}}{=} \frac{1}{2}(-iK_i - iK_i) = -iK_i \quad (3.182)$$

From this we get the $(0, \frac{1}{2})$ representation of the generators

$$-iK_i = \frac{1}{2}\sigma_i \rightarrow K_i = \frac{-1}{2i}\sigma_i = \frac{-i}{2i^2}\sigma_i = \frac{i}{2}\sigma_i. \quad (3.183)$$

And from Eq. 3.181 we get

$$J_i = -iK_i = \frac{1}{2}\sigma_i. \quad (3.184)$$

We conclude that in this representation a Lorentz rotation is given by

$$R_\theta = e^{i\vec{\theta}\vec{J}} = e^{i\vec{\theta}\frac{\vec{\sigma}}{2}} \quad (3.185)$$

and a Lorentz boost by

$$B_\theta = e^{i\vec{\phi}\vec{K}} = e^{-\vec{\phi}\frac{\vec{\sigma}}{2}}. \quad (3.186)$$

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**:

$$\chi_R = \begin{pmatrix} (\chi_R)^1 \\ (\chi_R)^2 \end{pmatrix} \quad (3.187)$$

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. We will learn in a moment that they are different, but not too different. In fact, 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

$$\chi_L = \chi_a \quad (3.188)$$

and a right-chiral spinor χ_R has an upper, dotted index

$$\chi_R = \chi^{\dot{a}}. \quad (3.189)$$

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 metric¹²⁷ as

$$\epsilon^{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (3.190)$$

and show that it has the desired properties. Furthermore, we define¹²⁸

$$\chi_L^C \equiv \epsilon \chi_L^* \quad (3.191)$$

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 behaviour 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

$$(-\epsilon)(\epsilon) = 1 \quad (3.192)$$

and

$$(\epsilon)\sigma_i^*(-\epsilon) = -\sigma_i \quad (3.193)$$

for each Pauli matrix σ_i , as you can check. Transforming yields¹²⁹

$$\begin{aligned} \chi_L^C &\rightarrow \chi'^C_L = \epsilon(\chi'_L)^* \\ &= \epsilon(e^{\frac{\vec{\phi}}{2}\vec{\sigma}}\chi_L)^* \\ &= \epsilon(e^{\frac{\vec{\phi}}{2}\vec{\sigma}}(-\epsilon)(\epsilon)\chi_L)^* \\ &\quad = 1 \text{ see Eq. 3.192} \\ &= \epsilon(e^{\frac{\vec{\phi}}{2}\vec{\sigma}^*}(-\epsilon)(\epsilon)\chi_L^*) \\ &\quad \text{Eq. 3.193: } = e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \\ &= e^{-\frac{\vec{\phi}}{2}\vec{\sigma}} \underbrace{\epsilon\chi_L^*}_{=\chi_L^C} \\ &= e^{-\frac{\vec{\phi}}{2}\vec{\sigma}}\chi_L^C, \end{aligned} \quad (3.194)$$

¹²⁷ Take note that this is the Levi-Civita symbol in two dimensions as defined in appendix B.5.5.

¹²⁸ 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 Sec. 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.

¹²⁹ We use the notation $\vec{\phi}\vec{\sigma} = \sum_i \sigma_i \phi_i \underbrace{=}_{\text{summation convention}} \sigma_i \phi_i$. The "vector" $\vec{\sigma}$ shouldn't be taken too seriously, because it's just a shorthand, conventional notation.

¹³⁰ The transformation behaviour of right-chiral spinors under boosts was derived in Eq. 3.186: $B_\theta = e^{i\vec{\phi}\vec{K}} = e^{-\vec{\phi}\frac{\vec{\sigma}}{2}}$. Compare this to how left-chiral spinors transform under boosts, as derived in Eq. 3.176: $B_\theta = e^{i\vec{\phi}\vec{K}} = e^{\vec{\phi}\frac{\vec{\sigma}}{2}}$

which is exactly the transformation behaviour of a right-chiral spinor¹³⁰. To get to the fifth line, we use the series expansion of $e^{\frac{\vec{\phi}}{2}\vec{\sigma}}$ and Eq. 3.193 on every term. You can check in the same way that the behaviour 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:

$$\chi_L^C \rightarrow \chi_L'^C = \epsilon(\chi_L')^* = \epsilon(e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\chi_L)^* = e^{\frac{i\vec{\theta}}{2}\vec{\sigma}}\epsilon(\chi_L)^*. \quad (3.195)$$

Furthermore, you can check that ϵ is invariant under all transformations and that if you want to go the other way round, i.e. transform a right-chiral spinor into a left-chiral spinor you have to use $(-\epsilon)$.

Therefore, we define in analogy with the tensor notation of special relativity that our "metric" raises and lowers indices

$$\epsilon\chi_L \underbrace{=}_{\text{written in index notation}} \epsilon^{ac}\chi_c = \chi^a \quad (3.196)$$

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

$$\chi_R = \epsilon\chi_L^* \quad (3.197)$$

This means that complex conjugation transforms an undotted index into a dotted index:

$$\chi_R = \epsilon\chi_L^* = \chi^{\dot{a}}. \quad (3.198)$$

Therefore, we can get a lower, dotted index by complex conjugating χ_L

$$\chi_L^* = \chi_a^* = \chi_{\dot{a}} \quad (3.199)$$

and an upper, undotted index, by complex conjugating χ_R

$$\chi_R^* = (\chi^{\dot{a}})^* = \chi^a \quad (3.200)$$

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 at all under Lorentz transformations¹³¹. From the transformation behaviour of a left-chiral spinor

$$\chi_L = \chi_a \rightarrow \chi'_a = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^b \chi_b \quad (3.201)$$

we can derive how a spinor with lower, dotted index transforms:

$$\begin{aligned} \chi_L^* = \chi_a^* = \chi_{\dot{a}} \rightarrow \chi'_{\dot{a}} &= (\chi'_a)^* = \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right)_a^b\right)^* \chi_b^* \\ &= \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}}\right)_{\dot{a}}^b \chi_b^* \end{aligned} \quad (3.202)$$

¹³¹ 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.

Analogously, we use the transformation behaviour of a right-chiral spinor:

$$\chi_R \rightarrow \chi'_R = \chi'^{\dot{a}} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^{\dot{a}}_b \chi^b \quad (3.203)$$

to derive how a spinor with upper, undotted index transforms:

$$\begin{aligned} \chi_R^* = (\chi^{\dot{a}})^* = \chi^a \rightarrow \chi'^a = (\chi'^{\dot{a}})^* &= \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^{\dot{a}}_b \right)^* (\chi^b)^* \\ &= \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2} - \vec{\phi}\frac{\vec{\sigma}^*}{2}} \right)^a_b \chi^b \end{aligned} \quad (3.204)$$

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 mustn't forget to 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 σ_i^* in the exponent, for example, $e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2}}$. Together with transposing this becomes the Hermitian conjugate: $\sigma_i^\dagger = (\sigma_i^*)^T$, where the symbol \dagger is called "dagger". The Hermitian conjugate of every Pauli matrix, is again the same Pauli matrix

$$\sigma_i^\dagger = (\sigma_i^*)^T = \sigma_i, \quad (3.205)$$

as you can easily check by looking at the explicit form of the Pauli matrices, as defined in Eq. 3.81.

By comparing Eq. 3.201 with Eq. 3.204 and using Eq. 3.205, we see that the transformation behaviour 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, because¹³²

$$\begin{aligned} (\chi^a)^T \chi_a &\rightarrow (\chi'^a)^T \chi'_a = \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}^*}{2} - \vec{\phi}\frac{\vec{\sigma}^*}{2}} \right)^a_b \chi^b \right)^T \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^c_a \chi_c \\ &= (\chi^b)^T \left(e^{-i\vec{\theta}\frac{(\sigma^*)^T}{2} - \vec{\phi}\frac{(\sigma^*)^T}{2}} \right)^a_b \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^c_a \chi_c \\ &\stackrel{\text{Eq. 3.205}}{=} (\chi^b)^T \underbrace{\left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^a_b \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)^c_a}_{=\delta_b^c} \chi_c \\ &= (\chi^c)^T \chi_c \end{aligned} \quad (3.206)$$

In the same way we can combine an upper, dotted index with a lower, dotted index as you can verify by comparing Eq. 3.202 with

¹³² 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 $b \neq c$.

Eq. 3.203. 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

$$\chi_R^T \chi_L = (\chi^{\dot{a}})^T \chi_a \rightarrow (\chi'^{\dot{a}})^T \chi'_a = \chi^{\dot{b}} \underbrace{\left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} - \vec{\phi} \frac{\vec{\sigma}}{2}} \right)^{\dot{a}}_{\dot{b}}}_{\neq \delta^{\dot{a}}_{\dot{b}}} \left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)^c_a \chi_c \quad (3.207)$$

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 type¹³³ in order to get Lorentz invariant terms. Or formulated differently, we must combine the complex conjugate of a right-chiral spinor with a left-chiral spinor $\chi_R^\dagger \chi_L = (\chi_R^*)^T \chi_L \hat{=} (\chi^a)^T \chi_a$, or the complex conjugate of a left-chiral spinor with a right-chiral spinor $\chi_L^\dagger \chi_R = (\chi_L^*)^T \chi_R = (\chi_{\dot{a}})^T \chi^{\dot{a}}$ to get Lorentz invariant terms. We will use this later, when we look for 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. 3.206, can be written as

$$\chi_a^T \chi^a \underbrace{=} \chi_a^T \epsilon^{ab} \chi_b. \quad (3.208)$$

Eq. 3.196

Compare this to how we defined in Eq. 2.31 the invariant product of Minkowski space, using the Minkowski metric $\eta^{\mu\nu}$:

$$x_\mu y^\mu = x_\mu \eta^{\mu\nu} y_\nu. \quad (3.209)$$

The spinor metric is indeed what the Minkowski metric is for four-vectors¹³⁴.

After setting up this notation we can now write the spinor "metric" with lowered indices

$$\epsilon_{ab} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad (3.210)$$

because we need¹³⁵ $(-\epsilon)$ 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:

$$\chi_R \rightarrow \chi'_R = \chi'^{\dot{a}} = \Lambda^{\dot{a}}_{\dot{b}} \chi^{\dot{b}} = \left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} - \vec{\phi} \frac{\vec{\sigma}}{2}} \right)^{\dot{a}}_{\dot{b}} \chi^{\dot{b}} \quad (3.211)$$

and analogous for left-chiral spinors

$$\chi_L \rightarrow \chi'_L = \chi'_a = \Lambda_a^b \chi_b = \left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)^b_a \chi_b. \quad (3.212)$$

¹³³ In this context dotted \dot{a} or undotted \dot{a}

¹³⁴ 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 .

¹³⁵ You can check this yourself, but it's not very important for what follows.

Therefore:

$$\Lambda_{(\frac{1}{2},0)} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right) \triangleq \Lambda_a^b \quad (3.213)$$

and

$$\Lambda_{(0,\frac{1}{2})} = \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} - \vec{\phi}\frac{\vec{\sigma}}{2}} \right) \triangleq \Lambda_{\dot{b}}^{\dot{a}} \quad (3.214)$$

This notation will prove to be very 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 $(\frac{1}{2}, \frac{1}{2})$ Representation

For this representation we use the 2-dimensional representation, for both copies of the $SU(2)$ Lie algebra¹³⁶ 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. $[N_i^+, N_j^-] = 0$. 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 $SU(2)$. Here the notation we introduced in the last section comes in handy.

We know from the fact that both indices can take on two values ($\frac{1}{2}$ and $-\frac{1}{2}$), because each representation is 2 dimensional, that our object v will have 4 components. Therefore, the objects can be 2×2 matrices, but it's also possible to enforce a four component vector form, as we will see¹³⁷.

But 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 and as explained in Sec. 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

¹³⁶ Mathematically we have $(\frac{1}{2}, \frac{1}{2}) = (\frac{1}{2}, 0) \otimes (0, \frac{1}{2})$

¹³⁷ Remember that when we talked about rotations of the plane we were in the same situation. The rotation could be described by complex numbers acting on complex numbers. Doing the map to real matrices we had real matrices acting on real matrices, but the same action could be described by a real matrix acting on a column vector.

¹³⁸ This means that an arbitrary Hermitian 2×2 matrix can be written as a linear combination of the form: $a_0 1 + a_i \sigma_i$

¹³⁹ Defined in Eq. 3.81

¹⁴⁰ 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 behaviour we derived earlier for vectors.

that our irreducible representation acts on Hermitian 2×2 matrices.

A basis¹³⁸ for Hermitian 2×2 matrices is given by the σ matrices¹³⁹ together with the identity matrix.

Instead of examining v_a^b , we will have a look at v_{ab} , because then we can use the Pauli matrices as defined in Eq. 3.81. Take note that v_a^b and v_{ab} can be transformed into each other by multiplication with $\epsilon^{b\dot{c}}$ and therefore if you want to work with v_a^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

$$v_{ab} = v_\nu \sigma_{ab}^\nu = v^0 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + v^1 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + v^2 \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + v^3 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.215)$$

As explained above, we could use¹⁴⁰ $v_a^b = v_\mu \sigma_{ac}^\mu \epsilon^{b\dot{c}}$ instead, which means we would use the basis $(\tilde{\sigma}^b)^\mu = \sigma_{ac}^\mu \epsilon^{b\dot{c}}$. We therefore write a general Hermitian matrix as

$$v_{ab} = \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix}. \quad (3.216)$$

Remember that we have learned in the last section that different indices transform differently. To be specific: A lower dotted index transforms differently than a lower undotted index.

Now we have a look at how v_{ab} transforms and use the transformation operator for an lower undotted index as derived in Eq. 3.202

$$\begin{aligned} v \rightarrow v' = v'_{ab} &= \left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(\left(e^{-i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)_b^{\dot{d}} \right)^T \\ &= \left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)_b^{\dot{d}} \\ &\stackrel{\sigma_i^\dagger = \sigma_i}{=} \left(e^{i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta} \frac{\vec{\sigma}}{2} + \vec{\phi} \frac{\vec{\sigma}}{2}} \right)_b^{\dot{d}} \end{aligned} \quad (3.217)$$

¹⁴¹ 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 $v_{cd}^\dagger = v_{cd}$ that $v_{cd}^\dagger = -v_{cd}$.

We can now see that a Hermitian matrix is after such a transformation still Hermitian, as promised above¹⁴¹

$$\begin{aligned}
\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d &\rightarrow \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d \right)^\dagger \\
&\stackrel{(ABC)^\dagger = C^\dagger B^\dagger A^\dagger}{=} \left(\left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d \right)^\dagger v_{cd}^\dagger \left(\left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c \right)^\dagger \\
&= \left(e^{i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2} + \vec{\phi}\frac{\vec{\sigma}^\dagger}{2}} \right)_b^d v_{cd}^\dagger \left(e^{-i\vec{\theta}\frac{\vec{\sigma}^\dagger}{2} + \vec{\phi}\frac{\vec{\sigma}^\dagger}{2}} \right)_a^c \\
&\stackrel{\text{if } v_{cd}^\dagger = v_{cd}}{=} \left(e^{i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_a^c v_{cd} \left(e^{-i\vec{\theta}\frac{\vec{\sigma}}{2} + \vec{\phi}\frac{\vec{\sigma}}{2}} \right)_b^d \quad \checkmark
\end{aligned} \tag{3.218}$$

The explicit computation¹⁴² for an arbitrary transformation is long and tedious so we will look at one specific example. Let's boost v along the z-axis¹⁴³

$$\begin{aligned}
v_{ab} \rightarrow v'_{ab} &= \left(e^{\phi\frac{\sigma_3}{2}} \right)_a^c v_{cd} \left(e^{\phi\frac{\sigma_3}{2}} \right)_b^d \\
&= \begin{pmatrix} e^{\frac{\phi}{2}} & 0 \\ 0 & e^{-\frac{\phi}{2}} \end{pmatrix} \begin{pmatrix} v_0 + v_3 & v_1 - iv_2 \\ v_1 + iv_2 & v_0 - v_3 \end{pmatrix} \begin{pmatrix} e^{\frac{\phi}{2}} & 0 \\ 0 & e^{-\frac{\phi}{2}} \end{pmatrix} \\
&= \begin{pmatrix} e^\phi(v_0 + v_3) & v_1 - iv_2 \\ v_1 + iv_2 & e^{-\phi}(v_0 - v_3) \end{pmatrix}
\end{aligned} \tag{3.219}$$

where we have used the fact that σ_3 is diagonal¹⁴⁴ and that

$e^A = \begin{pmatrix} e^{A_{11}} & 0 \\ 0 & e^{A_{22}} \end{pmatrix}$ holds for every diagonal matrix. Comparing the transformed object we computed in Eq. 3.219 with a generic object v' yields

$$v'_{ab} = \begin{pmatrix} v'_0 + v'_3 & v'_1 - iv'_2 \\ v'_1 + iv'_2 & v'_0 - v'_3 \end{pmatrix} = \begin{pmatrix} e^\phi(v_0 + v_3) & v_1 - iv_2 \\ v_1 + iv_2 & e^{-\phi}(v_0 - v_3) \end{pmatrix}$$

This tells us how the components of the transformed object are related to the untransformed components¹⁴⁵

$$\rightarrow v'_0 + v'_3 = e^\phi(v_0 + v_3) = (\cosh(\phi) + \sinh(\phi))(v_0 + v_3)$$

$$\rightarrow v'_0 - v'_3 = e^{-\phi}(v_0 - v_3) = (\cosh(\phi) - \sinh(\phi))(v_0 - v_3).$$

The addition and subtraction of both equations yields

$$\rightarrow v'_0 = \cosh(\phi)v_0 + \sinh(\phi)v_3$$

$$\rightarrow v'_3 = \sinh(\phi)v_0 + \cosh(\phi)v_3 \tag{3.220}$$

¹⁴² See, for example, page 128 in Matthew Robinson. *Symmetry and the Standard Model*. Springer, 1st edition, August 2011. ISBN 978-1-4419-8267-4

¹⁴³ This means $\vec{\phi} = (0, 0, \phi)^T$. Such a boost is the most easiest because σ_3 is diagonal. For boosts along other axes the exponential series must be evaluated in detail.

¹⁴⁴ $\sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$

¹⁴⁵ 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: $\cosh(\phi) \equiv \frac{1}{2}(e^\phi + e^{-\phi})$ and $\sinh(\phi) \equiv \frac{1}{2}(e^\phi - e^{-\phi})$ or rewrite the few equations here in terms of e^ϕ and $e^{-\phi}$, which is equally good.

¹⁴⁶ See 3.146 for the explicit form of the matrix for a boost along the z-axis.

which is exactly what we get using the 4-vector formalism¹⁴⁶

$$\begin{aligned} \begin{pmatrix} v'_0 \\ v'_1 \\ v'_2 \\ v'_3 \end{pmatrix} &= \begin{pmatrix} \cosh(\phi) & 0 & 0 & \sinh(\phi) \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \sinh(\phi) & 0 & 0 & \cosh(\phi) \end{pmatrix} \begin{pmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{pmatrix} \\ &= \begin{pmatrix} \cosh(\phi)v_0 + \sinh(\phi)v_3 \\ v_1 \\ v_2 \\ \sinh(\phi)v_0 + \cosh(\phi)v_3 \end{pmatrix}. \end{aligned} \quad (3.221)$$

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 tensors¹⁴⁷. 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.

When we started our studies of the Lorentz group, we noted that it consists of four components. These components are connected by the parity and the time-reversal operator¹⁴⁸. 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 will restrict to parity transformations, because it turns out that nature isn't always symmetric under parity transformations, which we will discuss in later chapters. Similar to what we discuss in the next section it's possible to derive representations of the time-reversal operator.

3.7.9 Spinors and Parity

Up to this point, there is no justification for why we called 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-

¹⁴⁷ A rank-2 tensor is simply a matrix $M_{\mu\nu}$.

¹⁴⁸ See Eq. 3.134

chiral. After talking a bit about parity transformation, this will make sense.

Recall that we already know the behaviour of the generators of the Lorentz group under parity transformations. The result was Eq. 3.153, which we recite here for convenience

$$J_i \xrightarrow{\text{P}} J_i \quad K_i \xrightarrow{\text{P}} -K_i \quad (3.222)$$

By looking at the definition of the generators N^\pm in Eq. 3.161, which we recite here, too

$$N_i^\pm = \frac{1}{2}(J_i \pm iK_i) \quad (3.223)$$

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 spinors¹⁴⁹. Just as a right-handed coordinate system changes into a left-handed coordinate system under parity transformations, these two representations change into each other.

Rotational transformations look the same for both representations, but boost transformations differ by a sign and it is easy to make the above statement explicit:

$$(\Lambda_{\vec{K}})_{(\frac{1}{2}, 0)} = e^{\vec{\phi} \vec{K}} \xrightarrow{\text{P}} e^{-\vec{\phi} \vec{K}} = (\Lambda_{\vec{K}})_{(0, \frac{1}{2})} \quad (3.224)$$

$$(\Lambda_{\vec{K}})_{(0, \frac{1}{2})} = e^{-\vec{\phi} \vec{K}} \xrightarrow{\text{P}} e^{\vec{\phi} \vec{K}} = (\Lambda_{\vec{K}})_{(\frac{1}{2}, 0)}. \quad (3.225)$$

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**

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_a \\ \xi^{\dot{a}} \end{pmatrix} \quad (3.226)$$

Recalling 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

$$\Psi_M = \begin{pmatrix} \chi_L \\ \chi_R \end{pmatrix} \quad (3.227)$$

¹⁴⁹ The conventional name is left- and right-handed spinors, but this can be quite confusing, because the notions left-handed and right-handed are directly related to a concept called helicity, which is different from chirality. Anyway the name should make some sense, because something left is changed into something right under parity transformations.

¹⁵⁰ This is a **reducible** representation, which is obvious because of the block-diagonal form of the transformation 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.

is a special case, called Majorana spinor. A Dirac or Majorana spinor is **not a four-vector**, because it transforms completely differently. A Dirac spinor transforms according to the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation¹⁵⁰ of the Lorentz group, which means nothing more than writing the corresponding transformations in block-diagonal form into one big matrix:

$$\Psi \rightarrow \Psi' = \Lambda_{(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})} \Psi = \begin{pmatrix} \Lambda_{(\frac{1}{2}, 0)} & 0 \\ 0 & \Lambda_{(0, \frac{1}{2})} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \quad (3.228)$$

For example, a boost transformation is in this representation

$$\Psi \rightarrow \Psi' = \begin{pmatrix} e^{\frac{\vec{\phi}}{2} \vec{\sigma}} & 0 \\ 0 & e^{-\frac{\vec{\phi}}{2} \vec{\sigma}} \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}. \quad (3.229)$$

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 $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation), because we know that under parity transformations $N^+ \leftrightarrow N^-$ and therefore

$$\left(0, \frac{1}{2}\right) \underbrace{\leftrightarrow}_{\text{P}} \left(\frac{1}{2}, 0\right). \quad (3.230)$$

We conclude, if a Dirac spinor transforms according to the $(\frac{1}{2}, 0) \oplus (0, \frac{1}{2})$ representation, the parity transformed object transforms according to the $(0, \frac{1}{2}) \oplus (\frac{1}{2}, 0)$ representation.

$$\Psi^P \rightarrow (\Psi^P)' = \Lambda_{(0, \frac{1}{2}) \oplus (\frac{1}{2}, 0)} \Psi^P = \begin{pmatrix} \Lambda_{(0, \frac{1}{2})} & 0 \\ 0 & \Lambda_{(\frac{1}{2}, 0)} \end{pmatrix} \begin{pmatrix} \xi_R \\ \chi_L \end{pmatrix}. \quad (3.231)$$

Therefore

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \rightarrow \Psi^P = \begin{pmatrix} \xi_R \\ \chi_L \end{pmatrix}. \quad (3.232)$$

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 $\xi_L \rightarrow \xi_R$, which is a different kind of transformation we will talk about in the next section.

3.7.10 Spinors and Charge Conjugation

In Sec. 3.7.7 stumbled upon a transformation, which yields $\chi_L \rightarrow \chi_R$ and $\tilde{\zeta}_R \rightarrow \tilde{\zeta}_L$. The transformation is $\chi_L \rightarrow \chi_L^C = \epsilon \chi_L^* = \chi_R$ and analogously for a right-chiral spinor $\tilde{\zeta}_R \rightarrow \tilde{\zeta}_R^C = (-\epsilon) \tilde{\zeta}_R^* = \tilde{\zeta}_L$. This transformation is not part of the Lorentz group and 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:

$$\Psi = \begin{pmatrix} \chi_L \\ \tilde{\zeta}_R \end{pmatrix} \rightarrow \tilde{\Psi} = \begin{pmatrix} \chi_L^C \\ \tilde{\zeta}_R^C \end{pmatrix} = \begin{pmatrix} \chi_R \\ \tilde{\zeta}_L \end{pmatrix}. \quad (3.233)$$

Unfortunately, this object does not transform like a Dirac spinor¹⁵¹, which transform under boosts

$$\Psi \rightarrow \Psi' = \begin{pmatrix} e^{\frac{\vec{\theta}}{2} \vec{\sigma}} & 0 \\ 0 & e^{-\frac{\vec{\theta}}{2} \vec{\sigma}} \end{pmatrix} \begin{pmatrix} \chi_L \\ \tilde{\zeta}_R \end{pmatrix}. \quad (3.234)$$

The object $\tilde{\Psi}$ we get from the naive operation, transform as

$$\tilde{\Psi} \rightarrow \tilde{\Psi}' = \begin{pmatrix} e^{-\frac{\vec{\theta}}{2} \vec{\sigma}} & 0 \\ 0 & e^{\frac{\vec{\theta}}{2} \vec{\sigma}} \end{pmatrix} \begin{pmatrix} \chi_L \\ \tilde{\zeta}_R \end{pmatrix}. \quad (3.235)$$

This is a different kind of object, because it transforms according to a different representation of the Lorentz group. Therefore we write

$$\Psi = \begin{pmatrix} \chi_L \\ \tilde{\zeta}_R \end{pmatrix} \rightarrow \Psi^C = \begin{pmatrix} \tilde{\zeta}_R^C \\ \chi_L^C \end{pmatrix} = \begin{pmatrix} \tilde{\zeta}_L \\ \chi_R \end{pmatrix}, \quad (3.236)$$

which incorporates the transformation behaviour 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 particles¹⁵². 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 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

¹⁵¹ 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.

¹⁵² 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 $-\frac{1}{2}$.

of this text. Nevertheless, there is another representation, the infinite-dimensional representation, that is especially interesting, because we need it to transform physical fields.

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 so far. 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

$$\Phi_a \rightarrow \Phi'_a = M_{ab}(\Lambda)\Phi_b \quad (3.237)$$

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 coordinates¹⁵³ $\Phi = \Phi(x)$ and these coordinates are affected by the Lorentz transformations, too. In general we have

$$x^\mu \rightarrow \Lambda^\mu_\nu x^\nu, \quad (3.238)$$

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 case¹⁵⁴

$$\Phi_a(x) \rightarrow M_{ab}(\Lambda)\Phi_b(\Lambda x). \quad (3.239)$$

Our transformation will therefore consist of two parts. One part, represented by a finite-dimensional representation, acting on Φ_a and a second part acting on the coordinates. This second part will act on an infinite-dimensional¹⁵⁵ vector space and we therefore need an infinite-dimensional representation. The infinite-dimensional representation of the Lorentz group is given by differential operators¹⁵⁶

$$M_{\mu\nu}^{\text{inf}} = i(x^\mu \partial^\nu - x^\nu \partial^\mu) \quad (3.240)$$

you can check by straightforward computation that $M_{\mu\nu}^{\text{inf}}$ satisfies the Lorentz algebra (Eq. 3.167) and transforms the coordinates as desired.

¹⁵³ Here x is a shorthand notation for all spacetime coordinates t, x, y, z

¹⁵⁴ Most books use the Wigner convention for symmetry operators: $\Phi_a(x) \rightarrow M_{ab}(\Lambda)\Phi_b(\Lambda^{-1}x)$, but unfortunately there is at this point no way to motivate this convention.

¹⁵⁵ 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.

¹⁵⁶ The symbols ∂^ν are a shorthand notation for the partial derivative $\frac{\partial}{\partial x^\nu}$.

The transformation of the coordinates is now given by¹⁵⁷

$$\Phi(\Lambda x) = e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}^{\text{inf}}}\Phi(x), \quad (3.241)$$

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:

$$\Phi_a(x) \rightarrow \left(e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}^{\text{fin}}} \right)_a^b e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}^{\text{inf}}} \Phi_b(x). \quad (3.242)$$

Because our matrices $M_{\mu\nu}^{\text{fin}}$ are finite-dimensional and constant we can put the two exponents together

$$\Phi_a(x) \rightarrow \left(e^{-i\frac{\omega^{\mu\nu}}{2}M_{\mu\nu}} \right)_a^b \Phi_b(x) \quad (3.243)$$

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 result in a mixing of components and therefore, we need no finite-dimensional representation, but 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 Poincare group, which is the topic of the next section. Nevertheless, we will introduce the infinite-dimensional representation for this kind of transformation 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

$$\Phi(x) \rightarrow \Phi(x + \epsilon) = \Phi(x) + \underbrace{\partial_x \Phi(x)}_{\text{"rate of change" along the x-axis}} \epsilon,$$

which is, of course, again 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

$$P_i \equiv -i\partial_i. \quad (3.244)$$

With this definition an arbitrary, finite translation is

$$\Phi(x) \rightarrow \Phi(x + a) = e^{-ia^i P_i} \Phi(x) = e^{a^i \partial_i} \Phi(x)$$

where a^i denotes the amount we want to translate in each direction. If we write the exponential function as Taylor series¹⁵⁸, this equation can simply be seen as the Taylor expansion¹⁵⁹ 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 $P_i = -i\partial_i$.

¹⁵⁷ Recall the definition of $M^{\mu\nu}$ in Eq. 3.165. The components of $\omega_{\mu\nu}$ can then be directly related to the usual rotation angles $\theta_i = \frac{1}{2}\epsilon_{ijk}\omega_{jk}$ and the boost parameters $\phi_i = \omega_{0i}$

¹⁵⁸ This is derived in appendix B.4.1.

¹⁵⁹ As derived in appendix B.3.

3.8 The Poincare Group

Let's move on to the full spacetime symmetry group of nature: the Poincare 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. Or equivalently, the speed of light does not depend on the choice of where we put the origin of the coordinate system we use to describe some process. When we add these symmetries to the Lorentz group we get the Poincare group¹⁶⁰

¹⁶⁰ The Poincare 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.

$$\begin{aligned}\text{Poincare group} &= \text{Lorentz group} \textbf{ plus } \text{translations} \\ &= \text{Rotations} \textbf{ plus } \text{boosts} \textbf{ plus } \text{translations} \quad (3.245)\end{aligned}$$

The generators of the Poincare group are the generators of the Lorentz group J_i, K_i plus the generators of translations in Minkowski space P_μ .

¹⁶¹ This is not very enlightening, but included for completeness.

In terms of J_i, K_i and P_μ the algebra reads¹⁶¹

$$[J_i, J_j] = i\epsilon_{ijk}J_k \quad (3.246)$$

$$[J_i, K_j] = i\epsilon_{ijk}K_k \quad (3.247)$$

$$[K_i, K_j] = -i\epsilon_{ijk}J_k \quad (3.248)$$

$$[J_i, P_j] = i\epsilon_{ijk}P_k \quad (3.249)$$

$$[J_i, P_0] = 0 \quad (3.250)$$

$$[K_i, P_j] = i\delta_{ij}P_0 \quad (3.251)$$

$$[K_i, P_0] = -iP_i \quad (3.252)$$

Because this looks like a huge mess it is conventional to write this in terms of $M_{\mu\nu}$, which was defined by

$$J_i = \frac{1}{2}\epsilon_{ijk}M_{jk} \quad (3.253)$$

$$K_i = M_{0i}. \quad (3.254)$$

With $M_{\mu\nu}$ the Poincare algebra reads

$$[P_\mu, P_\nu] = 0 \quad (3.255)$$

$$[M_{\mu\nu}, P_\rho] = i(\eta_{\mu\rho}P_\nu - \eta_{\nu\rho}P_\mu) \quad (3.256)$$

and of course again

$$[M_{\mu\nu}, M_{\rho\sigma}] = i(\eta_{\mu\rho}M_{\nu\sigma} - \eta_{\mu\sigma}M_{\nu\rho} - \eta_{\nu\rho}M_{\mu\sigma} + \eta_{\nu\sigma}M_{\mu\rho}) \quad (3.257)$$

For this quite complicated group it is very useful to label the representations by using the fixed scalar values of the Casimir operators. The Poincare group has two Casimir operators¹⁶². The first one is:

$$P_\mu P^\mu =: m^2. \quad (3.258)$$

We give the scalar value the suggestive name m^2 , because we will learn later that it coincides with the mass of particles¹⁶³.

The second Casimir operator is $W_\mu W^\mu$ with¹⁶⁴

$$W^\mu = \frac{1}{2}\epsilon^{\mu\nu\rho\sigma}P_\nu M_{\rho\sigma} \quad (3.259)$$

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 $j \equiv j_1 + j_2$, which 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 two¹⁶⁵ representations of the $SU(2)$ algebra.

For example, the $(j_1, j_2) = (0, 0)$ representation is called spin 0 representation¹⁶⁶. The $(j_1, j_2) = (\frac{1}{2}, 0)$ and $(j_1, j_2) = (0, \frac{1}{2})$ are both called spin $\frac{1}{2}$ representations¹⁶⁷ and analogously the $(j_1, j_2) = (\frac{1}{2}, \frac{1}{2})$ representation is called spin 1 representation¹⁶⁸.

The message to take away is that each representation is labelled by two scalar values: m and j . m can take on arbitrary values, but j is restricted to half-integer or integer values.

3.9 Elementary Particles

The labels for the irreducible representations of the Poincare group are how elementary particles are **labelled** in physics¹⁶⁹: by their **mass** m and by their **spin** ($= j$ here). 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 Poincare group.

¹⁶² Recall that a Casimir operator is defined as an operator, constructed from the generators, that commutes with all other generators.

¹⁶³ Don't worry, this will make much more sense later.

¹⁶⁴ $\epsilon^{\mu\nu\rho\sigma}$ is the four-dimensional Levi-Civita symbol, which is defined in appendix B.5.5.

¹⁶⁵ Recall that the 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) .

¹⁶⁶ $j_1 + j_2 = 0 + 0 = 0$

¹⁶⁷ $j_1 + j_2 = \frac{1}{2} + 0 = 0 + \frac{1}{2} = \frac{1}{2}$

¹⁶⁸ $j_1 + j_2 = \frac{1}{2} + \frac{1}{2} = 1$

¹⁶⁹ Some prefer to say: Elementary particles **are** the irreducible representations of the Poincare 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

- mass: $9,109 \cdot 10^{-32}$ 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 with¹⁷⁰

¹⁷⁰ 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.

- 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 Poincare group are the **mathematical tools we need to describe all elementary particles**. To describe **scalar** particles, like the Higgs Boson, we use mathematical objects, called **scalars**, that transform 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 Sec. 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

- **John Stillwell - Naive Lie Theory**¹⁷¹ is a very readable, math orientated introduction to Lie Theory.
- **N. Jeevanjee - An Introduction to Tensors and Group Theory for Physicists**¹⁷² is a very good introduction, with focus on the usage of Group Theory in physics.

¹⁷¹ John Stillwell. *Naive Lie Theory*. Springer, 1st edition, 8 2008b. ISBN 9780387782140

¹⁷² Nadir Jeevanjee. *An Introduction to Tensors and Group Theory for Physicists*. Birkhaeuser, 1st edition, August 2011. ISBN 978-0817647148

3.10 Appendix: Rotations in a Complex Vector Space

The concept of transformations that preserve the inner product can be used with complex vector spaces, too. 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, where a complex number would make little sense. Therefore, the inner product of complex vector spaces is defined with additional complex conjugation¹⁷³

$$a \cdot a = (a^T)^* a = a^\dagger a. \quad (3.260)$$

The symbol \dagger , called dagger, denotes Hermitian conjugation, which means complex conjugation and transposing. We see that a transformation that preserves this inner product must fulfil the condition $U^\dagger U = 1$:

$$(Ua) \cdot (Ua) = a^\dagger U^\dagger U a = a^\dagger a \quad (3.261)$$

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 extra condition $\det(U) = 1$.

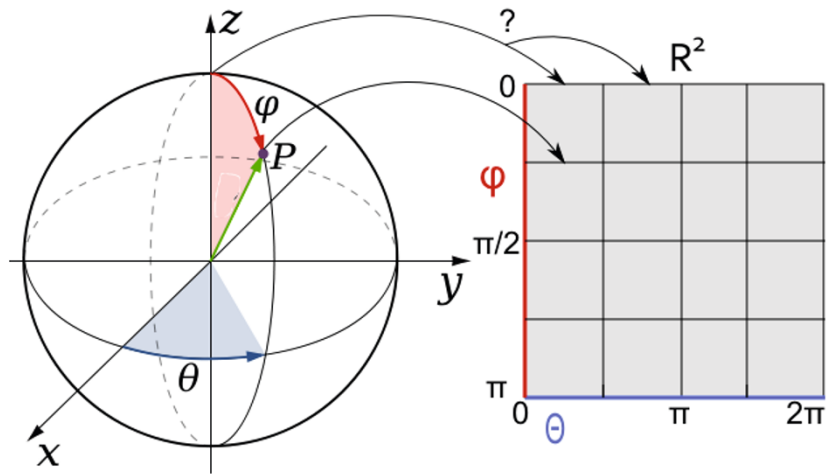
¹⁷³ Because for $z = a + ib$ we have $z^* = a - ib$ and therefore $z^* z = (a + ib)(a - ib) = a^2 + b^2$, which is real.

3.11 Appendix: Manifolds

A manifold M is a set of points if there exists 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 $(x_1(P), \dots, x_n(P))$ where the numbers $x_1(P), \dots, x_n(P)$ 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 sphere. The surface of the three-dimensional sphere 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 sphere 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. 3.8: 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 mustn'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 restricting 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 quantities¹. We will use this connection in the next chapter to understand how the quantities we measure in experiments can be described by the theory.

4.1 Lagrangian Formalism

The Lagrangian formalism is an incredibly powerful framework² that is used in most parts of fundamental physics. It is relatively simple, because the fundamental object, the **Lagrangian**, is a scalar³. The formalism is very useful if one wants to use symmetry considerations. If we demand the **action**, the integral over the Lagrangian, to be invariant under some symmetry transformation, we ensure that the

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 $\partial_t Q = 0 \rightarrow Q = \text{const.}$

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.

dynamics of the system in question respects 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."

⁴ "Recherche des loix du mouvement"
(1746)

- Pierre de Maupertius⁴

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. Mathematically we can write this, if we define the **action** of a given path $q(t)$ to be

$$S_{\text{light}}[\mathbf{q}(t)] = \int dt$$

and our task is to find the specific path $q(t)$ that minimizes the action⁵. To find the minimum⁶ of a given **function**, we take the derivative and set it to zero. Here we want to find the minimum of a **functional** $S[\mathbf{q}(t)]$, which means a function S of a function $\mathbf{q}(t)$ and we need a new mathematical idea, called **variational calculus**.

⁵ The action is here simply the integral over the time for a specific path, but in general the action will be a bit more complicated, as we will see in a moment.

⁶ In general, we want to find extremums, 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.

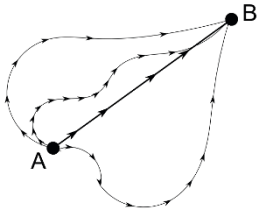


Fig. 4.1: Variations of a path with fixed starting- and end-point.

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 the neighbourhood of the minimum. 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 of a into our function $f(x)$:

$$f(a + \epsilon) = 3(a + \epsilon)^2 + (a + \epsilon) = 3(a^2 + 2a\epsilon + \epsilon^2) + a + \epsilon.$$

If a is a minimum, first order variations in ϵ must vanish, because 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 this to be zero

$$3 \cdot 2a\epsilon + \epsilon \stackrel{!}{=} 0 \rightarrow 6a + 1 \stackrel{!}{=} 0.$$

So we find the minimum

$$x_{\min} = a = \frac{-1}{6},$$

which is of course exactly the same result 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, but variational calculus is in addition able to find the extrema of functionals. We will see in a moment how this can be done 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, but we can make a more general ansatz

$$S[q(t)] = \int \mathcal{L} dt,$$

where \mathcal{L} is a, in general non-constant, parameter, called the Lagrangian. This parameter happens to be constant for light. In general, the Lagrangian is a function of the position $q(t)$ of the object on question and 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 section⁷. 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 Chap. 1.1 there are restrictions to our present theories we can't motivate from first principles. We only know that we must respect these restrictions in order to get a sensible theory.

One important restriction is that we are only allowed to use the lowest possible, non-trivial order 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. The lowest order of a given theory is determined by the condition that the Lagrangian must be Lorentz invariant⁸, 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 with such theories⁹. In addition, higher order derivatives in the Lagrangian lead to higher order derivatives in the equations of motion and therefore more initial conditions would be required.

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.

⁹ These problems are known as Ostrogradski instabilities. The energy in theories with higher order derivatives can be arbitrarily negative, which would mean that every state in such theory would always decay into lower energy states. There are no stable states in such theories.

¹⁰ Locality is a consequence of the basis postulates of special relativity, as shown in Sec. 2.4.

¹¹ We will discuss Lagrangians for particle theories, where we search for particle paths and Lagrangians for field theories, where we search for field configurations $\Phi(x)$. This is the topic of the next section.

¹² From another perspective, this means again that we only include the lowest possible, non-trivial terms. Terms with Φ^0 and Φ^1 are trivial, as we will see later and therefore we use once more only the lowest possible, non-trivial order, now in Φ .

¹³ Another possibility are spherical coordinates.

¹⁴ An especially beautiful feature of quantum field theory is how particles come into play. In Chap. 6 we will see that fields are able to create and destroy particles.

¹⁵ The usage of a different symbol \mathcal{L} here is intentional, because in field theories we will work most of the time with the Lagrangian density \mathcal{L} . The Lagrangian density is related to the Lagrangian via $\mathcal{L} = \int d^3x \mathcal{L}$.

It is sometimes claimed that the constraint to first order derivatives is a consequence of our demand to get a local¹⁰ theory, but this only rules out an **infinite** number of derivatives. A non-local interaction is of the form¹¹

$$\Phi(x-h)\Phi(x), \quad (4.1)$$

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

$$\Phi(x-h) = \sum_{k=0}^{\infty} \left(\left(\frac{\partial}{\partial x} \right)^k \Phi(x) \Big|_{x=h} \right) \frac{(x-h)^k}{k!} \quad (4.2)$$

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 the second power order. This means we only include the terms¹²

$$\Phi^0, \Phi^1, \Phi^2$$

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 coordinates¹³. For such theories the Lagrangian depends on the position \vec{q} , the velocity $\partial_t \vec{q}$ and the time t :

$$\mathcal{L} = \mathcal{L}(\vec{q}, \partial_t \vec{q}, t). \quad (4.3)$$

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 fields¹⁴. In such theories space and time form the stage the fields $\Phi(\vec{x}, t)$ act on. Using the restrictions we get¹⁵

$$\mathcal{L} = \mathcal{L}(\Phi(\vec{x}, t), \partial_\mu \Phi(\vec{x}, t), \vec{x}) \quad (4.4)$$

A famous example is the Lagrangian $\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)$, 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 will start with a 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** (e.g. maximum, minimum).

We use the notation

$$\dot{q}(t) \equiv \frac{dq(t)}{dt}.$$

Analogous to the example above, we fix $q(t) = a(t)$ and vary this function 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:

$$0 = \epsilon(t_1) = \epsilon(t_2), \quad (4.5)$$

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 \mathcal{L} as Taylor series¹⁶ and demand the first order terms to vanish

$$\int_{t_1}^{t_2} dt \left[\epsilon(t) \frac{\partial \mathcal{L}}{\partial q} + \left(\frac{d}{dt} \epsilon(t) \right) \frac{\partial \mathcal{L}}{\partial \dot{q}} \right] \stackrel{!}{=} 0. \quad (4.6)$$

We integrate the last term by parts¹⁷, which yields

$$\begin{aligned} & \int_{t_1}^{t_2} dt \left(\frac{d}{dt} \epsilon(t) \right) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \\ &= \epsilon(t) \frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} dt \epsilon(t) \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right) \end{aligned}$$

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

$$\int_{t_1}^{t_2} dt \epsilon(t) \left[\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial q} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right) \right] \stackrel{!}{=} 0$$

and we can see that this only vanishes for arbitrary variations $\epsilon(t)$ if the expression in the square bracket [] is identically zero. This yields¹⁸

$$\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial q} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}(q, \dot{q}, t)}{\partial \dot{q}} \right) = 0 \quad (4.7)$$

which is the famous **Euler-Lagrange equation**.

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.

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 $\frac{df}{dt} = \frac{\partial f}{\partial x} \frac{dx}{dt} + \frac{\partial f}{\partial y} \frac{dy}{dt} + \frac{\partial f}{\partial z} \frac{dz}{dt} + \frac{\partial f}{\partial t} \frac{dt}{dt}$. 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 $f(x(t), y(t)) = x^2 y + y^3$, 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, $\frac{df}{dt} = 2xy \frac{dx}{dt} + (x^2 + 3y^2) \frac{dy}{dt}$. In contrast for another function $g(x(t), y(t), t) = x^2 t + y$ we have $\frac{\partial g}{\partial t} = x^2$.

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 \mathcal{L} :

$$\mathcal{L} = \int d^3x \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \quad (4.8)$$

and we can write the action in terms of the Lagrangian density

$$S = \int dt \mathcal{L} = \int d^4x \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) \quad (4.9)$$

Following the same steps as above, we get the equations¹⁹ of motion for a field theory:

$$\frac{\partial \mathcal{L}}{\partial \Phi^i} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi^i)} \right) = 0 \quad (4.10)$$

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 quantities²⁰ to describe nature and from this theorem we learn how we can work with them in a theoretical context.

¹⁹ Plural because we get one equation for each field component, i.e. Φ_1, Φ_2, \dots

²⁰ We can see them as anchors in an otherwise extremely complicated world. While everything changes, the conserved quantities stay the same.

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 certainly 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 changes. The invariance of the Lagrangian under an infinitesimal transformation²¹ $q \rightarrow q' = q + \delta q$ can be expressed mathematically

$$\begin{aligned} \delta \mathcal{L} &= \mathcal{L} \left(q, \frac{dq}{dt}, t \right) - \mathcal{L} \left(q + \delta q, \frac{d(q + \delta q)}{dt}, t \right) \\ &= \mathcal{L} \left(q, \frac{dq}{dt}, t \right) - \mathcal{L} \left(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t \right) \stackrel{!}{=} 0. \end{aligned} \quad (4.11)$$

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:

$$\delta S = \int dt \mathcal{L} \left(q, \frac{dq}{dt}, t \right) - \int \mathcal{L} \left(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t \right) = \int dt \delta \mathcal{L} \stackrel{\substack{=0 \\ \text{if } \delta \mathcal{L}=0}}{=} 0 \quad (4.12)$$

How can the Lagrangian change, while the action stays the same? It turns out we can always add the total time derivative of an arbitrary function G to the Lagrangian

$$\mathcal{L} \rightarrow \mathcal{L} + \frac{dG}{dt}$$

without changing the action because²²

$$\delta S \rightarrow \delta S' = \delta S + \int_{t_1}^{t_2} dt \frac{d}{dt} \delta G = \delta S + \underbrace{\frac{\partial G}{\partial q} \delta q \Big|_{t_1}^{t_2}}_{=0 \text{ because } \delta q(t_1)=\delta q(t_2)=0}.$$

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 .

In the last step, we use that the variation δq vanishes at the initial and final moments of time (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 a less restrictive condition

$$\delta \mathcal{L} \stackrel{!}{=} \frac{dG}{dt}. \quad (4.13)$$

This means the Lagrangian can change without changing the action and therefore the equation of motion, 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

$$\delta \mathcal{L} = \mathcal{L} \left(q, \frac{dq}{dt}, t \right) - \mathcal{L} \left(q + \delta q, \frac{dq}{dt} + \frac{d\delta q}{dt}, t \right) \stackrel{!}{=} \frac{dG}{dt} \quad (4.14)$$

We expand the second term as Taylor series, keep only terms of first order in δq , because δq is infinitesimal and use the notation $\frac{dq}{dt} = \dot{q}$:

$$\begin{aligned} \rightarrow \delta \mathcal{L} &= \mathcal{L} - \mathcal{L} - \frac{\partial \mathcal{L}}{\partial q} \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \stackrel{!}{=} \frac{dG}{dt} \\ \rightarrow \delta \mathcal{L} &= - \frac{\partial \mathcal{L}}{\partial q} \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \stackrel{!}{=} \frac{dG}{dt} \end{aligned} \quad (4.15)$$

²³ Eq. 4.7: $\frac{\partial \mathcal{L}}{\partial q} = \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right)$

We can rewrite Eq. 4.15 by using the Euler-Lagrange equation²³. This yields

$$\rightarrow \delta \mathcal{L} = - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \delta q - \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} = \frac{dG}{dt}.$$

This can be rewritten, using the product rule²⁴

$$\begin{aligned} \rightarrow \delta \mathcal{L} &= -\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q \right) = \frac{dG}{dt} \\ \rightarrow \frac{d}{dt} \underbrace{\left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q + G \right)}_{\equiv J} &= 0. \end{aligned} \quad (4.16)$$

Therefore, we have found a quantity J that is conserved in time:

$$J = \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q + G, \quad (4.17)$$

because we have

$$\frac{d}{dt} J = 0 \rightarrow J = \text{const.}$$

To illustrate this, we will use one later result from Sec. 10.2: Newton's second law for a free particle with constant mass is²⁵

$$m\ddot{\vec{q}} = 0. \quad (4.18)$$

The Lagrangian that reproduces this famous equation of motion is²⁶

$$\mathcal{L} = \frac{1}{2} m \dot{\vec{q}}^2 \quad (4.19)$$

as you can check, by putting it into the Euler-Lagrange equation (Eq. 4.7).

Let us compute for this Lagrangian for different symmetries 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, with $G = 0$, which certainly fulfils the condition Eq. 4.13

$$J_{\text{trans}} = \frac{\partial \mathcal{L}}{\partial \dot{\vec{q}}} \vec{a} = m \dot{\vec{q}} \vec{a} = \vec{p} \vec{a}, \quad (4.20)$$

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 the **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. Instead of working with vectors like \vec{q} , we can use an index notation, which is quite useful here. We can simply rewrite all equations with $q \rightarrow q_i$. Let's now have a look at an infinitesimal rotation²⁷ $q_i \rightarrow q'_i = q_i + \epsilon_{ijk} q_j a_k$ and therefore $\delta q_i = \epsilon_{ijk} q_j a_k$. Our Lagrangian is

²⁴ If you're unsure about the product rule, have a look at appendix B.1.

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.

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} .

²⁷ Here we write the generator of rotations by using the Levi-Civita symbol. This was explained in the text below Eq. 3.63.

²⁸ This was derived in Eq. 4.16, again we work here with $G = 0$.

invariant under such transformations, because again, $\mathcal{L} = \frac{1}{2}m\dot{q}^2$ does not depend on q , and the corresponding conserved quantity is²⁸

$$\begin{aligned} J_{\text{rot}} &= \frac{\partial \mathcal{L}(q_i, \dot{q}_i, t)}{\partial \dot{q}_i} \delta q_i = \frac{\partial \mathcal{L}(q_i, \dot{q}_i, t)}{\partial \dot{q}_i} \epsilon_{ijk} q_j a_k \\ &= m \dot{q}_i \epsilon_{ijk} q_j a_k = p_i \epsilon_{ijk} q_j a_k \\ &\rightarrow J_{\text{rot}} = (\vec{p} \times \vec{q}) \cdot \vec{a} \equiv \vec{L} \cdot \vec{a} \end{aligned} \quad (4.21)$$

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.**

²⁹ 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.

Next let's have a look at invariance under time translations²⁹. An infinitesimal time displacement $t \rightarrow t' = t + \epsilon$ has the effect

$$\begin{aligned} \delta \mathcal{L} &= \mathcal{L}\left(q(t), \frac{dq(t)}{dt}, t\right) - \mathcal{L}\left(q(t+\epsilon), \frac{dq(t+\epsilon)}{dt}, t+\epsilon\right) \\ &= -\frac{\partial \mathcal{L}}{\partial q} \frac{\partial q}{\partial t} \epsilon - \frac{\partial \mathcal{L}}{\partial \dot{q}} \frac{\partial \dot{q}}{\partial t} \epsilon - \underbrace{\frac{\partial \mathcal{L}}{\partial t} \epsilon}_{\text{The left hand side is exactly the total derivative}} - \frac{d\mathcal{L}}{dt} \epsilon, \end{aligned} \quad (4.22)$$

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

$$\frac{d}{dt} \underbrace{\left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - L \right)}_{\equiv \mathcal{H}} = 0 \quad (4.23)$$

The conserved quantity \mathcal{H} is called the **Hamiltonian** and represents the total energy of the system. For our example Lagrangian, we have

$$\begin{aligned} \mathcal{H} &= \frac{\partial \mathcal{L}}{\partial \dot{q}} \dot{q} - \mathcal{L} = \underbrace{\left(\frac{\partial}{\partial \dot{q}} \frac{1}{2} m \dot{q}^2 \right) \dot{q}}_{=m\dot{q}} - \frac{1}{2} m \dot{q}^2 \\ &= m \dot{q}^2 - \frac{1}{2} m \dot{q}^2 = \frac{1}{2} m \dot{q}^2, \end{aligned} \quad (4.24)$$

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 Sec. 4.6. The resulting conserved quantity is

$$\tilde{J}_{\text{boost}} = pt - \underbrace{\frac{1}{2}mvt}_{\equiv \tilde{p}t} - mq = \tilde{p}t - mq. \quad (4.25)$$

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 invariance³⁰ \Rightarrow conservation of $\tilde{p}t - mq$
- Rotational invariance \Rightarrow conservation of angular momentum
- Translational invariance in time \Rightarrow conservation of energy

³⁰ Another name for a boost is translation in momentum space, because the transformation $q \rightarrow q + vt$, changes the momentum to $m\dot{q} \rightarrow m(\dot{q} + v)$.

Noether's theorem shows us why those notions³¹ 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.

³¹ Except for the conserved quantity following from boost invariance.

It is instructive to have a look at how those notions occur in field theories. For field theories we have two kinds of symmetries. On the one hand, our Lagrangian can be invariant under transformations of spacetime, 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 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 Sec. 3.7.11 the field representation, which we will be using now. The (infinite-dimensional) differential

³² 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 $A'_\mu = \begin{pmatrix} A'_0 \\ A'_1 \\ A'_2 \\ A'_3 \end{pmatrix} = \begin{pmatrix} A_0 \\ -A_2 \\ A_1 \\ A_3 \end{pmatrix} \cdot A'_\mu$ and

A_μ describe the same field in coordinate systems that are rotated by 90° around the z-axis relative to each other.

³³ 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.

³⁴ See Eq. 4.10: $\frac{\partial \mathcal{L}}{\partial \Phi} = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right)$

operator representation changes $x \rightarrow x'$. This means by using this representation we can compute the field components at a different point in spacetime or in a rotated frame. The finite-dimensional transformation of the Lorentz group changes $\Psi \rightarrow \Psi'$, i.e. mixes the field components³².

A complete transformation, for a field that depends on spacetime, needs to consider both parts. We will look at these parts separately, starting with the change $x \rightarrow x'$. For rotations the conserved quantity that follows is not really conserved, because we neglected the second part of the transformation. Only the sum of the two conserved quantities that follow from $x \rightarrow x'$ and $\Psi \rightarrow \Psi'$ is conserved.

To make this more concrete consider a general Lagrangian density $\mathcal{L}((\Phi(x_\mu), \partial_\mu \Phi(x_\mu), x_\mu)$. Symmetry means we have

$$\mathcal{L}((\Phi(x_\mu), \partial_\mu \Phi(x_\mu), x_\mu) = \mathcal{L}((\Phi'(x'_\mu), \partial_\mu \Phi'(x'_\mu), x'_\mu). \quad (4.26)$$

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 by³³

$$df(g(x), h(x), \dots) = \frac{\partial f}{\partial g} \delta g + \frac{\partial f}{\partial h} \delta h + \dots + \frac{\partial f}{\partial x} \delta x. \quad (4.27)$$

Applying this to the Lagrangian yields

$$\delta \mathcal{L} = \frac{\partial \mathcal{L}}{\partial \Phi} \delta \Phi + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \delta (\partial_\mu \Phi) + \frac{\partial \mathcal{L}}{\partial x_\mu} \delta x_\mu, \quad (4.28)$$

which we can rewrite using the Euler-Lagrange equations³⁴

$$\begin{aligned} \delta \mathcal{L} &= \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) \delta \Phi + \frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \underbrace{\delta (\partial_\mu \Phi)}_{=\partial_\mu \delta \Phi} + \frac{\partial \mathcal{L}}{\partial x_\mu} \delta x_\mu \\ &\underbrace{=}_{\text{Product rule}} \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \delta \Phi \right) + \frac{\partial \mathcal{L}}{\partial x_\mu} \delta x_\mu \end{aligned} \quad (4.29)$$

The variation $\delta \Phi$ has now two parts

$$\delta \Phi = \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x) - \frac{\partial \Phi(x)}{\partial x_\mu} \delta x_\mu, \quad (4.30)$$

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 $S_i = \frac{1}{2} \epsilon_{ijk} S_{jk}$ and to the generators of boosts by $K_i = S_{0i}$, analogous to the definition of $M_{\mu\nu}$ in Eq. 3.165. 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 lead to a mixing of 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, i.e.

$$x_\mu \rightarrow x'_\mu = x_\mu + \delta x_\mu = x_\mu + a_\mu. \quad (4.31)$$

From Eq. 4.30 we have for translations, with $\epsilon_{\mu\nu} = 0$, because field components do not mix under translations

$$\delta\Phi = -\frac{\partial\Phi}{\partial x_\mu}\delta x_\mu$$

and thus from Eq. 4.29, if we want to investigate the consequences of invariance ($\delta\mathcal{L} = 0$)

$$-\partial_\nu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi}{\partial x_\mu} \delta x_\mu \right) + \frac{\partial\mathcal{L}}{\partial x_\mu} \delta x_\mu = 0 \quad (4.32)$$

$$\rightarrow -\partial_\nu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi}{\partial x_\mu} - \delta_\mu^\nu \mathcal{L} \right) \delta x^\mu = 0 \quad (4.33)$$

From Eq. 4.31 we have $\delta x_\mu = a_\mu$, which we put into Eq. 4.33. This yields

$$-\partial_\nu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi}{\partial x_\mu} - \delta_\mu^\nu \mathcal{L} \right) a^\mu = 0, \quad (4.34)$$

and we define the energy-momentum tensor

$$T_\mu^\nu := \frac{\partial\mathcal{L}}{\partial(\partial_\nu\Phi)} \frac{\partial\Phi}{\partial x_\mu} - \delta_\mu^\nu \mathcal{L}. \quad (4.35)$$

Equation 4.34 tells us that T_μ^ν fulfils a continuity equation, because a^μ is arbitrary

$$\partial_\nu T_\mu^\nu = 0 \quad (4.36)$$

$$\rightarrow \partial_\nu T_\mu^\nu = \partial_0 T_\mu^0 - \partial_i T_\mu^i = 0 \quad (4.37)$$

for each component μ . This tells us directly that we have conserved quantities, because for example for $\mu = 0$ we get³⁵

$$\begin{aligned} \partial_0 T_0^0 + \partial_i T_0^i &= 0 \rightarrow \partial_0 T_0^0 = -\partial_i T_0^i \\ \partial_t T_0^0 &= -\nabla \vec{T} \cdot \underbrace{\int_V d^3x \partial_t T_0^0}_{\text{Integrating over some infinite volume } V} = -\int_V d^3x \nabla \vec{T} \\ &\xrightarrow{\text{Divergence theorem; } \delta V \text{ denotes the surface of the volume } V} -\int_{\delta V} d^2x \vec{T} \cdot \underbrace{\vec{n}}_{\text{Because fields vanish at infinity}} = 0 \end{aligned} \quad (4.38)$$

³⁵ Using $\partial_0 = \partial_t$ and $\partial^i T_i^0 = \nabla \vec{T}$ and the famous divergence theorem $\int_V d^3x \nabla A = \int_{\delta V} d^2x A$, which enables us to rewrite the integral over some volume V , as an integral over the corresponding surface δV . A very illuminating 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 available Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics: Volume 2*. Addison-Wesley, 1st edition, 2 1977. ISBN 9780201021172

$$\rightarrow \partial_t \int_V d^3x T_0^0 = 0 \quad (4.39)$$

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 $r = \infty$. We discovered in Sec. 2.3 that we have an upper speed limit for everything in physics. Therefore the field configuration infinitely far away, can't have any influence on physics at a finite x and we say that fields vanish at infinity.

We conclude: The invariance under translations in spacetime leads us to 4 conserved quantities³⁶, one for each component μ . Equation 4.39 tells us these are

$$E = \int d^3x T_0^0 \quad (4.40)$$

$$P_i = \int d^3x T_i^0, \quad (4.41)$$

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 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, which means $S_{\mu\nu} = 0$, because we use for scalars the 1-dimensional representation of the generators of the Lorentz group that were derived in Sec. 3.7.4. Therefore what we derive in this first section is the complete conserved quantity for a scalar field.

The second part of the transformation is given by

$$x_\mu \rightarrow x'_\mu = x_\mu + \delta x_\mu = x_\mu + M_\mu^\sigma x_\sigma$$

where M_μ^σ is the generator of rotations and boost in its infinite-dimensional representation, which means we use differential operators, as defined in Eq. 3.240.

³⁶ 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 component.

Putting $\delta x_\mu = M_\mu^\sigma x_\sigma$ into Eq. 4.33 yields

$$\partial_\nu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\nu \Phi)} \frac{\partial (\Phi)}{\partial x_\mu} - \delta_\mu^\nu \mathcal{L} \right) M^{\mu\sigma} x_\sigma = 0 \quad (4.42)$$

$$\xrightarrow{\text{Eq. 4.35}} \partial_\nu T_\mu^\nu M^{\mu\sigma} x_\sigma = 0$$

We can write this a little differently to bring it into the conventional form

$$\begin{aligned} &= \frac{1}{2} (\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma + \partial^\nu T_\nu^\mu \underbrace{M_{\mu\sigma}}_{=-M_{\sigma\mu}} x^\sigma) = \frac{1}{2} (\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma - \partial^\nu T_\nu^\mu M_{\sigma\mu} x^\sigma) \\ &\stackrel{\text{Renaming dummy indices}}{=} \frac{1}{2} (\partial^\nu T_\nu^\mu M_{\mu\sigma} x^\sigma - \partial^\nu T_\nu^\sigma M_{\mu\sigma} x^\mu) = \frac{1}{2} (\partial^\nu T_\nu^\mu x^\sigma - \partial^\nu T_\nu^\sigma x^\mu) M_{\mu\sigma} \\ &= \frac{1}{2} \partial_\nu (T^{\mu\nu} x^\sigma - T^{\sigma\nu} x^\mu) M_{\mu\sigma} = 0 \end{aligned}$$

$$\rightarrow \text{with } (J^\nu)^{\sigma\mu} \equiv T^{\mu\nu} x^\sigma - T^{\sigma\nu} x^\mu \rightarrow \partial_\nu (J^\nu)^{\sigma\mu} = 0. \quad (4.43)$$

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. 3.165, 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

$$Q^{\mu\nu} = \int d^3x (T^{\mu 0} x^\sigma - T^{\sigma 0} x^\mu) \quad (4.44)$$

For rotational³⁷ invariance, we therefore have the conserved quantities

$$L_{\text{orbit}}^i = \frac{1}{2} \epsilon^{ijk} Q^{jk} = \frac{1}{2} \epsilon^{ijk} \int d^3x (T^{k0} x^j - T^{j0} x^k), \quad (4.45)$$

which we call the orbital angular momentum of our field³⁸.

Equivalently, we have the conserved quantity for boosts³⁹

$$Q^{0i} = \int d^3x (T^{i0} x^0 - T^{00} x^i), \quad (4.46)$$

which is discussed a bit more in the appendix Sec. 4.7.

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

$$\delta\Phi = \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x), \quad (4.47)$$

³⁷ Recall the connection between the generator of rotations J_i and $M_{\mu\nu}$ from Eq. 3.165: $J_i = \frac{1}{2} \epsilon_{ijk} M_{jk}$, which means we now restrict to the spatial components $i = j = \{1, 2, 3\}$ here and write i, j instead of $\mu = \nu = \{0, 1, 2, 3\}$.

³⁸ The subscript orbit will become clear in a moment, because actually this is just one part of the conserved quantity and we will have a look at the second part in a moment.

³⁹ Recall the connection between the boost generators K_i and $M_{\mu\nu}$: $K_i = M_{0i}$.

⁴⁰ 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.

where $S^{\mu\nu}$ is the appropriate **finite-dimensional** representation of the transformation in question⁴⁰. This gives us in Eq. 4.29 the extra term

$$\partial_\rho \left(\frac{\partial \mathcal{L}}{\partial(\partial_\rho \Phi)} \epsilon_{\mu\nu} S^{\mu\nu} \Phi(x) \right) \quad (4.48)$$

which leads to an additional term in Eq. 4.45. The complete conserved quantity is

$$L^i = \frac{1}{2} \epsilon^{ijk} Q^{ij} = \frac{1}{2} \epsilon^{ijk} \int d^3x \left(\frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} S^{jk} \Phi(x) + (T^{k0} x^j - T^{j0} x^k) \right) \quad (4.49)$$

and therefore we write

$$L^i = L_{\text{spin}}^i + L_{\text{orbit}}^i. \quad (4.50)$$

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 internal⁴¹ angular momentum.

⁴¹ 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.

Recall that we label the representations of the Poincare group by something we call spin, too. Now this notion appears again. We will learn in the next chapter and in great detail in Sec. 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 Poincare group.

4.5.5 Noether's Theorem for Field Theories - Internal Symmetries

We now take a look at internal symmetries⁴². The invariance of a Lagrangian ($\delta \mathcal{L} = 0$) under some infinitesimal transformation of the field **itself**

$$\Phi_i \rightarrow \Phi'_i = \Phi_i + \delta \Phi_i \quad (4.51)$$

can be written mathematically

$$\begin{aligned} \delta \mathcal{L} &= \mathcal{L}(\Phi^i, \partial_\mu \Phi^i) - \mathcal{L}(\Phi_i + \delta \Phi_i, \partial_\mu (\Phi_i + \delta \Phi_i)) \\ &= - \frac{\partial \mathcal{L}(\Phi^i, \partial_\mu \Phi^i)}{\partial \Phi_i} \delta \Phi_i - \frac{\partial \mathcal{L}(\Phi^i, \partial_\mu \Phi^i)}{\partial (\partial_\mu \Phi_i)} \partial_\mu \delta \Phi_i \stackrel{!}{=} 0, \end{aligned} \quad (4.52)$$

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

⁴² 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.

infinitesimal. From the Euler-Lagrange equation (Eq. 4.10) for fields we know

$$\frac{\partial \mathcal{L}}{\partial \Phi_i} = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi_i)} \right).$$

Putting this into Eq. 4.52, we get

$$\delta \mathcal{L} = -\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi_i)} \right) \delta \Phi_i - \frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi_i)} \partial_\mu \delta \Phi_i = 0$$

which is, using the product rule, the same as

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi_i)} \delta \Phi_i \right) = 0 \quad (4.53)$$

This means that, 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

$$J^\mu \equiv \frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi_i)} \delta \Phi_i, \quad (4.54)$$

which fulfils the continuity equation

$$\text{Eq. 4.53} \rightarrow \partial_\mu J^\mu = 0. \quad (4.55)$$

Furthermore, following the same steps as in Eq. 4.39, we are able to derive a quantity that is conserved in time

$$\partial_t \int d^3x J^0 = 0. \quad (4.56)$$

Let's have a look at one very important example: Invariance under displacements of the field itself⁴³, which means $\delta \Phi = -i\epsilon$ in Eq. 4.51:

$$\Phi \rightarrow \Phi' = \Phi - i\epsilon. \quad (4.57)$$

or for more than one field component

$$\Phi_i \rightarrow \Phi'_i = \Phi_i - i\epsilon_i. \quad (4.58)$$

The generator of this transformation is $-i\frac{\partial}{\partial \Phi}$, because

$$\Phi' = e^{-i\epsilon \frac{\partial}{\partial \Phi}} \Phi = \left(1 - i\epsilon \frac{\partial}{\partial \Phi} + \dots\right) \Phi \approx \Phi - i\epsilon \quad (4.59)$$

The corresponding conserved quantity is called **conjugate momentum** Π . It's conventional to introduce the conjugate momentum density⁴⁴ $\pi \equiv J_0$ and with Eq. 4.53, Eq. 4.56 and using that $\delta \Phi = -i\epsilon$ doesn't need to be included in order to get a conserved quantity. we get⁴⁵

$$\partial_t \Pi = 0 \rightarrow \partial_t \Pi = \partial_t \int d^3x J_0 = \partial_t \int d^3x \pi = 0$$

⁴³ 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 complex 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 $\epsilon' := i\epsilon$.

⁴⁴ This rather abstract quantity is one of the most important notions in quantum field theory, as we will see in Chap. 6.

⁴⁵ The $\delta \Phi$ in Eq. 4.54 becomes a constant ϵ and because the quantity is conserved for arbitrary ϵ we define our conserved quantity without it.

$$\text{with } J_0 = \pi = \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} \quad (4.60)$$

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

$$p^i = T^{0i} = \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_i},$$

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 developing interaction theories. Now let's summarize what we learned about symmetries and conserved quantities in field theories:

- Translational invariance in space \Rightarrow conservation of physical momentum $P^i = \int d^3x T^{0i} = \int d^3x \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_i}$
- Boost invariance⁴⁶ \Rightarrow constant velocity of the center of energy.
- Rotational invariance \Rightarrow conservation of the total angular momentum $L^i = \frac{1}{2} \epsilon^{ijk} \int d^3x \left(\frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} S^{jk} \Phi(x) + (T^{k0} x^j - T^{j0} x^k) \right) = L_{\text{spin}}^i + L_{\text{orbit}}^i$ consisting of spin and orbital angular momentum.
- Translational invariance in time \Rightarrow conservation of energy $E = \int d^3x T^{00} = \int d^3x \left(\frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} \frac{\partial \Phi}{\partial x_0} - \mathcal{L} \right)$.
- Displacement invariance of the field itself \Rightarrow conservation of the conjugate momentum $\Pi = \int d^3x \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} = \int d^3x \pi$.

⁴⁶ Another name for a boost is a translation in momentum space.

Further Reading Tips

⁴⁷ Cornelius Lanczos. *The Variational Principles of Mechanics*. Dover Publications, 4th edition, 3 1986. ISBN 9780486650678

⁴⁸ 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**⁴⁷ is a brilliant book about the usage of the Lagrangian formalism in classical mechanics.
- **Herbert Goldstein - Classical Mechanics**⁴⁸ 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 $\mathcal{L} = \frac{1}{2} m \dot{q}^2$ from boost invariance. A boost transformation is $q \rightarrow q' = q + \delta q = q + vt$, 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 $q \rightarrow q' = q + vt \Rightarrow \underbrace{\dot{q}'}_{v=\text{const!}} = \dot{q} + v$ and therefore our Lagrangian

$$\mathcal{L} = \frac{1}{2}m\dot{q}^2$$

is changed by

$$\begin{aligned}\delta\mathcal{L} &= \mathcal{L}(q, \dot{q}, t) - \mathcal{L}(q', \dot{q}', t) = \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m\dot{q}'^2 \\ &= \frac{1}{2}m\dot{q}^2 - \frac{1}{2}m(\dot{q} + v)^2 = -m\dot{q}v - \frac{1}{2}mv^2,\end{aligned}$$

which is the total derivative of the function⁴⁹

$$-G \equiv -m\dot{q}v - \frac{1}{2}mv^2t.$$

We conclude that our equations of motion⁵⁰ stay unchanged⁵¹ and we can find a conserved quantity.

For boosts, the conserved quantity is (Eq. 4.16)

$$J_{\text{boost}} = \underbrace{\frac{\partial\mathcal{L}}{\partial\dot{q}}}_{\equiv p} \underbrace{\delta q}_{=vt} - G = pvt - mvq - \frac{1}{2}mv^2t,$$

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

$$\tilde{J}_{\text{boost}} = pt - \underbrace{\frac{1}{2}mv^2t - mvq}_{\equiv \tilde{p}t} = \tilde{p}t - mq \quad (4.61)$$

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.

4.7 Appendix: Conserved Quantity from Boost Invariance for Field Theories

For invariance under boosts in a field theory we have the conserved quantities

$$Q^{0i} = \int d^3x (T^{i0}x^0 - T^{00}x^i), \quad (4.62)$$

⁴⁹ $\frac{dG}{dt} = \frac{d}{dt}(-m\dot{q}v - \frac{1}{2}mv^2t) = -m\dot{q}v - \frac{1}{2}mv^2$, because v and m are constant.

⁵⁰ Another way to see this is by looking directly at the equation of motion: $m\ddot{q} = 0$. We have the transformation $q \rightarrow q' = q + vt$, with $v = \text{const}$ and therefore $\dot{q} \rightarrow \dot{q}' = \dot{q} + v$ and $\ddot{q} \rightarrow \ddot{q}' = \ddot{q}$.

⁵¹ The Lagrangian is changed, i.e. $\delta\mathcal{L} \neq 0$, but the action is only changed up to a total derivative and we therefore get the same equation of motion.

which we derived in Eq. 4.46 and using $K_i = M_{0i}$.

How can we interpret this quantity? Q^{0i} is conserved, which means

$$\begin{aligned} 0 = \frac{\partial Q^{0i}}{\partial t} &= \underbrace{\int d^3x \frac{\partial T^{i0}}{\partial t} x^0}_{=x^0 \frac{\partial}{\partial t} \int d^3x T^{i0}} + \underbrace{\int d^3x T^{i0} \frac{\partial x_0}{\partial t}}_{=1 \text{ because } x_0=t} - \frac{\partial}{\partial t} \int d^3x \frac{\partial T^{00}}{\partial t} x^i \\ &= t \frac{\partial P^i}{\partial t} + P^i - \frac{\partial}{\partial t} \int d^3x T^{00} x^i. \end{aligned}$$

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

$$\frac{\partial}{\partial t} \int d^3x T^{00} x^i = P^i = \text{const.} \quad (4.63)$$

⁵² 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 us⁵² 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. 4.46 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."

Paul A. M. Dirac
in The evolution of the
physicist's picture of nature.
Scientific American, vol. 208
Issue 5, pp. 45-53.
Publication Date: 05/1963

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. We will utilize this connection in this chapter.

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 momentum, energy or angular momentum. Noether's theorem hints us towards and incredibly important idea: We identify the quantities we use to describe nature with the corresponding generators:

physical quantity \Rightarrow generator of the corresponding symmetry. (5.1)

As we will see, this identification will naturally guide us towards quantum theory. Let's make this concrete by considering a particle theory.

5.1 The Operators of Quantum Mechanics

It is conventional to use a hat: \hat{O} to denote an operator.

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

momentum $\hat{p}_i \rightarrow$ generator of spatial-translations $-i\partial_i$

Analogous, the invariance under the action of the generator of time-translations leads us to the conservation of energy. Conse-

¹ Alternatively, we can see this by looking at the conserved quantity corresponding to invariance under boosts. Take note that we derived in Sec. 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 $M_i = (tp_i - x_i E)$. The generator of boosts is (see Eq. 3.240 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 $M_i = (tp_i - x_i E) \leftrightarrow K_i = i(t\partial_i - x_i \partial_0)$. 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 Sec. 8.3 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: $\vec{F} = m\vec{\ddot{x}}$, could be written as $\vec{F}\vec{C} = m\vec{\ddot{x}}\vec{C}$, 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.

quently

energy $\hat{E} \rightarrow$ 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 have¹

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 important² equations of quantum mechanics. As explained above, we will assume that our operators act on something abstract Ψ . Then we have

$$[\hat{p}_i, \hat{x}_j]\Psi = (\hat{p}_i \hat{x}_j - \hat{x}_j \hat{p}_i)\Psi = (\partial_i \hat{x}_j - \hat{x}_j \partial_i)\Psi$$

$$\underbrace{= (\partial_i \hat{x}_j)\Psi + \hat{x}_j (\partial_i \Psi) - \hat{x}_j \partial_i \Psi}_{\text{product rule}} = \underbrace{i\delta_{ij}\Psi}_{\text{because } \partial_i \hat{x}_j = \frac{\partial x_j}{\partial x_i}}. \quad (5.2)$$

This equation holds for arbitrary Ψ , because we made no assumptions about Ψ and therefore we can write the equation without it³:

$$[\hat{p}_i, \hat{x}_j] = i\delta_{ij}. \quad (5.3)$$

5.1.1 Spin and Angular Momentum

In the last chapter (Sec. 4.5.4) we saw that the conserved quantity that resulted from rotational invariance has two parts, at least for a non-scalar theory. The second part could be identified with the orbital angular momentum and we therefore make the identification with the **infinite** dimensional representation of the generator

orb. angular mom. $\hat{L}_i \rightarrow$ 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 generators⁴

spin $\hat{S}_i \rightarrow$ generators of rotations (fin. dim. rep.) S_i .

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 we derived in Sec. 3.7.5:

$$\hat{S}_i = i\frac{\sigma_i}{2} \quad (5.4)$$

with the Pauli matrices σ_i . We will return to this very interesting and very strange type of angular momentum in Sec. 8.5.5, after we learned how to work with the operators we derive in this chapter. Always keep in mind that only the **sum** of both parts is conserved.

5.2 The Operators of Quantum Field Theory

The central object in a field theory is, of course, the field as 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 $\Pi = \int dx^3 \pi(x) \neq \Pi(x)$.

⁵ Here $x = x_0, x_1, x_2, x_3$ includes time $x_0 = t$.

We discovered in the last chapter for invariance under displacements of the field itself $\Phi \rightarrow \Phi - i\epsilon$ 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. 4.59)

conj. mom. density $\pi(x) \rightarrow$ gen. of displ. of the field itself : $-i\frac{\partial}{\partial\Phi(x)}$

We'll 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. Here we will work again with the abstract Ψ , that we will specify in a later chapter. We are again able to derive an incredibly important equation, this time of quantum field theory⁶

$$[\Phi(x), \pi(y)]\Psi = \left[\Phi(x), -i\frac{\partial}{\partial\Phi(y)} \right] \Psi$$

$$\underbrace{=}_{\text{product rule}} -i\Phi(x)\frac{\partial\Psi}{\partial\Phi(y)} + i\Phi(x)\left(\frac{\partial\Psi}{\partial\Phi(y)}\right) + i\left(\frac{\partial\Phi(x)}{\partial\Phi(y)}\right)\Psi = i\delta(x-y)\Psi \quad (5.5)$$

Again, the equations hold for arbitrary Ψ and we can therefore write

$$[\Phi(x), \pi(y)] = i\delta(x-y) \quad (5.6)$$

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.

Analogous we have for more than one field component

$$[\Phi_i(x), \pi_j(y)] = i\delta(x-y)\delta_{ij}. \quad (5.7)$$

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**, which means non-interacting, fields¹ 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 equation from the vector $(\frac{1}{2},\frac{1}{2})$ representation of the Lorentz Group which in the massless limit reduces to 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 we have. We want these equations to look the same in all inertial frames, because if this was not the case we 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 is **Lorentz covariance**. 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 $A_\mu \rightarrow A'_\mu$, 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

$$\Lambda_\mu^\nu A_\nu + 7\Lambda_\mu^\nu B_\nu + \underbrace{C_\nu A^\nu \Lambda_\mu^\nu D_\nu}_{= \text{a Lorentz scalar, transforming according to the } (0,0) \text{ representation}} = A'_\mu + 7B'_\mu + C_\nu A^\nu D'_\mu = 0.$$

= a Lorentz scalar, transforming according to the (0,0) representation

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 action² 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.

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.

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 restrictions³ is

$$\mathcal{L} = A\Phi^0 + B\Phi + C\Phi^2 + D\partial_\mu\Phi + E\partial_\mu\Phi\partial^\mu\Phi + F\Phi\partial_\mu\Phi \quad (6.1)$$

Firstly, take note that we are considering the Lagrangian density \mathcal{L} and not L itself and we get our physical theory from the action

$$S = \int dx \mathcal{L}, \quad (6.2)$$

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

³ Discussed in Sec. 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.

equivalent to the term $\partial_\mu \Phi \partial^\mu \Phi$, as we can see if we integrate by parts⁴.

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 motion⁵. In addition, we can 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 motion⁶. What remains is

$$\mathcal{L} = C\Phi^2 + E\partial_\mu \Phi \partial^\mu \Phi \quad (6.3)$$

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.

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 physics⁷. Nevertheless, it is conventional to include an overall factor $\frac{1}{2}$ into the Lagrangian and call the remaining constant⁸ $-m^2$. Therefore, we are finally left with

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \quad (6.4)$$

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

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) &= 0 \\ \rightarrow \frac{\partial}{\partial \Phi} \left(\frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \right) - \partial_\mu \left(\frac{\partial}{\partial (\partial_\mu \Phi)} \left(\frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \right) \right) \\ \rightarrow (\partial_\mu \partial^\mu + m^2)\Phi &= 0 \end{aligned} \quad (6.5)$$

This is the famous **Klein-Gordon equation**, which is the correct equation to describe free spin 0 fields/particles.

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.

⁴ The boundary terms, as usual, vanish, because fields vanish at infinity. Recall that this follows, because we have an upper speed limit in physics (Sec. 2.3). Therefore fields infinitely far away can't have any influence at finite x .

⁵ See Eq. 4.10: $\frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) = 0$ and therefore $\mathcal{L} \rightarrow \mathcal{L} + A$ with some constant A does not change anything: $\frac{\partial (\mathcal{L} + A)}{\partial \Phi} - \partial_\mu \left(\frac{\partial (\mathcal{L} + A)}{\partial (\partial_\mu \Phi)} \right) = \frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right)$

⁶ $\mathcal{L} \rightarrow \mathcal{L} + B\Phi$ yields $\frac{\partial (\mathcal{L} + B\Phi)}{\partial \Phi} - \partial_\mu \left(\frac{\partial (\mathcal{L} + B\Phi)}{\partial (\partial_\mu \Phi)} \right) = \frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) + B = 0$, which is just an additional constant in the equation of motion.

⁷ $\mathcal{L} \rightarrow C\mathcal{L}$ yields $\frac{\partial (C\mathcal{L})}{\partial \Phi} - \partial_\mu \left(\frac{\partial (C\mathcal{L})}{\partial (\partial_\mu \Phi)} \right) = 0$
 \uparrow
 $\frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) = 0$

⁸ 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.

⁹ 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.

This will not be the case for spin $\frac{1}{2}$ fields Ψ and this curious fact will have very interesting consequences⁹. Nevertheless, nothing prevents us from investigating the equally Lorentz invariant Lagrangian

$$L = \partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi$$

as many textbooks do. This is simply the same as investigating the interaction between **two** scalar fields of equal mass:

$$L = \frac{1}{2} \partial_\mu \phi_1 \partial^\mu \phi_1 - \frac{1}{2} m^2 \phi_1^2 + \frac{1}{2} \partial_\mu \phi_2 \partial^\mu \phi_2 - \frac{1}{2} m^2 \phi_2^2$$

because we have

$$\phi \equiv \frac{1}{\sqrt{2}} (\phi_1 + i\phi_2).$$

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, but this Lagrangian can be used to describe composite systems like mesons. Therefore, 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 Zee¹⁰

¹⁰ Anthony Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 1st edition, 3 2003. ISBN 9780691010199

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 time¹¹. The objects transforming under this representation are called Dirac spinor, which combine right-chiral and left-chiral spinors into one object. As already discussed in Sec. 3.7.9 a Dirac spinor is defined by

$$\Psi \equiv \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_a \\ \xi^{\dot{a}} \end{pmatrix}. \quad (6.6)$$

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 spinors.

¹¹ We showed in Sec. 3.7.9 that a parity transformation transforms the $(\frac{1}{2}, 0)$ representation, into the $(0, \frac{1}{2})$ representation.

We will use the Van-der-Waerden notation introduced in Sec. 3.7.7. Two possibilities are¹²

$$I_1 := \chi_a^T \zeta^{\dot{a}} = (\chi_a^*)^T \zeta^{\dot{a}} = (\chi_a)^\dagger \zeta^{\dot{a}} = (\chi_L)^\dagger \zeta_R \quad (6.7)$$

and

$$I_2 := (\zeta^a)^T \chi_a = ((\zeta^a)^*)^T \chi_a = (\zeta_R)^\dagger \chi_L, \quad (6.8)$$

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, as was shown explicitly in Sec. 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 Sec. 3.7.8 how we can construct four-vectors from spinors

$$v_{ab} = v_\nu \sigma_{ab}^\nu$$

where v_ν transforms like of a four-vector. The derivation operator is therefore in the spinor formalism

$$\partial_{ab} = \partial_\nu \sigma_{ab}^\nu \quad (6.9)$$

It is conventional to define $\bar{\sigma}^0 = I_{2 \times 2}$, $\bar{\sigma}^i = -\sigma^i$ and then we can construct the Lorentz invariant terms

$$I_3 := (\chi_a)^T \partial_\mu (\sigma^\mu)^{\dot{a}b} \chi_b = (\chi_L)^\dagger \partial_\mu \bar{\sigma}^\mu \chi_L \quad (6.10)$$

and

$$I_4 := (\zeta^a)^T \partial_\mu (\sigma^\mu)_{ab} \zeta^{\dot{b}} = (\zeta_R)^\dagger \partial_\mu \sigma^\mu \zeta_R \quad (6.11)$$

Instead of $(\sigma^\mu)_{ab}$, we need here $(\sigma^\mu)^{\dot{a}b}$. 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:

$$\begin{aligned} (\sigma^\mu)^{\dot{a}b} &= ((\sigma^\mu)^T)^{\dot{b}a} = \epsilon^{\dot{b}\dot{c}} ((\sigma^\mu)^T)_{\dot{c}d} (\epsilon^{ad})^T \\ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} (\sigma^\mu)^T \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}^T \\ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} (\sigma^\mu)^T \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \bar{\sigma}_\mu \end{aligned} \quad (6.12)$$

for example, for σ_3

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{=\sigma_3^T} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}}_{=\bar{\sigma}_3 = -\sigma_3}.$$

¹² 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. ζ_R and χ_L both have two components, and therefore we have four degrees of freedom in a term like $(\zeta_R)^\dagger \chi_L$. 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. Just one more thing: A Majorana spinor is a "real" Dirac spinor. I put real into quotation marks, because usually real means $\Psi^* \stackrel{!}{=} \Psi$. For spinors this condition is not Lorentz invariant (because the Lorentz transformations are complex in this representation). If we impose the standard condition ($\Psi^* \stackrel{!}{=} \Psi$) 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: $\begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \Psi^* \stackrel{!}{=} \Psi$, which is commonly interpreted as charge conjugation. This interpretation will be explained in Sec. 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: $\Psi_M \equiv \begin{pmatrix} \chi_L \\ \epsilon \chi_L^* \end{pmatrix}$ or $\Psi_M \equiv \begin{pmatrix} -\epsilon \zeta_R^* \\ \zeta_R \end{pmatrix}$

¹³ Remember that we get the equations of motion from the action, which is the integral over the Lagrangian.

¹⁴ 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 give us anything new.

¹⁵ 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 $\eta_{\mu\nu}\sigma^\nu = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} \sigma^0 \\ \sigma^1 \\ \sigma^2 \\ \sigma^3 \end{pmatrix} = \begin{pmatrix} \sigma^0 \\ -\sigma^1 \\ -\sigma^2 \\ -\sigma^3 \end{pmatrix} = \bar{\sigma}_\mu$

¹⁶ Remember that σ_0 is just the unit matrix and we have $\bar{\sigma}_0 = \sigma_0$.

¹⁷ Take note that $\sigma^\mu\partial_\mu = \partial_\mu\sigma^\mu$, because σ^μ are constant matrices.

¹⁸ Recall, these were maximum order two in Ψ and the lowest possible, non-trivial order in ∂_μ , which is here order 1.

¹⁹ Therefore the left-chiral and right-chiral spinors inside each Dirac spinor, are complex, too.

Don't let yourself get confused why ∂_μ acts only on one spinor, because we are going to use these invariants in Lagrangians, which we always evaluate inside of integrals¹³ and therefore, we can always integrate by parts to get the other possibility. Therefore, our choice here is no restriction¹⁴.

If we introduce the matrices¹⁵

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma_\mu \\ \bar{\sigma}_\mu & 0 \end{pmatrix} \rightarrow \gamma_\mu = \begin{pmatrix} 0 & \bar{\sigma}_\mu \\ \sigma_\mu & 0 \end{pmatrix} \quad (6.13)$$

we can write the invariants we just found using the Dirac spinor formalism. Our invariants can be written using the matrices γ_μ and Dirac spinors as

$$\Psi^\dagger\gamma_0\Psi \quad \text{and} \quad \Psi^\dagger\gamma_0\gamma^\mu\partial_\mu\Psi, \quad (6.14)$$

because

$$\Psi^\dagger\gamma_0\Psi = \begin{pmatrix} (\chi_L)^\dagger & (\xi_R)^\dagger \end{pmatrix} \begin{pmatrix} 0 & \bar{\sigma}_0 \\ \sigma_0 & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \underbrace{(\chi_L)^\dagger\bar{\sigma}_0\xi_R}_{=I_1} + \underbrace{(\xi_R)^\dagger\sigma_0\chi_L}_{=I_2}$$

which are exactly¹⁶ the first two invariants we found earlier and

$$\begin{aligned} \Psi^\dagger\gamma_0\gamma^\mu\partial_\mu\Psi &= \begin{pmatrix} (\chi_L)^\dagger & (\xi_R)^\dagger \end{pmatrix} \begin{pmatrix} 0 & \bar{\sigma}_0 \\ \sigma_0 & 0 \end{pmatrix} \begin{pmatrix} 0 & \sigma^\mu\partial_\mu \\ \bar{\sigma}^\mu\partial_\mu & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \\ &= \underbrace{(\chi_L)^\dagger\bar{\sigma}_0\bar{\sigma}^\mu\partial_\mu\chi_L}_{=I_3} + \underbrace{(\xi_R)^\dagger\sigma_0\sigma^\mu\partial_\mu\xi_R}_{=I_4} \end{aligned}$$

gives the other two invariants¹⁷, as promised. It is conventional to introduce the notation

$$\bar{\Psi} := (\Psi)^\dagger\gamma_0. \quad (6.15)$$

Now we have everything we need to construct a Lorentz-invariant Lagrangian that is in agreement with the restrictions¹⁸ discussed in Sec. 4.2 using Dirac spinors:

$$\mathcal{L} = A\Psi^\dagger\gamma_0\Psi + B\Psi^\dagger\gamma_0\gamma^\mu\partial_\mu\Psi = A\bar{\Psi}\Psi + B\bar{\Psi}\gamma^\mu\partial_\mu\Psi.$$

Putting in the constants ($A = -m$, $B = i$) gives us the final **Dirac-Lagrangian**

$$\mathcal{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma^\mu\partial_\mu\Psi = \bar{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi. \quad (6.16)$$

Take note that what appears here in our Lagrangian are two distinct fields, because Ψ is complex¹⁹. This is a requirement, because other-

wise 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

$$\frac{\partial \mathcal{L}}{\partial \Psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Psi)} \right) = 0$$

we get

$$-m\bar{\Psi} - i\partial_\mu \bar{\Psi} \gamma^\mu = 0 \rightarrow (i\partial^\mu \bar{\Psi} \gamma_\mu + m\bar{\Psi}) = 0 \quad (6.17)$$

and with the Euler-Lagrange equation for the field $\bar{\Psi}$

$$\frac{\partial \mathcal{L}}{\partial \bar{\Psi}} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \bar{\Psi})} \right) = 0$$

we get the equation of motion for Ψ

$$(i\gamma_\mu \partial^\mu - m)\Psi = 0, \quad (6.18)$$

which is the famous **Dirac equation**. Take note that this is exactly what we get if we integrate the Lagrangian by parts

$$-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu \partial^\mu \Psi \underbrace{=} -m\bar{\Psi}\Psi - (i\partial^\mu \bar{\Psi})\gamma_\mu \Psi$$

Integrate by parts. Just imagine here the action integral.

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 asymmetry²⁰.

²⁰ You are free to use the longer Lagrangian that includes both possibilities, but the results are the same.

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 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 Sec. 4.2. We must combine an upper with a lower index, because we defined the scalar product of Minkowski space in Sec. 2.4 this way and scalars are what we need in the Lagrangian. The possible invariants are²¹

²¹ Again, a term of the form $\partial_\mu \partial^\mu A^\nu A_\nu$ is redundant to $\partial^\mu A^\nu \partial_\mu A_\nu$, because we can integrate by parts.

$$\begin{aligned} I_1 &= \partial^\mu A^\nu \partial_\mu A_\nu & , & & I_2 &= \partial^\mu A^\nu \partial_\nu A_\mu \\ I_3 &= A^\mu A_\mu & , & & I_4 &= \partial^\mu A_\mu \end{aligned}$$

and the Lagrangian reads

$$\mathcal{L}_{\text{Proka}} = C_1 \partial^\mu A^\nu \partial_\mu A_\nu + C_2 \partial^\mu A^\nu \partial_\nu A_\mu + C_3 A^\mu A_\mu + C_4 \partial^\mu A_\mu. \quad (6.19)$$

We can neglect the term $\partial^\mu A_\mu$, because it gives, using the Euler-Lagrange equation, a constant C_4 in our equations of motion, which does not have any influence. 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

$$\frac{\partial \mathcal{L}}{\partial A_\rho} = \partial_\sigma \left(\frac{\partial \mathcal{L}}{\partial (\partial_\sigma A_\rho)} \right),$$

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 :

$$\begin{aligned} & \partial_\sigma \left(\frac{\partial}{\partial (\partial_\sigma A_\rho)} (C_1 \partial^\mu A^\nu \partial_\mu A_\nu) \right) \\ & \stackrel{\text{product rule}}{=} C_1 \partial_\sigma \left((\partial_\mu A_\nu) \frac{\partial (\partial^\mu A^\nu)}{\partial (\partial_\sigma A_\rho)} + (\partial^\mu A^\nu) \frac{\partial (\partial_\mu A_\nu)}{\partial (\partial_\sigma A_\rho)} \right) \\ & \stackrel{\text{lowering indices with the metric}}{=} C_1 \partial_\sigma \left((\partial_\mu A_\nu) g^{\mu\kappa} g^{\nu\lambda} \frac{\partial (\partial_\kappa A_\lambda)}{\partial (\partial_\sigma A_\rho)} + (\partial^\mu A^\nu) \frac{\partial (\partial_\mu A_\nu)}{\partial (\partial_\sigma A_\rho)} \right) \\ & = C_1 \partial_\sigma \left((\partial_\mu A_\nu) g^{\mu\kappa} g^{\nu\lambda} \delta_\kappa^\sigma \delta_\lambda^\rho + (\partial^\mu A^\nu) \delta_\mu^\sigma \delta_\nu^\rho \right) \\ & = C_1 \partial_\sigma (\partial^\sigma A^\rho + \partial^\sigma A^\rho) = 2C_1 \partial_\sigma \partial^\sigma A^\rho \end{aligned}$$

Following similar steps we can compute

$$\partial_\sigma \left(\frac{\partial}{\partial (\partial_\sigma A_\rho)} (C_2 \partial^\mu A^\nu \partial_\nu A_\mu) \right) = 2C_2 \partial^\rho (\partial_\sigma A^\sigma)$$

and 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, if we put in the conventional constants

$$\rightarrow m^2 A^\rho = \frac{1}{2} \partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma). \quad (6.20)$$

²² Plural, because we have one equation for each component ρ .

These are called the **Proca equations**²², which are the wave equa-

tions for massive spin 1 particles. For massless ($m = 0$) spin 1 particles, e.g. photons, the equation reads

$$\rightarrow 0 = \frac{1}{2} \partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma). \quad (6.21)$$

These are the **inhomogeneous Maxwell equations** in absence of electric currents. It is conventional to define the electromagnetic tensor

$$F^{\sigma\rho} := \partial^\sigma A^\rho - \partial^\rho A^\sigma. \quad (6.22)$$

Then the inhomogeneous Maxwell-equations read

$$\partial_\rho F^{\sigma\rho} = 0. \quad (6.23)$$

Take note that we can rewrite the Lagrangian for massless spin 1 fields

$$\mathcal{L}_{\text{Maxwell}} = \frac{1}{2} (\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu)$$

as

$$\begin{aligned} \mathcal{L}_{\text{Maxwell}} &= \frac{1}{4} F^{\mu\nu} F_{\mu\nu} \\ &= \frac{1}{4} (\partial^\mu A^\nu - \partial^\nu A^\mu) (\partial_\mu A_\nu - \partial_\nu A_\mu) \\ &= \frac{1}{4} (\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\nu A^\mu \partial_\mu A_\nu + \partial^\nu A^\mu \partial_\nu A_\mu) \\ &\stackrel{\text{renaming dummy indices}}{=} \frac{1}{4} (\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu A_\mu + \partial^\mu A^\nu \partial_\mu A_\nu) \\ &= \frac{1}{4} (2\partial^\mu A^\nu \partial_\mu A_\nu - 2\partial^\mu A^\nu \partial_\nu A_\mu) \\ &= \frac{1}{2} (\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu) \quad \checkmark. \end{aligned} \quad (6.24)$$

$\mathcal{L}_{\text{Maxwell}} = \frac{1}{4} F^{\mu\nu} F_{\mu\nu}$ is the conventional way to write the Lagrangian. Equivalently, the Lagrangian for a massive spin 1 field can be written as

$$\mathcal{L}_{\text{Proca}} = \frac{1}{4} F^{\mu\nu} F_{\mu\nu} + m^2 A_\mu A^\mu = \frac{1}{2} (\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu) + m^2 A_\mu A^\mu. \quad (6.25)$$

In this chapter we derived the equations of motion that describe free fields/particles. We want to understand what we can do with these equations in order to get predictions for experiments, but 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/particles.

7

Interaction Theory

Summary

In this chapter we will derive how different fields interact with each other. This will enable us, for example, to describe how electrons, interact with photons¹.

We will be guided to the correct form of the Lagrangians by internal symmetries, which are in this context often called gauge² symmetries. The starting point will be local³ $U(1)$ symmetry and we end up with the Lagrangian

$$\mathcal{L} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + A_\mu\bar{\Psi}\gamma^\mu\Psi + \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu,$$

which 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 Lagrangian⁴

¹ From a different point of view: How electron (= massive spin $\frac{1}{2}$) fields interact with photon (=massless spin 1) fields.

² This strange name will be explained in a moment.

³ This means that instead of $e^{i\alpha}$, we transform each point in spacetime with a different factor, which is mathematically expressed as $e^{i\alpha(x)}$. 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 $W_j^{\mu\nu}$ will be defined similar for the three fields W_j^μ as $F_{\mu\nu}$ for the $U(1)$ gauge field A^μ .

$$\mathcal{L} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\sigma_j W_j^\mu\Psi - \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i,$$

which includes three spin 1 fields W_j^μ . We need three fields to make the Lagrangian locally $SU(2)$ invariant, because we have three basis generators of the $SU(2)$ group $J_i = \frac{\sigma_i}{2}$. We will see that local $SU(2)$ symmetry can only be achieved without "mass terms" of the form $m\bar{\Psi}\Psi$, $mW_\mu W^\mu$, 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_j^μ 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_j^μ 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 same formalism enables us to add arbitrary mass terms for the spin $\frac{1}{2}$ fields, as required by experiments. The final interaction Lagrangian describes a new interaction, called the **weak interaction**, which is mediated by three⁵ massive spin 1 fields, called W^+, W^- and Z . 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 $SU(3)$ 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 contains 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 bosons⁶, called gluons, are massless⁷. From experiments we know that only quarks (spin $\frac{1}{2}$) and gluons (spin 1) carry color. The resulting Lagrangian

⁵ To be precise: We start with the symmetry $U(1) \otimes SU(2)$, which is broken to $U(1)$. This process results in three massive vector bosons: W^+, W^-, Z , and one massless vector boson: The photon γ . Therefore it is often said that electromagnetism (derived from $U(1)$ symmetry) and the theory of weak-interactions (derived from $SU(2)$ symmetry) are unified. The $U(1)$ symmetry we start with, is different from the $U(1)$ symmetry that remains at the end. The photon and the Z -boson can then be seen as linear combinations of two vector bosons, commonly named B and W^3 , that were introduced to make the Lagrangian locally $U(1)$ (B -boson) and $SU(2)$ (W^1, W^2, W^3 -bosons) invariant.

⁶ 8 because $SU(3)$ has 8 basis generators.

⁷ 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 the $SU(3)$ symmetry is unbroken.

$$\mathcal{L} = -\frac{1}{4}F_{\alpha\beta}^A F_A^{\alpha\beta} + \bar{Q}(iD_\mu\gamma^\mu - m)Q,$$

will only be cited, because the derivation is tedious and 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 symmetries, often 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 electromagnetism⁸

"as a consequence of spacetime symmetry, specifically symmetry under local changes of length scale."

⁸ Frank Wilczek. Riemann-einstein structure from volume and gauge symmetry. *Phys. Rev. Lett.*, 80:4851–4854, Jun 1998. DOI: 10.1103/PhysRevLett.80.4851

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

Have a look again at the Lagrangian, we derived for a free spin $\frac{1}{2}$ theory (Eq. 6.16)

$$\mathcal{L}_{\text{Dirac}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi = \bar{\Psi}(i\gamma_\mu\partial^\mu - m)\Psi. \quad (7.1)$$

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 according to

$$\Psi \rightarrow \Psi' = e^{ia}\Psi$$

$$\Rightarrow \bar{\Psi} \rightarrow \bar{\Psi}' = \Psi'^{\dagger} \gamma_0 = (e^{ia} \Psi)^{\dagger} \gamma_0 = \bar{\Psi} e^{-ia}, \quad (7.2)$$

⁹ Remember $\bar{\Psi} = \Psi^{\dagger} \gamma_0$

where the minus sign comes from complex conjugation⁹ and a is an arbitrary real number. To see this we write the Lagrangian for the transformed fields explicitly

$$\begin{aligned} \mathcal{L}'_{\text{Dirac}} &= -m \bar{\Psi}' \Psi' + i \bar{\Psi}' \gamma_{\mu} \partial^{\mu} \Psi' \\ &= -m (\bar{\Psi} e^{-ia}) (e^{ia} \Psi) + i (\bar{\Psi} e^{-ia}) \gamma_{\mu} \partial^{\mu} (e^{ia} \Psi) \\ &= -m \bar{\Psi} \Psi \underbrace{e^{-ia} e^{ia}}_{=1} + i \bar{\Psi} \gamma_{\mu} \partial^{\mu} \Psi \underbrace{e^{-ia} e^{ia}}_{=1} \\ &= -m \bar{\Psi} \Psi + i \bar{\Psi} \gamma_{\mu} \partial^{\mu} \Psi = \mathcal{L}_{\text{Dirac}}, \end{aligned} \quad (7.3)$$

¹⁰ 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 freely¹⁰. Remembering that we learned in Chap. 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. It may seem at a first glance like a cute, but rather useless, mathematical side note. Stay tuned, because we will see in a moment that this observation is incredibly important!

Let's take now 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 Sec. 2.4. The global symmetry 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

$$\begin{aligned} \Psi &\rightarrow \Psi' = e^{ia(x)} \Psi \\ \Rightarrow \bar{\Psi} &\rightarrow \bar{\Psi}' = e^{-ia(x)} \bar{\Psi}, \end{aligned} \quad (7.4)$$

where the factor $a = a(x)$ now depends on the position, we get the transformed Lagrangian¹¹

¹¹ Maybe you wonder if the Lagrangian that includes both possible derivatives, which we neglected for brevity, is locally $U(1)$ invariant: $\mathcal{L} = -m \bar{\Psi} \Psi + i \bar{\Psi} \gamma_{\mu} \partial^{\mu} \Psi + i (\partial^{\mu} \bar{\Psi}) \gamma_{\mu} \Psi$. 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:

$$i \bar{\Psi} \gamma_{\mu} \partial^{\mu} \Psi + i (\partial^{\mu} \bar{\Psi}) \gamma_{\mu} \Psi$$

$$\underbrace{=}_{\text{integration by parts}} i (\partial^{\mu} \bar{\Psi}) \gamma_{\mu} \Psi - i (\partial^{\mu} \bar{\Psi}) \gamma_{\mu} \Psi = 0.$$

integration by parts

The correct Lagrangian that includes both possible derivatives has a minus sign between those terms: $\mathcal{L}_{\text{Dirac}} = -m \bar{\Psi} \Psi + i \bar{\Psi} \gamma_{\mu} \partial^{\mu} \Psi - i (\partial^{\mu} \bar{\Psi}) \gamma_{\mu} \Psi$ and is therefore **not** locally $U(1)$ invariant.

$$\begin{aligned}
\mathcal{L}'_{\text{Dirac}} &= -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' \\
&= -m(\underbrace{\bar{\Psi}e^{-ia(x)}}_{=1})(e^{ia(x)}\Psi) + i(\bar{\Psi}e^{-ia(x)})\gamma_\mu\partial^\mu(e^{ia(x)}\Psi) \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu(\partial^\mu\Psi)\underbrace{e^{-ia(x)}e^{ia(x)}}_{=1} + i(e^{-ia(x)}\bar{\Psi})\gamma_\mu\Psi(\partial^\mu e^{ia(x)}) \\
&\quad \text{Product rule} \\
&= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + i^2(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi \neq \mathcal{L}_{\text{Dirac}} \quad (7.5)
\end{aligned}$$

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 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 particles¹²

$$\mathcal{L}_{\text{Proca}} = \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu + m^2 A^\mu A_\mu. \quad (7.6)$$

¹² See Eq. 6.25 and take note that we neglect, for brevity, a conventional factor $\frac{1}{2}$ here.

We can discover a **global** internal symmetry here, too. If we transform

$$A_\mu \rightarrow A'_\mu = A_\mu + a_\mu \quad (7.7)$$

with some arbitrary constants a_μ , the Lagrangian reads

$$\begin{aligned}
\mathcal{L}'_{\text{Proca}} &= (\partial^\mu A'^\nu \partial_\mu A'_\nu - \partial^\mu A'^\nu \partial_\nu A'_\mu) + m^2 A'^\mu A'_\mu \\
&= \partial^\mu (A^\nu + a^\nu) \partial_\mu (A_\nu + a_\nu) - \partial^\mu (A^\nu + a^\nu) \partial_\nu (A_\mu + a_\mu) + m^2 (A_\mu + a_\mu)(A^\mu + a^\mu) \\
&= \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu + m^2 (A_\mu + a_\mu)(A^\mu + a^\mu). \quad (7.8)
\end{aligned}$$

We conclude this transformation is a **global** symmetry transformation of this Lagrangian, if we restrict to **massless** fields, i.e. $m = 0$.

What about **local** symmetry here? We transform

$$A_\mu \rightarrow A'_\mu = A_\mu + a_\mu(x) \quad (7.9)$$

and the transformed **massless** Lagrangian reads

$$\begin{aligned}
\mathcal{L}'_{\text{Maxwell}} &= (\partial^\mu A'^\nu \partial_\mu A'_\nu - \partial^\mu A'^\nu \partial_\nu A'_\mu) \\
&= \partial^\mu (A^\nu + a^\nu(x)) \partial_\mu (A_\nu + a_\nu(x)) - \partial^\mu (A^\nu + a^\nu(x)) \partial_\nu (A_\mu + a_\mu(x)) \\
&= \partial^\mu A^\nu \partial_\mu A_\nu + \partial^\mu a^\nu \partial_\mu A_\nu + \partial^\mu A^\nu \partial_\mu a_\nu(x) + \partial^\mu a^\nu(x) \partial_\mu A_\nu(x) \\
&\quad - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu a_\mu(x) - \partial^\mu a^\nu(x) \partial_\nu A_\mu - \partial^\mu a^\nu(x) \partial_\nu a_\mu(x) \quad (7.10)
\end{aligned}$$

which shows this is no local internal symmetry .

Nevertheless, we can find a local internal symmetry if we transform instead $A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu a(x)$. This means we add the derivative $\partial_\mu a(x)$ of some arbitrary function instead of an arbitrary function. This yields

$$\begin{aligned}
 \mathcal{L}'_{\text{Maxwell}} &= \partial^\mu A'^\nu \partial_\mu A'_\nu - \partial^\mu A'^\nu \partial_\nu A'_\mu \\
 &= \partial^\mu (A^\nu + \partial^\nu a(x)) \partial_\mu (A_\nu + \partial_\nu a(x)) - \partial^\mu (A^\nu + \partial^\nu a(x)) \partial_\nu (A_\mu + \partial_\mu a(x)) \\
 &= \partial^\mu A^\nu \partial_\mu A_\nu + \partial^\mu (\partial^\nu a(x)) \partial_\mu A_\nu + \partial^\mu A^\nu \partial_\mu (\partial_\nu a(x)) + \partial^\mu (\partial^\nu a(x)) \partial_\mu (\partial_\nu a(x)) \\
 &\quad - \partial^\mu A^\nu \partial_\nu A_\mu - \partial^\mu A^\nu \partial_\nu (\partial_\mu a(x)) - \partial^\mu (\partial^\nu a(x)) \partial_\nu A_\mu - \partial^\mu (\partial^\nu a(x)) \partial_\nu (\partial_\mu a(x)) \\
 &\quad \underbrace{= \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu}_{\partial_\nu \partial_\mu = \partial_\mu \partial_\nu \text{ and renaming dummy indices}} = \mathcal{L}_{\text{Maxwell}}
 \end{aligned} \tag{7.11}$$

and we see this is indeed an internal local symmetry transformation. This may look again like a technical side note. Okay, we found some internal, local symmetry, so now 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 $\Psi \rightarrow \Psi' = e^{ia} \Psi$. 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

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu a(x), \tag{7.13}$$

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_μ .

This really looks like two pieces of a puzzle we should put together: An extra term $A_\mu \bar{\Psi} \gamma^\mu \Psi$ in the Lagrangian transforms into

$$A_\mu \bar{\Psi} \gamma^\mu \Psi \rightarrow (A_\mu + \partial_\mu a(x)) \bar{\Psi} \gamma^\mu \Psi = A_\mu \bar{\Psi} \gamma^\mu \Psi + \partial_\mu a(x) \bar{\Psi} \gamma^\mu \Psi. \tag{7.14}$$

Compare the second term 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 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

$$-(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi \rightarrow -g(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi. \quad (7.15)$$

This extra factor g accounts for an arbitrary coupling constant¹³ g , 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, because otherwise A_μ has an unmatched index μ and therefore wouldn't be Lorentz invariant, and inserted the **coupling constant**¹⁴ g . This yields the Lagrangian

$$\mathcal{L}_{\text{Dirac+Extra-Term}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi.$$

Transforming this Lagrangian according to the rules for local transformations of Ψ , $\bar{\Psi}$ and A_μ yields¹⁵

$$\begin{aligned} \mathcal{L}'_{\text{Dirac+Extra-Term}} &= -m\bar{\Psi}'\Psi' + i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' + gA'_\mu\bar{\Psi}'\gamma^\mu\Psi' \\ &\stackrel{\text{See Eq. 7.5}}{=} \underbrace{-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - g(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi}_{\text{See Eq. 7.5}} + gA'_\mu\bar{\Psi}'\gamma^\mu\Psi' \\ &= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - g(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi + g(A_\mu + \partial_\mu a(x))(e^{-iga(x)}\bar{\Psi})\gamma^\mu(e^{iga(x)}\Psi) \\ &= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \underbrace{g(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi}_{\text{See Eq. 7.5}} + gA_\mu\bar{\Psi}\gamma^\mu\Psi + g\underbrace{(\partial_\mu a(x))\bar{\Psi}\gamma^\mu\Psi}_{=(\partial^\mu a(x))\bar{\Psi}\gamma_\mu\Psi} \\ &= -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi = \mathcal{L}_{\text{Dirac+Extra-Term}}. \end{aligned} \quad (7.16)$$

Therefore, by adding an extra term we get, as promised, 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

$$\mathcal{L}_{\text{Dirac+Extra-Term+Maxwell}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + gA_\mu\bar{\Psi}\gamma^\mu\Psi + \partial^\mu A^\nu\partial_\mu A_\nu - \partial^\mu A^\nu\partial_\nu A_\mu. \quad (7.17)$$

It is conventional to introduce a new symbol

$$D_\mu \equiv i\partial_\mu + gA_\mu, \quad (7.18)$$

¹³ A coupling constant always tells us how strong a given interaction is. Here we are talking about electromagnetic interactions and g determines its strength.

¹⁴ We can see here that g determines how strong Ψ , $\bar{\Psi}$ and A_μ couple together.

¹⁵ The combined transformation of Ψ , $\bar{\Psi}$ and A_μ is called $U(1)$ **gauge transformation**.

called **covariant derivative**. The Lagrangian then reads

$$\begin{aligned}\mathcal{L}_{\text{Dirac+Extra-Term+Maxwell}} &= -m\bar{\Psi}\Psi + \bar{\Psi}\gamma_\mu \underbrace{(i\partial^\mu + gA_\mu)}_{\equiv D^\mu} \Psi + \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu \\ &= -m\bar{\Psi}\Psi + \bar{\Psi}\gamma_\mu D^\mu \Psi + \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu.\end{aligned}\quad (7.19)$$

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 observing 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?

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

$$\mathcal{L}_{\text{Dirac+Extra-Term+Proca}} = -m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu \partial^\mu \Psi + gA_\mu \bar{\Psi}\gamma^\mu \Psi + \partial^\mu A^\nu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu$$

into the Euler-Lagrange equation for each field

$$\frac{\partial \mathcal{L}}{\partial \Psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Psi)} \right) = 0$$

$$\frac{\partial \mathcal{L}}{\partial \bar{\Psi}} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \bar{\Psi})} \right) = 0$$

$$\frac{\partial \mathcal{L}}{\partial A_\rho} - \partial_\sigma \left(\frac{\partial \mathcal{L}}{\partial (\partial_\sigma A_\rho)} \right) = 0.$$

This yields

$$\bar{\Psi}(i\gamma_\mu \partial^\mu + m) + gA_\mu \bar{\Psi}\gamma^\mu = 0 \quad (7.20)$$

$$(i\gamma_\mu \partial^\mu - m)\Psi - gA_\mu \gamma^\mu \Psi = 0 \quad (7.21)$$

$$\partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma) + g\bar{\Psi}\gamma^\mu \Psi = 0 \quad (7.22)$$

The first two equations describe the behaviour 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

$$\partial_\mu \rightarrow D_\mu = i\partial_\mu + A_\mu \quad (7.23)$$

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 Sec. 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 Sec. 3.7.10 the transformation (Eq. 3.236)

$$\Psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \rightarrow \Psi^C = \begin{pmatrix} \tilde{\xi}_L \\ \chi_R \end{pmatrix}. \quad (7.24)$$

This transformation can now be described easily using one of the γ_μ matrices. Using the definition of γ_2 in Eq. 6.13, we have

$$\Psi^C = i\gamma_2\Psi^* = i \begin{pmatrix} 0 & \sigma_2 \\ -\sigma_2 & 0 \end{pmatrix} \begin{pmatrix} \chi_L^* \\ \xi_R^* \end{pmatrix}, \quad (7.25)$$

because we can rewrite this, using that $i\sigma_2 = \epsilon$ is exactly the spinor metric

$$= \begin{pmatrix} 0 & \epsilon \\ -\epsilon & 0 \end{pmatrix} \begin{pmatrix} \chi_L^* \\ \xi_R^* \end{pmatrix} = \begin{pmatrix} \epsilon\xi_R^* \\ -\epsilon\chi_L^* \end{pmatrix}. \quad (7.26)$$

This is equivalent to

$$= \begin{pmatrix} \tilde{\xi}_L \\ \chi_R \end{pmatrix}, \quad (7.27)$$

as was shown in Sec. 3.7.7, specifically Eq. 3.194. Therefore we start with Eq. 7.21:

$$(i\gamma_\mu\partial^\mu - m)\Psi - gA_\mu\gamma^\mu\Psi = ((\gamma_\mu(i\partial^\mu - gA^\mu) - m)\Psi = 0, \quad (7.28)$$

and complex conjugate this equation, as a first step towards an equation for Ψ^C :

$$\rightarrow (\gamma_\mu^*(-i\partial^\mu - gA^\mu) - m)\Psi^* = 0. \quad (7.29)$$

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 Ψ :

$$\rightarrow \gamma_2(\gamma_\mu^*(-i\partial^\mu - gA^\mu) - m) \underbrace{\gamma_2^{-1}\gamma_2}_{=1}\Psi^* = 0. \quad (7.30)$$

$$\rightarrow (\underbrace{\gamma_2\gamma_\mu^*\gamma_2^{-1}}_{=-\gamma_\mu}(-i\partial^\mu - gA^\mu) - m\gamma_2\gamma_2^{-1})\gamma_2\Psi^* = 0. \quad (7.31)$$

$$\underbrace{\rightarrow}_{\text{Multiplying the equation with } i} (-\gamma_\mu(-i\partial^\mu - gA^\mu) - m) \underbrace{i\gamma_2\Psi^*}_{=\Psi^C \text{ see Eq. 7.25}} = 0. \quad (7.32)$$

$$\rightarrow ((\gamma_\mu(i\partial^\mu + gA^\mu) - m)\Psi^C = 0, \quad (7.33)$$

This is exactly the same equation of motion as for Ψ , but with opposite coupling strength $g \rightarrow -g$. This justifies the name charge conjugation¹⁶.

Next, we turn to the third equation of motion derived the last section, Eq. 7.22, which is called **inhomogeneous Maxwell equation in the presence of an electric current**. To make this precise we need again Noether's theorem.

7.1.6 Noether's Theorem for Internal $U(1)$ Symmetry

In Sec. 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

$$J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Psi)} \delta\Psi$$

which fulfils a continuity equation

$$\partial_\mu J^\mu = 0. \quad (7.34)$$

A global¹⁷ $U(1)$ transformation is

$$\Psi \rightarrow \Psi' = e^{iga}\Psi = (1 + iga + \dots)\Psi.$$

¹⁶ It is important to note that $\Psi^c \neq \Psi$. $\Psi^c = i\gamma_2\Psi^*$ and $\bar{\Psi} = \Psi^\dagger\gamma_0 = (\Psi^*)^T\gamma_0$. Charge conjugation is the correct transformation that enables us to interpret things in terms of antiparticles, as we will discuss later in detail.

¹⁷ 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 $a = \text{const}$. Therefore, if we have a locally $U(1)$ invariant Lagrangian, it is automatically globally $U(1)$ invariant, too. Considering **global** symmetry $U(1)$ here will give us a quantity that is conserved for free **and** interacting fields.

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

$$\Psi \rightarrow \Psi' = \Psi + ig a \Psi.$$

Therefore we have $\delta\Psi = ig a \Psi$ and as we derived in Sec. 4.5.5 the corresponding Noether current is

$$\begin{aligned} J^\mu &= \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Psi)} \delta\Psi \\ &= \frac{\partial(-m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_\mu\partial^\mu\Psi)}{\partial(\partial_\mu \Psi)} ig a \Psi \\ &= -\bar{\Psi}\gamma^\mu ga\Psi = -ga\bar{\Psi}\gamma^\mu\Psi. \end{aligned} \quad (7.35)$$

We can ignore¹⁸ the arbitrary constant a , because the continuity equation holds for arbitrary a and therefore, we define

$$J^\mu \equiv -g\bar{\Psi}\gamma^\mu\Psi \quad (7.36)$$

This is usually called the **electric four-current**. The zeroth component is the **electric charge density**, which gives us if we integrate over all space a quantity that is conserved in time¹⁹

$$Q = \int d^3x \underbrace{\rho}_{\text{Charge density}} = \int d^3x J^0 = -g \int d^3x \bar{\Psi}\gamma^0\Psi. \quad (7.37)$$

In the quantum framework the objects Ψ will be related to probability amplitudes which will require 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 proportional to electric charge. **Therefore, global $U(1)$ symmetry leads us to 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

$$\begin{aligned} \partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) + \underbrace{g\bar{\Psi}\gamma^\rho\Psi}_{=-J^\rho} &= 0 \\ \rightarrow \partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) &= J^\rho. \end{aligned} \quad (7.38)$$

Using the electromagnetic tensor as defined in Eq. 6.22 this equation reads

$$\partial_\sigma F^{\rho\sigma} = J^\rho. \quad (7.39)$$

These are the inhomogeneous Maxwell equations in the presence of an external electromagnetic current. These equations²⁰, together

¹⁸ We keep the conventional constant g , which is not arbitrary but has one fixed value that is determined in experiments.

¹⁹ This can be seen by following the same steps as in Eq. 4.39.

²⁰ Plural, because we have an equation for each component ρ .

²¹ We will see this in Chap. 11.

with the homogeneous Maxwell equations, which follow immediately²¹ from the definition of $F^{\rho\sigma}$, are the basis for the classical theory of electrodynamics.

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

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2)$$

is not $U(1)$ invariant, as we can see by transforming $\Phi \rightarrow \Phi' = e^{ia}\Phi$. Nevertheless, the complex scalar theory

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi^* \partial^\mu \Phi - m^2 \Phi^* \Phi). \quad (7.40)$$

has $U(1)$ symmetry, because then we have $\Phi \rightarrow \Phi' = e^{ia}\Phi$ and $\Phi^* \rightarrow (\Phi^*)' = e^{-ia}\Phi^*$. Therefore it is possible to derive, analogous to what we did in Sec. 7.1 for spin $\frac{1}{2}$ fields, an interaction theory for this Lagrangian. The derivation is completely analogous²² and one gets

$$\mathcal{L} = \frac{1}{2}((\partial_\mu - iqA_\mu)\Phi^*)((\partial^\mu + iqA^\mu)\Phi) - m^2 \Phi^* \Phi \quad (7.41)$$

Using the Euler-Lagrange equations

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \Phi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right) &= 0 \\ \frac{\partial \mathcal{L}}{\partial \Phi^*} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi^*)} \right) &= 0, \end{aligned}$$

we find the corresponding equations of motion

$$(\partial_\mu - iqA_\mu)(\partial^\mu + iqA^\mu)\Phi^* - m^2 \Phi^* = 0 \quad (7.42)$$

$$(\partial_\mu - iqA_\mu)(\partial^\mu + iqA^\mu)\Phi - m^2 \Phi = 0, \quad (7.43)$$

which describe each 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

$$\mathcal{L}_{\text{Maxwell}} = \frac{1}{4}F^{\mu\nu}F_{\mu\nu}.$$

²² The correct Lagrangian can be computed by substituting $\partial_\mu \rightarrow D_\mu = i\partial_\mu + A_\mu$ as introduced in Eq. 7.23.

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)

$$\mathcal{L}_{\text{Proca}} = \frac{1}{2} G^{\mu\nu} G_{\mu\nu} + m^2 B^\mu B_\mu.$$

Lorentz symmetry dictates the interaction term in the Lagrangian to be of the form

$$\mathcal{L}_{\text{Proca-interaction}} = C G_{\mu\nu} F^{\mu\nu},$$

with the coupling constant C 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 addition of two copies of the Lagrangian we derived in Sec. 6.3. The final Lagrangian we derived can be found in Eq. 6.16:

$$\mathcal{L}_{\text{Dirac}} = \bar{\psi}(i\gamma_\mu \partial^\mu - m)\psi.$$

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

$$\mathcal{L}_{\text{D1+D2}} = i\bar{\psi}_1 \gamma_\mu \partial^\mu \psi_1 + i\bar{\psi}_2 \gamma_\mu \partial^\mu \psi_2 \quad (7.44)$$

and this can be rewritten, if we define

$$\begin{aligned} \Psi &:= \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \\ \rightarrow \bar{\Psi} &:= (\bar{\psi}_1 \quad \bar{\psi}_2) \end{aligned}$$

where the newly defined object Ψ is called a **doublet**. Thus

$$\mathcal{L}_{\text{D1+D2}} = i\bar{\Psi} \gamma_\mu \partial^\mu \Psi. \quad (7.45)$$

This Lagrangian is invariant under **global** $SU(2)$ transformations

$$\Psi \rightarrow \Psi' = e^{ia_i \frac{\sigma_i}{2}} \Psi \quad (7.46)$$

$$\Rightarrow \bar{\Psi} \rightarrow \bar{\Psi}' = \bar{\Psi} e^{-ia_i \frac{\sigma_i}{2}}, \quad (7.47)$$

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 .

²³ We neglect mass terms here, because these 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 defining

$$m := \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}$$

as

$$\mathcal{L}_{D1+D2} = -\bar{\Psi} m \Psi.$$

Unfortunately, this term is not invariant under $SU(2)$ transformations, because

$$\mathcal{L}_{D1+D2} = -\bar{\Psi}' m \Psi' = \bar{\Psi} \underbrace{e^{-ia_i \frac{\sigma_i}{2}} m e^{ia_i \frac{\sigma_i}{2}}}_{\neq m} \Psi.$$

For equal masses $m_1 = m_2$ it would be invariant again, 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.

²⁴ This is known as **choosing a gauge**.

To see the invariance we take a look at the transformed Lagrangian²³

$$\begin{aligned} \mathcal{L}'_{D1+D2} &= i\bar{\Psi}' \gamma_\mu \partial^\mu \Psi' \\ &= i\bar{\Psi} e^{-ia_i \frac{\sigma_i}{2}} \gamma_\mu \partial^\mu e^{ia_i \frac{\sigma_i}{2}} \Psi \\ &= i\bar{\Psi} \gamma_\mu \partial^\mu \Psi = \mathcal{L}_{D1+D2}, \quad \checkmark \end{aligned} \quad (7.48)$$

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

$$\left[(e^{-ia_i \frac{\sigma_i}{2}})_{ab} \delta_{\alpha\beta} \right] \left[\delta_{bc} \gamma_\mu^{\beta\delta} \right] \left[(e^{ia_i \frac{\sigma_i}{2}})_{cd} \delta_{\delta\epsilon} \right] = \left[\delta_{ad} \gamma_\mu^{\alpha\epsilon} \right]$$

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 choice²⁴, 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:

$$\begin{aligned} \mathcal{L}'_{D1+D2} &= i\bar{\Psi}' \gamma_\mu \partial^\mu \Psi' \\ &= i\bar{\Psi} e^{-ia_i(x) \frac{\sigma_i}{2}} \gamma_\mu \partial^\mu e^{ia_i(x) \frac{\sigma_i}{2}} \Psi \\ &= \underbrace{i\bar{\Psi} \gamma_\mu \partial^\mu \Psi}_{\text{product rule}} - \bar{\Psi} \gamma_\mu (\partial_\mu a_i(x) \frac{\sigma_i}{2}) \Psi \neq \mathcal{L}_{D1+D2}. \end{aligned} \quad (7.49)$$

We already know what to do next from the experience with $U(1)$, which we discussed in Sec. 7.1. We will, step by step, derive a locally $SU(2)$ invariant Lagrangian, which will take some time. To avoid confusion, the steps that follow are summarized in the following list:

- The final result will be that we get a locally $SU(2)$ invariant Lagrangian, by adding the extra term $i\bar{\Psi}\gamma_\mu W_i^\mu \sigma_i \Psi$ to the Lagrangian, which describes interactions between two spin $\frac{1}{2}$ fields $\begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$ and three massive spin 1 fields W_i^μ .
- In addition, this is only invariant if we transform

$$(W_\mu)_i \rightarrow (W'_\mu)_i = (W_\mu)_i + \partial_\mu a_i(x) + \epsilon_{ijk} a_j(x) (W_\mu)_k$$

instead of

$$(W_\mu)_i \rightarrow (W'_\mu)_i = (W_\mu)_i + \partial_\mu a_i(x),$$

which we used to fix local $U(1)$ symmetry.

- This is only a symmetry for the free $(W_\mu)_i$ Lagrangian

$$L = \frac{1}{4} (W_{\mu\nu})_i (W^{\mu\nu})_i$$

if we define

$$(W_{\mu\nu})_i = \partial_\mu (W_\nu)_i - \partial_\nu (W_\mu)_i + \epsilon_{ijk} (W_\mu)_j (W_\nu)_k$$

instead of

$$(W_{\mu\nu})_i = \partial_\mu (W_\nu)_i - \partial_\nu (W_\mu)_i.$$

In other words: In order to get a locally $SU(2)$ invariant Lagrangian, we must redefine the free Lagrangian for our three W_i^μ fields, such that it is invariant under

$$(W_\mu)_i \rightarrow (W'_\mu)_i = (W_\mu)_i + \partial_\mu a_i(x) + \epsilon_{ijk} a_j(x) (W_\mu)_k.$$

Now we will go through these steps in detail. We know the internal symmetry of spin 1 fields from Eq. 7.11

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu a(x)$$

and here we are going to use three spin 1 fields, W_1^μ , W_2^μ and W_3^μ , because we have three generators $\frac{\sigma_i}{2}$ for $SU(2)$ and need one field for each extra term, produced by the sum $a_i(x) \frac{\sigma_i}{2}$ in the exponent. These possess the local internal symmetry

$$W_i^\mu \rightarrow W'^\mu_i = W_i^\mu + \partial^\mu a_i(x)$$

and therefore, we try adding the term

$$\bar{\Psi} \gamma_\mu \frac{\sigma_j}{2} W_j^\mu \Psi$$

to the Lagrangian. Unfortunately, this does not lead to a locally $SU(2)$ invariant Lagrangian:

$$\begin{aligned}
\mathcal{L}'_{D1+D2+Extra} &= i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' + \bar{\Psi}'\gamma_\mu\frac{\sigma_j}{2}W_j'^\mu\Psi' \\
&= i\bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\partial^\mu e^{ia_i(x)\frac{\sigma_i}{2}}\Psi \\
&\quad + \bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\frac{\sigma_j}{2}(W_j^\mu + \partial_\mu a_j(x))e^{ia_i(x)\frac{\sigma_i}{2}}\Psi \\
&= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \bar{\Psi}\gamma_\mu(\partial_\mu a_i(x)\frac{\sigma_i}{2})\Psi \\
&\quad + \underbrace{\bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\frac{\sigma_j}{2}W_j^\mu e^{ia_i(x)\frac{\sigma_i}{2}}\Psi}_{\neq \gamma_\mu\frac{\sigma_j}{2}W_j^\mu \text{ because } [\frac{\sigma_i}{2}, \frac{\sigma_j}{2}] = \epsilon_{ijk}\frac{\sigma_k}{2} \neq 0} \\
&\quad + \bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\frac{\sigma_j}{2}(\partial_\mu a_j(x))e^{-ia_i(x)\frac{\sigma_i}{2}}\Psi \\
&= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \cancel{\bar{\Psi}\gamma_\mu(\partial_\mu a_i(x)\frac{\sigma_i}{2})\Psi} + \bar{\Psi}\gamma_\mu\sigma_i W_i^\mu\Psi \\
&\quad + \cancel{\bar{\Psi}\gamma_\mu\sigma_i(\partial_\mu a_i(x))\Psi} \\
&= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\sigma_i W_i^\mu\Psi = \mathcal{L}_{D1+D2+Extra} \quad \checkmark \quad (7.50)
\end{aligned}$$

²⁵ Mathematicians call a group with non commuting elements non-abelian. In contrast, $U(1)$ is abelian and therefore everything was easier.

We can see that the reason here is that the generators of $SU(2)$ do not commute²⁵. Let us take a closer look at the difficulty. We will look at, as usual for Lie groups, infinitesimal transformations, ignoring higher order terms:

$$\begin{aligned}
e^{-ia_i(x)\frac{\sigma_i}{2}}\frac{\sigma_j}{2}e^{ia_i(x)\frac{\sigma_i}{2}} &\approx \left(1 - ia_i(x)\frac{\sigma_i}{2}\right)\frac{\sigma_j}{2}\left(1 + ia_i(x)\frac{\sigma_i}{2}\right) \\
&= \frac{\sigma_j}{2} - a_i(x)\frac{\sigma_i}{2}\frac{\sigma_j}{2} + a_i(x)\frac{\sigma_j}{2}\frac{\sigma_i}{2} + \mathcal{O}(a^2) \\
&= \frac{\sigma_j}{2} + a_i(x)\left[\frac{\sigma_j}{2}, \frac{\sigma_i}{2}\right] + \mathcal{O}(a^2) \\
&= \frac{\sigma_j}{2} - a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2} + \mathcal{O}(a^2) \neq \frac{\sigma_j}{2} \quad (7.51)
\end{aligned}$$

Therefore, the infinitesimal transformed, which means we can ignore

terms of order $\mathcal{O}(a^2)$ and higher, Lagrangian is

$$\begin{aligned}
\mathcal{L}'_{D1+D2+Extra} &= i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' + \bar{\Psi}'\gamma_\mu\frac{\sigma_j}{2}W_j'^\mu\Psi' \\
&\stackrel{\text{Eq. 7.50}}{=} i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \bar{\Psi}\gamma_\mu(\partial_\mu a_i(x)\frac{\sigma_i}{2})\Psi + \bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\frac{\sigma_j}{2}W_j^\mu e^{ia_i(x)\frac{\sigma_i}{2}}\Psi \\
&\quad + \bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\frac{\sigma_j}{2}(\partial_\mu a_j(x))e^{-ia_i(x)\frac{\sigma_i}{2}}\Psi \\
&\stackrel{\text{Eq. 7.51}}{\approx} i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \bar{\Psi}\gamma_\mu(\partial_\mu a_i(x)\frac{\sigma_i}{2})\Psi + \bar{\Psi}\gamma_\mu(\frac{\sigma_j}{2} - a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2})W_j^\mu\Psi \\
&\quad + \bar{\Psi}\gamma_\mu(\frac{\sigma_j}{2} - a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2})(\partial_\mu a_j(x))\Psi \\
&= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi - \bar{\Psi}\gamma_\mu(\partial_\mu a_i(x)\frac{\sigma_i}{2})\Psi + \bar{\Psi}\gamma_\mu\frac{\sigma_j}{2}W_j^\mu\Psi - \bar{\Psi}\gamma_\mu a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2}W_j^\mu\Psi + \bar{\Psi}\gamma_\mu\frac{\sigma_j}{2}(\partial_\mu a_j(x))\Psi \\
&\quad - \underbrace{\bar{\Psi}\gamma_\mu a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2}(\partial_\mu a_j(x))\Psi}_{\mathcal{O}(a^2)} \\
&= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\frac{\sigma_j}{2}W_j^\mu\Psi - \bar{\Psi}\gamma_\mu a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2}W_j^\mu\Psi \quad (7.52)
\end{aligned}$$

The internal symmetry of $(W_\mu)_i$ that would fix local $SU(2)$ invariance is

$$(W_\mu)_i \rightarrow (W'_\mu)_i = (W_\mu)_i + \partial_\mu a_i(x) + \epsilon_{ijk}a_j(x)(W_\mu)_k$$

because this gives an extra term in the Lagrangian that cancels exactly the symmetry-destroying term

$$\begin{aligned}
\mathcal{L}'_{D1+D2+Extra} &= i\bar{\Psi}'\gamma_\mu\partial^\mu\Psi' + \bar{\Psi}'\gamma_\mu\frac{\sigma_j}{2}W_j'^\mu\Psi' \\
&\stackrel{\text{Eq. 7.52}}{=} i\bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\partial^\mu e^{ia_i(x)\frac{\sigma_i}{2}}\Psi + \bar{\Psi}e^{-ia_i(x)\frac{\sigma_i}{2}}\gamma_\mu\frac{\sigma_j}{2}(W_j^\mu + \partial_\mu a_j(x) + \epsilon_{jlm}a_l(x)W_m^\mu)e^{ia_i(x)\frac{\sigma_i}{2}}\Psi \\
&= i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\frac{\sigma_j}{2}W_j^\mu\Psi - \bar{\Psi}\gamma_\mu a_i(x)\epsilon_{ijk}\frac{\sigma_k}{2}W_j^\mu\Psi + \bar{\Psi}\gamma_\mu\frac{\sigma_j}{2}\epsilon_{jlm}a_l(x)W_m^\mu\Psi \\
&= \mathcal{L}_{D1+D2+Extra} \quad \checkmark \quad (7.53)
\end{aligned}$$

We therefore have to take a step back now and have a look if this is an internal symmetry of the three free spin 1 field Lagrangian. The Lagrangian for the three massless spin 1 fields is

$$\begin{aligned}
\mathcal{L}_{3\text{xMaxwell}} &= \frac{1}{4}(W_{\mu\nu})_1(W^{\mu\nu})_1 + \frac{1}{4}(W_{\mu\nu})_2(W^{\mu\nu})_2 + \frac{1}{4}(W_{\mu\nu})_3(W^{\mu\nu})_3 \\
&= \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i \quad (7.54)
\end{aligned}$$

with

$$(W_{\mu\nu})_i = \partial_\mu(W_\nu)_i - \partial_\nu(W_\mu)_i.$$

We want to see if this Lagrangian is invariant under

$$(W_\mu)_i \rightarrow (W'_\mu)_i = ((W_\mu)_i + \partial_\mu a_i(x) + \epsilon_{ijk}a_j(x)(W_\mu)_k) \quad (7.55)$$

The transformed Lagrangian is

$$\begin{aligned}
\mathcal{L}'_{3\text{xMaxwell}} &= \frac{1}{4} (W'_{\mu\nu})_i (W'^{\mu\nu})_i \\
&\stackrel{\text{See Eq. 6.25}}{=} (\partial^\mu (W'^\nu)_i \partial_\mu (W'_\nu)_i - \partial^\mu (W'^\nu)_i \partial_\nu (W'_\mu)_i) \\
&= (\partial^\mu ((W^\nu)_i + \partial^\nu a_i(x) + \epsilon_{ijk} a_j(x) (W^\nu)_k) \partial_\mu ((W_\nu)_i + \partial_\nu a_i(x) + \epsilon_{ijk} a_j(x) (W_\nu)_k) \\
&\quad - \partial^\mu ((W^\nu)_i + \partial^\nu a_i(x) + \epsilon_{ijk} a_j(x) (W^\nu)_k) \partial_\nu ((W_\mu)_i + \partial_\mu a_i(x) + \epsilon_{ijk} a_j(x) (W_\mu)_k)) \\
&\stackrel{\text{Ignoring terms of order } \mathcal{O}(a^2)}{=} \partial^\mu (W^\nu)_i \partial_\mu (W_\nu)_i + \cancel{\partial^\mu (W^\nu)_i \partial_\mu \partial_\nu a_i(x)} + (\partial^\mu (W^\nu)_i) \partial_\mu \epsilon_{ijk} a_j(x) (W_\nu)_k \\
&\quad + \cancel{\partial^\mu \partial^\nu a_i(x) \partial_\mu (W_\nu)_i} + \partial^\mu \epsilon_{ijk} a_j(x) (W^\nu)_k \partial_\mu (W_\nu)_i \\
&\quad - \partial^\mu (W^\nu)_i \partial_\nu (W_\mu)_i - \cancel{\partial^\mu (W^\nu)_i \partial_\nu \partial_\mu a_i(x)} - \partial^\mu (W^\nu)_i \partial_\nu \epsilon_{ijk} a_j(x) (W_\mu)_k \\
&\quad - \cancel{\partial^\mu \partial^\nu a_i(x) \partial_\nu (W_\mu)_i} - \partial^\mu \epsilon_{ijk} a_j(x) (W^\nu)_k \partial_\nu (W_\mu)_i \\
&= \underbrace{\partial^\mu (W^\nu)_i \partial_\mu (W_\nu)_i - \partial^\mu (W^\nu)_i \partial_\nu (W_\mu)_i}_{=\mathcal{L}_{3\text{xMaxwell}}} + (\partial^\mu (W^\nu)_i) \partial_\mu \epsilon_{ijk} a_j(x) (W_\nu)_k \\
&\quad + \partial^\mu \epsilon_{ijk} a_j(x) (W^\nu)_k \partial_\mu (W_\nu)_i - \partial^\mu (W^\nu)_i \partial_\nu \epsilon_{ijk} a_j(x) (W_\mu)_k - \partial^\mu \epsilon_{ijk} a_j(x) (W^\nu)_k \partial_\nu (W_\mu)_i, \\
&\stackrel{\text{product rule}}{=} \mathcal{L}_{3\text{xSpin1}} - (\partial^\mu (W^\nu)_i) \epsilon_{ijk} (\partial_\mu a_j(x)) (W_\nu)_k - (\partial^\mu (W^\nu)_i) \epsilon_{ijk} a_j(x) (\partial_\mu (W_\nu)_k) \\
&\quad - \epsilon_{ijk} (\partial^\mu a_j(x)) (W^\nu)_k \partial_\mu (W_\nu)_i - \epsilon_{ijk} a_j(x) (\partial^\mu (W^\nu)_k) \partial_\mu (W_\nu)_i \\
&\quad + \partial^\mu (W^\nu)_i \epsilon_{ijk} (\partial_\nu a_j(x)) (W_\mu)_k + \partial^\mu (W^\nu)_i \epsilon_{ijk} a_j(x) (\partial_\nu (W_\mu)_k) \\
&\quad + \epsilon_{ijk} (\partial^\mu a_j(x)) (W^\nu)_k \partial_\nu (W_\mu)_i + \epsilon_{ijk} a_j(x) (\partial^\mu (W^\nu)_k) \partial_\nu (W_\mu)_i
\end{aligned} \tag{7.56}$$

which show that this is no internal symmetry of the Lagrangian.

However we can make this an internal symmetry by adding the term

$$-\epsilon_{ijk} (W_\mu)_j (W_\nu)_k$$

to the field tensor $W_{\mu\nu}$. We have then

$$(W_{\mu\nu})_i = \partial_\mu (W_\nu)_i - \partial_\nu (W_\mu)_i - \epsilon_{ijk} (W_\mu)_j (W_\nu)_k,$$

which gives us a new Lagrangian for the three free spin 1 fields

$$\begin{aligned}
\mathcal{L}_{3\text{xMaxwell}} &= \frac{1}{4} (W_{\mu\nu})_i (W^{\mu\nu})_i \\
&= \left(\partial_\mu (W_\nu)_i - \partial_\nu (W_\mu)_i - \epsilon_{ijk} (W_\mu)_j (W_\nu)_k \right) \times \\
&\quad \times \left(\partial^\mu (W^\nu)_i - \partial^\nu (W^\mu)_i - \epsilon_{ijk} (W^\mu)_j (W^\nu)_k \right)
\end{aligned} \tag{7.57}$$

You can check the invariance of this Lagrangian by yourself, but be warned the computation is quite long and tedious. We have therefore found the locally $SU(2)$ invariant Lagrangian, which we write

$$\mathcal{L}_{\text{D1+D2+Interaction+ } 3\text{xMaxwell}} = i\bar{\Psi} \gamma_\mu \partial^\mu \Psi + \bar{\Psi} \gamma_\mu \frac{\sigma_j}{2} W_j^\mu \Psi - \frac{1}{4} (W_{\mu\nu})_i (W^{\mu\nu})_i \tag{7.58}$$

7.3 Mass Terms and Unification of $SU(2)$ and $U(1)$

In the last chapter 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 particles²⁶ have mass and this is conventionally interpreted as the $SU(2)$ symmetry being **broken**. This means the symmetry exists at high energy and spontaneously breaks at lower energies.

²⁶ For example, the electron e^- and the electron-neutrino ν_e , described by Ψ and the three bosons, described by $(W^\mu)_i$.

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 :

$$\mathcal{L}_{\text{locally } U(1) \text{ invariant}} = -m\bar{\psi}\psi + \bar{\psi}\gamma_\mu(i\partial^\mu + gB_\mu)\psi - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} \quad (7.59)$$

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 is²⁷

$$\mathcal{L}_{\text{locally } SU(2) \text{ invariant}} = \bar{\Psi}\gamma_\mu(i\partial^\mu + g'\sigma_j W_j^\mu)\Psi - \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i$$

²⁷ The coupling constant for the three W_i^μ fields, we neglected so far for brevity, is here called g' .

with

$$(W_{\mu\nu})_i = \partial_\mu (W_\nu)_i - \partial_\nu (W_\mu)_i + \epsilon_{ijk}(W_\mu)_j (W_\nu)_k$$

and the doublet

$$\Psi := \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

As above, the three spin 1 fields $(W_\nu)_i$ are often called $SU(2)$ 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

$$\mathcal{L}_{SU(2) \text{ and } U(1)} = \bar{\Psi}\gamma_\mu(i\partial^\mu + gB^\mu + g'\sigma_j W_j^\mu)\Psi - \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \quad (7.60)$$

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 Lagrangian for a complex spin 0 field is given by, as we derived in Eq. 7.40

$$\mathcal{L}_{\text{spin } 0} = \frac{1}{2}(\partial_\mu \phi^\dagger \partial^\mu \phi - m^2 \phi^\dagger \phi). \quad (7.61)$$

We can add to this Lagrangian the next higher power in ϕ without violating any symmetry constraints²⁸. Thus we write, renaming the

²⁸ Recall that only higher order derivatives 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.

constants to their conventional names

$$\mathcal{L}_{\text{spin0+extraTerm}} = \partial_\mu \phi^\dagger \partial^\mu \phi + \rho^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2. \quad (7.62)$$

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:

$$\begin{aligned} \mathcal{L}_{\text{spin0+extraTerm+spin1Coupling}} &= \left((\partial_\mu - i\frac{1}{2}gB_\mu)\phi^\dagger \right) \left((\partial^\mu + i\frac{1}{2}gB^\mu)\phi \right) \\ &\quad + \rho^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \end{aligned} \quad (7.63)$$

²⁹ See Eq. 7.13 and recall that an overall constant in the Lagrangian has no influence.

with the symmetries²⁹

$$B_\mu \rightarrow B'_\mu = B_\mu + \partial_\mu a(x) \quad (7.64)$$

$$\phi(x) \rightarrow \phi'(x) = e^{ia(x)} \phi(x). \quad (7.65)$$

In the same way 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 fields³⁰ as

$$^{30} \Phi = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$$

$$\begin{aligned} \mathcal{L}_{SU(2) \text{ and } U(1)} &= \left((\partial_\mu - ig'\sigma_i(W_\mu)_i - i\frac{1}{2}gB_\mu)\Phi^\dagger \right) \left((\partial^\mu + ig'\sigma_i(W^\mu)_i + i\frac{1}{2}gB^\mu)\Phi \right) \\ &\quad + \underbrace{\rho^2 \Phi^\dagger \Phi - \lambda (\Phi^\dagger \Phi)^2}_{\equiv -V(\Phi)} \end{aligned} \quad (7.66)$$

with the doublet $\Phi := \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}$ and the symmetries (Eq. 7.55)

$$(W_\mu)_i \rightarrow (W'_\mu)_i = (W_\mu)_i + \partial_\mu b_i(x) + \epsilon_{ijk} b_j(x) (W_\mu)_k \quad (7.67)$$

and

$$\Phi \rightarrow \Phi' = e^{ib_i(x)\sigma_i} \Phi, \quad \bar{\Phi} \rightarrow \bar{\Phi}' = \bar{\Phi} e^{-ib_i(x)\sigma_i} \quad (7.68)$$

We start with this locally $SU(2)$ invariant Lagrangian and investigate in the following how this Lagrangian gives us mass terms for the fields W_i^μ and B_i^μ . Adding mass terms "by hand" to the Lagrangian does not work, because these terms spoil the symmetry and this leads to an insensible theory³¹.

The term we defined above

$$\begin{aligned} V(\Phi) &= -\rho^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \\ &= -\rho^2 \phi_1^\dagger \phi_1 + \lambda (\phi_1^\dagger \phi_1)^2 - \rho^2 \phi_2^\dagger \phi_2 + \lambda (\phi_2^\dagger \phi_2)^2 \\ &= V_1(\phi_1) + V_2(\phi_2) \end{aligned} \quad (7.69)$$

³¹ 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.

is often called Higgs potential. A two-dimensional plot for different values of ρ , with $\lambda > 0$ can be seen in Fig. 7.1.

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 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 minimum of the potential can be computed in the usual way

$$V(\phi) = -\rho^2|\phi|^2 + \lambda|\phi|^4 \quad (7.70)$$

$$\frac{\partial V(\phi)}{\partial \phi} = -2\rho^2|\phi| + 4\lambda|\phi|^3 \stackrel{!}{=} 0 \quad (7.71)$$

$$\rightarrow |\phi|(-2\rho^2 + 4\lambda|\phi|^2) \stackrel{!}{=} 0 \quad (7.72)$$

$$\rightarrow |\phi|^2 \stackrel{!}{=} \frac{\rho^2}{2\lambda} \quad (7.73)$$

$$\rightarrow |\phi| \stackrel{!}{=} \sqrt{\frac{\rho^2}{2\lambda}} \quad (7.74)$$

$$\phi_{\min} = \sqrt{\frac{\rho^2}{2\lambda}} e^{i\varphi}. \quad (7.75)$$

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 and maybe randomly, one new vacuum value is chosen out of the infinite possibilities.

From Eq. 7.69 we see that for the doublet, both components have this choice to make. We therefore have for the doublet the minimum

$$\Phi_{\min} = \begin{pmatrix} \phi_{1\min} \\ \phi_{2\min} \end{pmatrix} \quad (7.76)$$

An economical **choice**³² for the minimum is

$$\Phi_{\min} = \begin{pmatrix} 0 \\ \sqrt{\frac{\rho^2}{2\lambda}} \end{pmatrix} \equiv \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}, \quad (7.77)$$

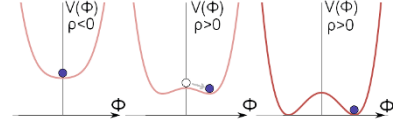


Fig. 7.1: Two-dimensional illustration of the Higgs potential for different values of ρ , which is believed to have changed as the universe cooled down as a result of the expansion of the universe. Figure adapted from "Spontaneous symmetry breaking" by FT2 (Wikimedia Commons) released under a CC BY-SA 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](http://commons.wikimedia.org/wiki/File:Spontaneous_symmetry_breaking_(explanatory_diagram).png), Accessed: 8.12.2014

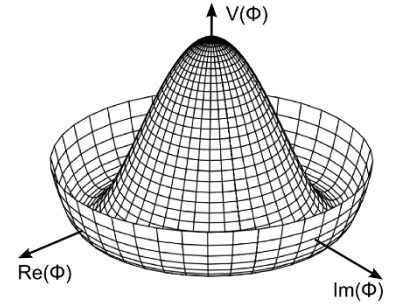


Fig. 7.2: 3-dimensional plot of the Higgs potential. Figure adapted from "Mexican hat potential polar" by Rupert Millard (Wikimedia Commons) released under a public domain licence. URL: http://commons.wikimedia.org/wiki/File:Mexican_hat_potential_polar.svg, Accessed: 7.5.2014

³² Recall that symmetry breaking means that **one** minimum is chosen out of the infinite possibilities.

where the factor $\frac{1}{2}$ is just a convention to make computations easier and we define for brevity $v \equiv \sqrt{\frac{\rho^2}{\lambda}}$. The next step is that we expand the field Φ around this minimum in order to learn something about its physical particle content. 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 shift the field to the new minimum. We therefore consider the field

$$\Phi = \begin{pmatrix} \phi_{1r} + i\phi_{1c} \\ \frac{v}{\sqrt{2}} + \phi_{2r} + i\phi_{2c} \end{pmatrix}. \quad (7.78)$$

³³ This form is very useful as we will see in a moment.

This can be rewritten as³³

$$\Phi = e^{i\theta_i \frac{\sigma_i}{2}} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}, \quad (7.79)$$

because if we consider the series expansion of the exponential function and the explicit form of the Pauli matrices σ_i , we can see that in first order

$$\begin{aligned} e^{i\theta_i \frac{\sigma_i}{2}} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} &\approx (1 + i\frac{1}{2}\theta_i \sigma_i) \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \\ &= (1 + i\frac{1}{2}\theta_1 \sigma_1 + i\frac{1}{2}\theta_2 \sigma_2 + i\frac{1}{2}\theta_3 \sigma_3) \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} 1 + i\frac{1}{2}\theta_3 & \frac{1}{2}\theta_1 - i\frac{1}{2}\theta_2 \\ \frac{1}{2}\theta_1 + i\frac{1}{2}\theta_2 & 1 - i\frac{1}{2}\theta_3 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \\ &= \begin{pmatrix} (\frac{1}{2}\theta_1 - i\frac{1}{2}\theta_2) \frac{v+h}{\sqrt{2}} \\ (1 - i\frac{1}{2}\theta_3) \frac{v+h}{\sqrt{2}} \end{pmatrix} \\ \text{redefinitions} \rightarrow &\equiv \begin{pmatrix} \phi_{1r} + i\phi_{1c} \\ \frac{v}{\sqrt{2}} + \phi_{2r} + i\phi_{2c} \end{pmatrix}. \end{aligned} \quad (7.80)$$

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 easiest³⁴.

A general local $SU(2)$ transformation is

$$\Phi \rightarrow \Phi' = e^{ib_i(x) \frac{\sigma_i}{2}} \Phi, \quad (7.81)$$

which enables us to eliminate the exponential factor in Eq. 7.79, by choosing appropriate $b_i(x)$. The complex scalar doublet is then, in

³⁴ 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.

this **unitary gauge**

$$\Phi_{un} = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}. \quad (7.82)$$

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 freedom³⁵. Therefore, these three fields aren't physical³⁶ 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. 7.63, and rename all constants to the standard choice

$$\begin{aligned} \mathcal{L} = & \left((\partial_\mu - ig' \frac{1}{2} \sigma_i (W_\mu)_i - i \frac{1}{2} g B_\mu) \Phi^\dagger \right) \left((\partial^\mu + ig' \sigma_i (W^\mu)_i + i \frac{1}{2} g B^\mu) \Phi \right) \\ & - V(\Phi) \end{aligned} \quad (7.83)$$

We substitute now the field Φ with the shifted field in the unitary gauge, which was defined in Eq. 7.82. 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 $\Phi \rightarrow \Phi_{\min} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}$, which means we ignore h , we get

$$\begin{aligned} & \left((\partial_\mu - ig' \sigma_i (W_\mu)_i - i \frac{1}{2} g B_\mu) \Phi_{\min}^\dagger \right) \left((\partial^\mu + ig' \frac{1}{2} \sigma_i (W^\mu)_i + i \frac{1}{2} g B^\mu) \Phi_{\min} \right) \\ &= \left| \left((\partial^\mu + ig' \frac{1}{2} \sigma_i (W^\mu)_i + i \frac{1}{2} g B^\mu) \Phi_{\min} \right) \right|^2 \\ &= \left| \left((\partial^\mu + ig' \frac{1}{2} \sigma_i (W^\mu)_i + i \frac{1}{2} g B^\mu) \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \right) \right|^2 \\ &= \frac{v^2}{8} \left| \left((g' \sigma_i (W^\mu)_i + g B^\mu) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) \right|^2. \end{aligned}$$

Now using that we have behind B^μ an implicit 2×2 identity matrix and the explicit form of the Pauli matrices³⁷ σ_i yields

³⁵ 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.

³⁶ 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.

³⁷ $\sigma_i W_i = \begin{pmatrix} W_3 & W_1 - iW_2 \\ W_1 + iW_2 & -W_3 \end{pmatrix}$

$$\begin{aligned}
&= \frac{v^2}{8} \left| \begin{pmatrix} g'W_3^\mu + gB^\mu & g'W_1^\mu - ig'W_2^\mu \\ g'W_1^\mu + ig'W_2^\mu & -g'W_3^\mu + gB^\mu \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right|^2 \\
&= \frac{v^2}{8} \left| \begin{pmatrix} g'W_1^\mu - ig'W_2^\mu \\ -g'W_3^\mu + gB^\mu \end{pmatrix} \right|^2 \\
&= \frac{v^2}{8} \left((g')^2 \left((W_1^\mu)^2 + (W_2^\mu)^2 \right) + (g'W_3^\mu - gB^\mu)^2 \right) \quad (7.84)
\end{aligned}$$

Next we define two new spin 1 fields from the old ones we have been using so far

$$W_+^\mu \equiv \frac{1}{\sqrt{2}}(W_1^\mu - iW_2^\mu) \quad (7.85)$$

$$W_-^\mu \equiv \frac{1}{\sqrt{2}}(W_1^\mu + iW_2^\mu), \quad (7.86)$$

where W_+^μ is the complex conjugate of W_-^μ . The first term in Eq. 7.84 is then

$$(W_1^\mu)^2 + (W_2^\mu)^2 = 2(W^+)_\mu (W^-)^\mu \quad (7.87)$$

and thus we have, including the constants,

$$\left(\underbrace{\frac{g'v}{2}}_{\equiv m_W} \right)^2 (W^+)_\mu (W^-)^\mu \quad (7.88)$$

which then looks like a typical "mass" term.

The second term in Eq. 7.84 can be written in matrix form

$$(g'W_3^\mu - gB^\mu)^2 = \begin{pmatrix} W_3^\mu & B^\mu \end{pmatrix} \underbrace{\begin{pmatrix} g'^2 & -gg' \\ -gg' & g'^2 \end{pmatrix}}_{\equiv G} \begin{pmatrix} W_3^\mu \\ B^\mu \end{pmatrix} \quad (7.89)$$

In order to be able to interpret this as mass-terms, we need to diagonalize³⁸ the matrix G . The standard linear-algebra way to do this needs the eigenvalues λ_1, λ_2 and normalized³⁹ eigenvectors v_1, v_2 of the matrix G , which are

$$\lambda_1 = 0 \rightarrow v_1 = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g \\ g' \end{pmatrix}$$

$$\lambda_2 = (g^2 + g'^2) \rightarrow v_2 = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g' \\ -g \end{pmatrix}$$

³⁸ 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.

³⁹ Which means length 1, i.e. $v \cdot v = 1$.

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

$$M = \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \quad (7.90)$$

and

$$G_{diag} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & (g^2 + g'^2) \end{pmatrix} \quad (7.91)$$

The matrix M is orthogonal ($M^T = M^{-1}$), because we work with normalized eigenvectors:

$$\begin{aligned} M^T M &= \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \\ &= \frac{1}{(g^2 + g'^2)} \begin{pmatrix} g^2 + g'^2 & gg' - gg' \\ gg' - gg' & g^2 + g'^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \end{aligned} \quad (7.92)$$

We therefore add two unit matrices $1 = M^T M$, into Eq. 7.89:

$$\left(W_3^\mu, B_\mu \right) \underbrace{MM^T}_{=1} G \underbrace{MM^T}_{=1} \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix} = \left(W_3^\mu, B_\mu \right) \underbrace{M M^T G M M^T}_{=G_{diag}} \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix}. \quad (7.93)$$

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:

$$\begin{aligned} M^T \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix} &= \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} g & g' \\ g' & -g \end{pmatrix} \begin{pmatrix} W_3^\mu \\ B_\mu \end{pmatrix} \\ &= \frac{1}{\sqrt{g^2 + g'^2}} \begin{pmatrix} (gW_3^\mu + g'B^\mu) \\ (g'W_3^\mu - gB^\mu) \end{pmatrix} \equiv \begin{pmatrix} A^\mu \\ Z_\mu \end{pmatrix} \end{aligned} \quad (7.94)$$

We can therefore write the second term as

$$\begin{aligned} \begin{pmatrix} A^\mu & Z_\mu \end{pmatrix} G_{diag} \begin{pmatrix} A^\mu \\ Z_\mu \end{pmatrix} &= \begin{pmatrix} A^\mu & Z_\mu \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & (g^2 + g'^2) \end{pmatrix} \begin{pmatrix} A^\mu \\ Z_\mu \end{pmatrix} \\ &= (g^2 + g'^2)(Z^\mu)^2 + 0 \cdot (A^\mu)^2. \end{aligned} \quad (7.95)$$

To summarize: We started with a Lagrangian, without mass terms for the spin 1 fields W_i^μ and B^μ

$$\left((\partial_\mu - iq' \frac{1}{2} \sigma_i (W_\mu)_i - i \frac{1}{2} q B_\mu) \Phi^\dagger \right) \left((\partial^\mu + iq' \sigma_i (W^\mu)_i + i \frac{1}{2} q B^\mu) \Phi \right). \quad (7.96)$$

After the process of spontaneous symmetry breaking, we have new terms in the Lagrangian that are interpreted as mass terms

$$\underbrace{\frac{1}{8}v^2 g'^2 (W^+)_\mu (W^-)^\mu}_{=\frac{1}{2}M_W^2} + \underbrace{\frac{1}{8}v^2 (g^2 + g'^2) Z_\mu^2}_{=\frac{1}{2}M_Z^2} + \underbrace{\frac{1}{8}v^2 0 \cdot A_\mu^2}_{\text{photon mass}=0} \quad (7.97)$$

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 massless⁴⁰. 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.

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** and we will discuss at the end of this section why this means that parity is violated. 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 neutrinos⁴¹. 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 helicity⁴².

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

⁴⁰ 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 Chap. 9.

⁴¹ 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 there should be a right-chiral neutrino component. Nevertheless, this right-chiral component does not participate in any known interaction.

⁴² 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

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)

$$\psi = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \quad (7.98)$$

and Ψ doublets of Dirac spinors

$$\Psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}. \quad (7.99)$$

Then this "something" is the projection operator P_L :

$$P_L \psi = P_L \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} = \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} \equiv \psi_L. \quad (7.100)$$

Such a projection operator can be constructed using the matrix⁴³

$$\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (7.101)$$

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 is then⁴⁴

$$P_L = \frac{1 - \gamma_5}{2} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad (7.102)$$

and we can define analogously

$$P_R = \frac{1 + \gamma_5}{2} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \quad (7.103)$$

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_μ^\pm and Z_μ with different fields. The corresponding terms were derived in Sec. 7.2, and the final result was Eq. 7.58, which we recite here for convenience:

$$\mathcal{L} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\sigma_j W_j^\mu\Psi - \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i. \quad (7.104)$$

⁴³ 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 .

⁴⁴ 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 Sec. 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 $P_L^{\text{Weyl}}\Psi^{\text{Weyl}} = \Psi_L^{\text{Weyl}} \Rightarrow P_L'\Psi' = \frac{1 - \gamma'_5}{2}\Psi' = \underbrace{U\Psi^{\text{Weyl}}}_{\Psi_L^{\text{Weyl}}} = \frac{1 - U i \gamma_0 U^{-1} U \gamma_1 U^{-1} U \gamma_2 U^{-1} U \gamma_3 U^{-1}}{2} U \Psi^{\text{Weyl}} = \frac{U - U i \gamma_0 \gamma_1 \gamma_2 \gamma_3}{2} \Psi^{\text{Weyl}} = U \left(\frac{1 - \gamma_5}{2} \right) \Psi^{\text{Weyl}} = U \Psi_L^{\text{Weyl}} = \Psi'_L \quad \checkmark$

The relevant term is $\bar{\Psi}\gamma_\mu\sigma_j W_j^\mu\Psi$ and we simply add P_L :

$$\rightarrow \mathcal{L} = i\bar{\Psi}\gamma_\mu\partial^\mu\Psi + \bar{\Psi}\gamma_\mu\sigma_j W_j^\mu P_L\Psi - \frac{1}{4}(W_{\mu\nu})_i(W^{\mu\nu})_i. \quad (7.105)$$

Here P_L acts on a doublet and is therefore

$$P_L\Psi = \begin{pmatrix} P_L & 0 \\ 0 & P_L \end{pmatrix} \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} = \begin{pmatrix} (\psi_1)_L \\ (\psi_2)_L \end{pmatrix}. \quad (7.106)$$

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 property⁴⁵. Projecting twice must be the same as projecting one time.
- $\{\gamma_5, \gamma_\mu\} = \gamma_5\gamma_\mu + \gamma_\mu\gamma_5 = 0$, which you can check by brute force computation⁴⁶.
- $(P_L)^\dagger = P_L$, because γ_5 is real, as can be seen from the explicit matrix form: $\gamma_5 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$

⁴⁵ 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 .

⁴⁶ 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 of $\gamma_5 = i\gamma_0\gamma_1\gamma_2\gamma_3$.

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

$$\gamma_\mu P_L = \gamma_\mu \frac{1 - \gamma_5}{2} = \frac{1 + \gamma_5}{2} \gamma_\mu = P_R \gamma_\mu. \quad (7.107)$$

We can now rewrite the relevant term of Eq. 7.105:

$$\begin{aligned} \bar{\Psi}\gamma_\mu\sigma_j W_j^\mu P_L\Psi &= \bar{\Psi}\gamma_\mu\sigma_j W_j^\mu (P_L)^2\Psi \\ &= \underbrace{\bar{\Psi}}_{\Psi^\dagger\gamma_0} \underbrace{\gamma_\mu P_L}_{P_R\gamma_\mu} \sigma_j W_j^\mu \underbrace{P_L\Psi}_{\Psi_L} \\ &= \Psi^\dagger \underbrace{\gamma_0 P_R}_{P_L\gamma_0} \gamma_\mu \sigma_j W_j^\mu \Psi_L \\ &= \underbrace{(P_L\Psi)^\dagger}_{\Psi_L^\dagger} \gamma_0 \gamma_\mu \sigma_j W_j^\mu \Psi_L \\ \text{Using } P_L^\dagger &= P_L \text{ and } (AB)^\dagger = ((AB)^T)^* = (B^T A^T)^* = B^\dagger A^\dagger \\ &= \underbrace{(P_L\Psi)^\dagger}_{\Psi_L^\dagger} \gamma_0 \gamma_\mu \sigma_j W_j^\mu \Psi_L \\ &= \bar{\Psi}_L \gamma_\mu \sigma_j W_j^\mu \Psi_L \quad \checkmark \end{aligned} \quad (7.108)$$

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 image⁴⁷. Here we need the parity operator for spinors⁴⁸ P_{spinor} and vectors⁴⁹ P_{vector} . The transformation yields

$$\begin{aligned} \underbrace{\bar{\Psi}}_{=\Psi^\dagger\gamma_0} \gamma_\mu \sigma_j (W_j^\mu) P_L \Psi &\rightarrow (P_{\text{spinor}} \Psi)^\dagger \gamma_0 \gamma_\mu \sigma_j (P_{\text{vector}} W^\mu)_j P_L (P_{\text{spinor}} \Psi) \\ &= (\Psi)^\dagger \gamma_0 \gamma_0 \gamma_\mu \sigma_j (P_{\text{vector}} W^\mu)_j P_L \gamma_0 \Psi \\ &\stackrel{\text{using } \{\gamma_5, \gamma_0\} = 0 \text{ and } P_L = \frac{1-\gamma_5}{2}}{=} (\Psi)^\dagger \gamma_0 \gamma_0 \gamma_\mu \gamma_0 \sigma_j (P_{\text{vector}} W^\mu)_j P_R \Psi. \end{aligned} \quad (7.109)$$

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:

$$(P_{\text{spinor}} \Psi)^\dagger \gamma_0 \gamma_\mu \sigma_j (P_{\text{vector}} W^\mu)_j P_L (P_{\text{spinor}} \Psi) = \bar{\Psi} \gamma_\mu \sigma_j W_j^\mu P_R \Psi \neq \bar{\Psi} \gamma_\mu \sigma_j W_j^\mu P_L \Psi \quad (7.110)$$

Therefore this term isn't 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 other⁵⁰. 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 $\bar{\Psi} \gamma_\mu \sigma_j W_j^\mu P_L \Psi = \bar{\Psi}_L \gamma_\mu \sigma_j W_j^\mu \Psi_L$. In physical terms this term means that the components of the left-chiral doublets, which means the two spin $\frac{1}{2}$ fields $(\psi_1)_L, (\psi_2)_L$, 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, as explained in Sec. 3.6.3. 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 $SU(2)$:

$$\Psi_L \rightarrow \Psi'_L = e^{i\vec{a} \cdot \frac{\vec{\sigma}}{2}} \Psi_L \quad (7.111)$$

⁴⁷ 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.

⁴⁸ The parity operator for spinors was derived in Sec. 3.7.9. Using the γ_μ matrices, we can write the parity operator derived there as $P = \gamma_0 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

⁴⁹ 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. 3.132.

⁵⁰ This is explained in appendix A.

- 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 $SU(2)$:

$$\begin{aligned}(\psi_1)_R &\rightarrow (\psi_1)'_R = e^0(\psi_1)_R = (\psi_1)_R \\(\psi_2)_R &\rightarrow (\psi_2)'_R = e^0(\psi_2)_R = (\psi_2)_R\end{aligned}\quad (7.112)$$

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 symmetry.

7.5 Lepton Mass Terms

At the beginning of Sec. 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:

$$I_1 := (\chi_a)^\dagger \zeta^{\dot{a}} = (\chi_L)^\dagger \zeta_R \quad \text{and} \quad I_2 := (\zeta^a)^T \chi_a = (\zeta_R)^\dagger \chi_L \quad (7.113)$$

We can write these invariants using Dirac spinors as⁵¹

$$\begin{aligned}\bar{\psi}\psi &= \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_L \\&= \begin{pmatrix} \chi_L^\dagger & 0 \end{pmatrix} \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \zeta_R \end{pmatrix} + \begin{pmatrix} 0 & \zeta_R^\dagger \end{pmatrix} \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ 0 \end{pmatrix} \\&= \chi_L^\dagger \zeta_R + \zeta_R^\dagger \chi_L \quad \checkmark\end{aligned}\quad (7.114)$$

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

$$\underbrace{\bar{\Psi}_L}_{\text{doublet}} \underbrace{\psi_R}_{\text{singlet}} \rightarrow \bar{\Psi}'_L \psi'_R = \bar{\Psi}_L e^{-ib_i \frac{\sigma_i}{2}} \psi_R \neq \bar{\Psi}_L \psi_R \quad (7.115)$$

⁵¹ 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 $\bar{\psi} = \psi^\dagger\gamma_0$.

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 a spin 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

$$\bar{\Psi}_L \Phi \psi_R. \quad (7.116)$$

To see the invariance we transform this term with a $SU(2)$ transformation⁵²

$$\bar{\Psi}^L \Phi \psi^R \rightarrow \bar{\Psi}'^L \Phi' \psi^R = \bar{\Psi}^L e^{-ib_i(x)\sigma_i} e^{ib_i(x)\sigma_i} \Phi \psi^R = \bar{\Psi}^L \Phi \psi^R \quad \checkmark$$

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 transformations⁵³ and therefore the term is Lorentz invariant, because we have the same Lorentz invariant terms as in Eq. 7.114.

This kind of term is called **Yukawa coupling** and we add it, with the equally allowed Hermitian conjugate to the Lagrangian, including a coupling constant⁵⁴ $-\lambda_2$

$$\mathcal{L} = -\lambda_2 (\bar{\Psi}^L \Phi \psi_2^R + \bar{\psi}_2^R \Phi \Psi^L). \quad (7.117)$$

This extra term does not only describe the interaction between the fermions and the Higgs field, but also leads to finite mass terms for the spin $\frac{1}{2}$ fields after the $SU(2)$ symmetry breaking. We put the expansion around the vacuum value, we chose (Eq. 7.77)

$$\Phi = \sqrt{\frac{1}{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix}$$

into the Lagrangian, which yields

$$\begin{aligned} \mathcal{L} &= -\frac{\lambda_2}{\sqrt{2}} \left((\bar{\Psi}_1^L, \bar{\Psi}_2^L) \begin{pmatrix} 0 \\ v+h \end{pmatrix} \psi_2^R + \bar{\psi}_2^R (0, v+h) \begin{pmatrix} \Psi_1^L \\ \Psi_2^L \end{pmatrix} \right) \\ &= -\frac{\lambda_2(v+h)}{\sqrt{2}} (\bar{\Psi}_2^L \psi_2^R + \bar{\psi}_2^R \Psi_2^L). \end{aligned}$$

Equation 7.114 tells us this is equivalent to

$$= -\frac{\lambda_2(v+h)}{\sqrt{2}} \bar{\psi}_2 \psi_2 \quad (7.118)$$

⁵² Remember $\Phi^L \rightarrow \Phi'^L = e^{ib_i(x)\sigma_i} \Phi^L$ and $\sigma_i^\dagger = \sigma_i$

⁵³ 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 Sec. 3.7.4.

⁵⁴ The strange name $-\lambda_2$ and why we are only adding ψ_2^R here will become clear in a moment, because terms including ψ_1^R and $-\lambda_1$ will be discussed afterwards.

$$= \underbrace{-\frac{\lambda_2 v}{\sqrt{2}}(\bar{\psi}_2 \psi_2)}_{\text{Fermion mass term}} - \underbrace{\frac{\lambda_f h}{\sqrt{2}}(\bar{\psi}_2 \psi_2)}_{\text{Fermion-Higgs interaction}}. \quad (7.119)$$

We see that we get indeed through the Higgs mechanism the required mass terms. Again, we used symmetry constraints to add a term to the Lagrangian, which 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-conjugated⁵⁵ Higgs field $\tilde{\Phi} = \epsilon\Phi^*$, because

⁵⁵ Charge conjugation is explained in Sec. 3.7.10.

$$\Phi = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} \rightarrow \tilde{\Phi} = \epsilon\Phi^* = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix}. \quad (7.120)$$

Following the same steps as above with the charge conjugated Higgs field leads to mass terms for Ψ_1 :

$$\begin{aligned} \mathcal{L} &= -\lambda_f(\bar{\Psi}^L \tilde{\Phi} \psi_1^R + \bar{\psi}_1^R \tilde{\Phi} \Psi^L). \\ &= -\frac{\lambda_1}{\sqrt{2}} \left((\bar{\Psi}_1^L, \bar{\Psi}_2^L) \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix} \Psi_1^R + \Psi_1^R \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix} \begin{pmatrix} \Psi_1^L \\ \Psi_2^L \end{pmatrix} \right) \\ &= -\frac{\lambda_1(v+h)}{\sqrt{2}} (\bar{\Psi}_1^L \Psi_1^R + \bar{\Psi}_1^R \Psi_1^L) \end{aligned}$$

To understand the rather abstract doublets better, we rewrite them more suggestively⁵⁶

⁵⁶ A neutrino is always denoted by a ν . In this step we simply give the two fields in the doublet ψ_1 and ψ_2 their conventional names: electron field e and electron-neutrino field ν_e .

$$\Psi = \begin{pmatrix} \nu_e \\ e \end{pmatrix} \quad (7.121)$$

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 a 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 $\bar{\Psi} \gamma_\mu \sigma_j W_j^\mu P_L \Psi$. As discussed in the last section this can be rewritten using the explicit matrix form of the Pauli matrices⁵⁷ σ_i , which then gives us terms coupling the components of the doublets together:

⁵⁷ This gives us once more

$$\sigma_i W_i^\mu = \begin{pmatrix} W_3^\mu & W_1^\mu - iW_2^\mu \\ W_1^\mu + iW_2^\mu & -W_3^\mu \end{pmatrix}$$

which we can rewrite using

$$W_\pm = \frac{1}{\sqrt{2}}(W_1 \mp iW_2):$$

$$\Rightarrow \sigma_i W_i^\mu = \begin{pmatrix} W_3^\mu & \sqrt{2}W_+^\mu \\ \sqrt{2}W_-^\mu & -W_3^\mu \end{pmatrix}$$

$$\begin{aligned}
\bar{\Psi} \gamma_\mu \sigma_j W_j^\mu P_L \Psi &= \begin{pmatrix} \bar{\nu}_e & \bar{e} \end{pmatrix} \gamma_\mu \begin{pmatrix} W_3^\mu & \sqrt{2}W_+ \\ \sqrt{2}W_- & -W_3^\mu \end{pmatrix} P_L \begin{pmatrix} \nu_e \\ e \end{pmatrix} \\
&\stackrel{\text{using Eq. 7.105}}{=} \begin{pmatrix} (\bar{\nu}_e)_L & (\bar{e})_L \end{pmatrix} \gamma_\mu \begin{pmatrix} W_3^\mu & \sqrt{2}W_+ \\ \sqrt{2}W_- & -W_3^\mu \end{pmatrix} \begin{pmatrix} (\nu_e)_L \\ (e)_L \end{pmatrix} \\
&= (\bar{\nu}_e)_L \gamma_\mu W_3^\mu (\nu_e)_L + (\bar{\nu}_e)_L \gamma_\mu \sqrt{2}W_+ (e)_L \\
&\quad + (\bar{e})_L \gamma_\mu \sqrt{2}W_- (\nu_e)_L - (\bar{e})_L \gamma_\mu W_3^\mu (e)_L. \tag{7.122}
\end{aligned}$$

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:

$$\bar{\Psi}_e \gamma_\mu \sigma_j W_j^\mu P_L \Psi_e + \bar{\Psi}_\mu \gamma_\mu \sigma_j W_j^\mu P_L \Psi_\mu + \bar{\Psi}_\tau \gamma_\mu \sigma_j W_j^\mu P_L \Psi_\tau, \tag{7.123}$$

which can be written more compact 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 mass terms read

$$\underbrace{-\frac{\lambda_l v}{\sqrt{2}}(\bar{l}l)}_{\text{Fermion mass term}} - \underbrace{\frac{\lambda_f h}{\sqrt{2}}(\bar{l}l)}_{\text{Fermion-Higgs interaction}}$$

and equivalently for the neutrinos.

This Lagrangian enables us to predict something about the Higgs field h that can be tested in experiments. For a given lepton, the mass is given by

$$m_l = \frac{\lambda_l v}{\sqrt{2}} \rightarrow \lambda_l = \frac{m_l \sqrt{2}}{v} \tag{7.124}$$

and the coupling strength of this lepton to the Higgs is given by

$$c_l = \frac{\lambda_l h}{\sqrt{2}} \stackrel{\text{Eq. 7.124}}{=} \frac{m_l \sqrt{2} h}{\sqrt{2} v} = \frac{m_l h}{v}. \tag{7.125}$$

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 Sec. 7.8, but first we want to talk about mass terms for quarks. Happily these can be incorporated analogously 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 quarks⁵⁸ these are the up- and down quark:

$$q = \begin{pmatrix} u \\ d \end{pmatrix} \quad (7.126)$$

and equally for the strange and charm or top and bottoms 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:

$$\underbrace{q_L}_{\text{doublet}} = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \rightarrow e^{ia_i \frac{\sigma_i}{2}} q_L \quad (7.127)$$

$$\begin{aligned} \underbrace{u_R}_{\text{singlet}} &\rightarrow u_R \\ \underbrace{d_R}_{\text{singlet}} &\rightarrow d_R. \end{aligned} \quad (7.128)$$

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

$$\bar{q}_L u_R + \bar{q}_L d_R + \bar{u}_R q_L + \bar{d}_R q_L, \quad (7.129)$$

which aren't $SU(2)$ invariant, we consider the coupling of the quarks to a spin 0 field doublet Φ :

⁵⁹ This is defined in Eq. 7.120:

$$\tilde{\Phi} = \epsilon \Phi^* \underbrace{=}_{\text{def}} \begin{pmatrix} \frac{v+h}{\sqrt{2}} \\ 0 \end{pmatrix}.$$

⁶⁰ 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 .

$$\lambda_u \bar{q}_L \tilde{\Phi} u_R + \lambda_d \bar{q}_L \Phi d_R + \lambda_u \bar{u}_R \tilde{\Phi} q_L + \lambda_d \bar{d}_R \Phi q_L, \quad (7.130)$$

with coupling constants λ_u, λ_d and the charge conjugated Higgs doublet⁵⁹ which is needed in order to get mass terms for the up quarks⁶⁰

Then, everything is analogous to the lepton case: We put the expansion of the Higgs field around its minimum⁶¹ into the Lagrangian, which gives us mass terms plus quark-Higgs coupling terms.

⁶¹ $\Phi = \begin{pmatrix} 0 \\ \frac{v+h}{\sqrt{2}} \end{pmatrix}$ and equivalently for the charge-conjugated Higgs field.

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 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 (Sec. 4.5.5, especially Eq. 4.56) tells us that

$$\partial_0 \int d^3x \underbrace{\frac{\partial \mathcal{L}}{\partial(\partial_0 \Psi)}}_{=Q} \delta \Psi = 0 \quad (7.131)$$

The Lagrangian is invariant under transformations of the form

$$\Psi \rightarrow e^{ia_i \frac{\sigma_i}{2}} \Psi = (1 + ia_i \frac{\sigma_i}{2} + \dots) \Psi \quad (7.132)$$

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. 7.45) is

$$\mathcal{L}_{D1+D2} = i\bar{\Psi} \gamma_\mu \partial^\mu \Psi.$$

The corresponding conserved quantities Q_i , for example for the electron-neutrino doublet, are⁶²

⁶² See Eq. 7.131 and as always defined without the arbitrary constants a_i .

$$\begin{aligned}
Q_i &= i\bar{\Psi}\gamma_0\frac{\sigma_i}{2}\Psi \\
&= \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \underbrace{\gamma_0\gamma_0}_{=1}\frac{\sigma_i}{2}\begin{pmatrix} v_e \\ e \end{pmatrix}.
\end{aligned} \tag{7.133}$$

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

$$\begin{aligned}
Q_3 &= \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \frac{\sigma_3}{2} \begin{pmatrix} v_e \\ e \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} v_e \\ e \end{pmatrix} \\
&= \frac{1}{2}v_e^\dagger v_e - \frac{1}{2}e^\dagger e
\end{aligned} \tag{7.134}$$

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

$$\begin{aligned}
Q_1 &= \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \frac{\sigma_1}{2} \begin{pmatrix} v_e \\ e \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} v_e \\ e \end{pmatrix}^\dagger \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} v_e \\ e \end{pmatrix} \\
&= \frac{1}{2}v_e^\dagger e + \frac{1}{2}e^\dagger v_e
\end{aligned} \tag{7.135}$$

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 Sec. 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 doublet⁶³ object.

⁶³ In a later section we will learn that the same can be done for triplets and the conserved quantities following from $SU(3)$ invariance.

A typical $SU(2)$ doublet is of the form

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}. \quad (7.136)$$

$SU(2)$ has exactly one Cartan generator $J_3 = \frac{1}{2}\sigma_3$, with eigenvalues $+\frac{1}{2}$ and $-\frac{1}{2}$. A (left-chiral) neutrino $\begin{pmatrix} \nu_e \\ 0 \end{pmatrix}$ 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.

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 zero⁶⁴: $J_i = 0$. 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

$$W^\mu = \begin{pmatrix} W_+^\mu \\ W_-^\mu \\ W_3^\mu \end{pmatrix}, \quad (7.137)$$

which transforms according to the three dimensional representation of $SU(2)$. In this representation the Cartan generator J_3 has eigenvalues⁶⁵ $+1, -1, 0$ 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.

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^μ .

If this is unclear, take 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 : $W^\mu = \sigma_i W_i^\mu = \sigma_1 W_1^\mu + \sigma_2 W_2^\mu + \sigma_3 W_3^\mu$ 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

⁶⁴ The right-chiral singlets do not transform at all as explained in Sec. 3.7.4.

⁶⁵ This can be seen directly from the explicit matrix form of J_3 in Eq. 3.121:

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

⁶⁶ The Lie algebra is a vector space!

⁶⁷ To be precise: A homomorphism, which is a map that satisfies some special conditions.

⁶⁸ A group itself is in general no vector space. Although we can take a look at how the group acts on itself, this is not a representation, but a realization of the group.

$$^{69} \vec{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

$$^{70} W^\mu = \begin{pmatrix} W_1^\mu \\ W_2^\mu \\ W_3^\mu \end{pmatrix}$$

in the Lie algebra⁶⁶ of $SU(2)$ and consequently our object W^μ lives there, too. Therefore, if we want to know how W^μ 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 map⁶⁷ from the group to the space of linear operators over a vector space. So far we only looked at "external" vector spaces like Minkowski space. The only⁶⁸ intrinsic vector space that comes with a group is its Lie algebra! 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 generators and therefore the adjoint representation is three dimensional. Exactly how we are able to write the components of a vector between two brackets⁶⁹, we can write the component of W^μ between two brackets⁷⁰, 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 describing strong interaction.

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.

$$U^\dagger U = U U^\dagger = 1 \quad \det U = 1. \quad (7.138)$$

As usual for Lie groups we can write this as an exponential function⁷¹

$$U = e^{iT_A \theta_A}. \quad (7.139)$$

The defining equations 7.138 of the group require, as for⁷² $SU(2)$, the generators to be Hermitian and traceless

$$T_A^\dagger = T_A \quad (7.140)$$

⁷¹ As already noted in Sec. 2.4, Capital Roman letters A, B, \dots are always summed from 1 to 8.

⁷² See 3.80 and the following text plus equations, where the basis was given by the 2×2 Pauli matrices.

$$\text{tr}(T_A) = 0 \quad (7.141)$$

A basis for those traceless, Hermitian generators is, at least in **one** representation, given by eight⁷³ 3×3 matrices, called Gell-Mann matrices:

$$\lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (7.142a)$$

$$\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \lambda_5 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \quad \lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (7.142b)$$

$$\lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}. \quad (7.142c)$$

The generators of the group are connected to these Gell-Mann matrices, just as the Pauli matrices were connected to the generators⁷⁴ of the $SU(2)$ group via $T_A = \frac{1}{2}\lambda_A$. The Lie algebra for this group is given by

$$[T_A, T_B] = if^{ABC}T^C, \quad (7.143)$$

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 yields⁷⁵

$$f^{123} = 1 \quad (7.144)$$

$$f^{147} = -f^{156} = f^{246} = f^{257} = f^{345} = -f^{367} = \frac{1}{2} \quad (7.145)$$

$$f^{458} = f^{678} = \frac{\sqrt{3}}{2}, \quad (7.146)$$

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

$$f^{ABC} = -f^{BAC} = -f^{CBA}. \quad (7.147)$$

All other possibilities, which cannot be computed by permutation, vanish.

⁷³ It can be shown that in general for $SU(N)$ there are $N^2 - 1$ generators. The number of generators is often called the **rank** of a group.

⁷⁴ $J_i = \frac{\sigma_i}{2}$, see Eq. 3.82 and the explanations there.

⁷⁵ This is not very enlightening, but we list it here for completeness.

Analogously to what we did for $SU(2)$ in Sec. 7.2, we introduce triplets of spin $\frac{1}{2}$ fields

$$Q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \quad (7.148)$$

and exactly as for $SU(2)$ we get new labels for the objects inside this triplet, which we will discuss in the next section.

To make the Lagrangian

$$\mathcal{L} = i\bar{Q}\partial_\mu\gamma^\mu Q - \bar{Q}mQ \quad (7.149)$$

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 Lagrangian⁷⁶

$$\mathcal{L} = -\frac{1}{4}F_{\alpha\beta}^A F_A^{\alpha\beta} + \bar{Q}(iD_\mu\gamma^\mu - m)Q, \quad (7.150)$$

and the field strength tensor $F_{\alpha\beta}^A$ for the spin 1 gluon field G_α^A is defined as

$$F_{\alpha\beta}^A = \partial_\alpha G_\beta^A - \partial_\beta G_\alpha^A - gf^{ABC}G_\alpha^B G_\beta^C \quad (7.151)$$

where f^{ABC} are the structure constants of $SU(3)$ that already appeared in the commutator of the generators. Furthermore, D_α is defined as

$$D_\alpha = \partial_\alpha + igT^C G_\alpha^C \quad (7.152)$$

where T^C are the generators of $SU(3)$ defined at the beginning of this section. 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, which is analogous to what we discussed for $SU(2)$ in Sec. 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 Cartan⁷⁷ 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.

⁷⁶ Remember: the sum over capital letters (A, B, C, \dots) runs from 1 to 8

⁷⁷ Recall, Cartan generators = diagonal generators.

⁷⁸ The eigenvectors are of course $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

The eigenvalues⁷⁸ 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 eigenvalues⁷⁹ are $\frac{1}{2\sqrt{3}}, \frac{1}{2\sqrt{3}}, \frac{-1}{\sqrt{3}}$

Therefore 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 :

$$\left(+\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \text{ for } \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \psi,$$

where one usually defines **red** := $(\frac{1}{2}, \frac{1}{2\sqrt{3}})$. This means something of

the form $\begin{pmatrix} \Psi \\ 0 \\ 0 \end{pmatrix}$ is called red.

Analogous

$$\left(-\frac{1}{2}, \frac{1}{2\sqrt{3}}\right) \text{ for } \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \psi$$

with **blue** := $(-\frac{1}{2}, \frac{1}{2\sqrt{3}})$ and analogously **green** := $(0, \frac{-1}{\sqrt{3}})$. The color idea comes from the fact that if we add the three colors, i.e.

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \psi,$$

we get a state with charge zero (a colorless state), because

$$\lambda_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0 \quad \text{and} \quad \lambda_8 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 0,$$

which is analogous to sunlight that contains all colors of light, but is colorless, nonetheless.

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-dimensional⁸⁰) adjoint representation of $SU(3)$ to assign color to the gauge fields $G_{A'}^\mu$, i.e. the gluons, completely analogous to how we assigned isospin to the W-Bosons in Sec. 7.7.

7.8.2 Quark Description

Recall that spin $\frac{1}{2}$ particles⁸¹, which interact via the strong force are

⁷⁹ Corresponding to the same eigenvectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

⁸⁰ The adjoint representation of $SU(3)$ is 8 dimensional, because we have 8 generators.

⁸¹ Which are created by spin $\frac{1}{2}$ fields, as we will learn in Chap. 6.

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, quarks are $SU(2)$ doublets, because they interact via the weak force, too. Each object in this doublet⁸² (=each quark) is a triplet: $q = \begin{pmatrix} u_c \\ d_c \end{pmatrix}$. This can become very confusing, 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 is again a two component object $u_c = \begin{pmatrix} (\chi_u^L)_c \\ (\xi_u^R)_c \end{pmatrix}$. The upper components 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

$$\bar{Q}mQ \quad (7.153)$$

is $SU(3)$ invariant, as long as all particles in the triplet have equal mass. This means m is proportional to the unit matrix⁸³. 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

$$U = \begin{pmatrix} u_r \\ u_b \\ u_g \end{pmatrix}, \quad (7.154)$$

where u_r denotes a red, u_b a blue and u_g green up-quark, which all have the same mass.

⁸² Each quark doublet consists of two different quarks, for example an up- and a down-quark or a top- and a bottom-quark.

⁸³ $m = m \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, instead of $m = m \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}$

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.

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 Sec. 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 Chap. 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^μ, G_2^μ, \dots . The final Lagrangian describes correctly strong interactions. $SU(3)$ symmetry tells us that color is conserved.

Different 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

*"There is nothing truly beautiful but that which can never be of any use
whatsoever; everything useful is ugly."*

Theophile Gautier
in *Mademoiselle de Maupin*.
Wildside Press, 11 2007 .
ISBN 9781434495556

8

Quantum Mechanics

Summary

In this chapter we will talk about quantum mechanics. The foundation for everything here are the identifications we discussed in Chap. 5. As a first result, we derive the **relativistic energy-momentum relation**.

After discussing 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 very useful for our understanding of the structure of quantum mechanics. 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 multiplication 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 then $|A|^2$. 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 take a look at one of the examples, we already solved using wave-mechanics, in the Dirac notation.

8.1 Particle Theory Identifications

¹ The Klein-Gordon, Dirac, Proka and Maxwell equations

² See Eq. 3.240, Eq. 3.244 and Chap. 5

The equations¹ we derived so far 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 energy and the momentum of the particle or particles in question. As explained in Chap. 5, we identify these with the generators of the corresponding symmetry²

- 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 Sec. 6.2 we derived the equation of motion for a free spin 0 field, the Klein-Gordon equation:

$$(\partial_\mu\partial^\mu + m^2)\Phi = 0.$$

With the identifications reiterated above this equation reads³

$$\begin{aligned} (\partial_\mu\partial^\mu + m^2)\Phi &= (\partial_0\partial_0 - \partial_i\partial_i + m^2)\Phi \\ &= \left(\left(\frac{1}{i}E \right) \left(\frac{1}{i}E \right) - \left(-\frac{1}{i}p_i \right) \left(-\frac{1}{i}p_i \right) + m^2 \right) \Phi \\ &= (-E^2 + \vec{p}^2 + m^2)\Phi = 0 \end{aligned} \quad (8.1)$$

$$\rightarrow E^2 = \vec{p}^2 + m^2 \quad \text{or using four-vectors} \quad p_\mu p^\mu = m^2 \quad (8.2)$$

which is the famous **energy-momentum relation** of special-relativity. For a particle at rest, i.e. $\vec{p} = 0$ this gives us Einstein's famous equation

$$E^2 = m^2 \rightarrow E = mc^2$$

where we restored c^2 for clarity. We can now understand why we gave the scalar value of the first Casimir operator of the Poincare group $p_\mu p^\mu$ in Eq. 3.258 the suggestive name m^2 . 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 Sec. 6.2 is conventionally called m^2 .

$$\text{}^3 \text{ Using } p_\mu = \begin{pmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{pmatrix} = \begin{pmatrix} p_0 \\ \vec{p} \end{pmatrix} = \begin{pmatrix} E \\ \vec{p} \end{pmatrix}$$

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 eigenfunction equation for the momentum operator we must solve is

$$\underbrace{-i\partial_i}_{\text{operator}} \Psi = \underbrace{p_i}_{\text{eigenvalue}} \underbrace{\Psi}_{\text{eigenfunction}}, \quad (8.3)$$

with some number p_i . A solution is

$$\underbrace{C}_{=const} e^{ip_i x_i} \quad \text{because} \\ \rightarrow -i\partial_i C e^{ip_i x_i} = p_i C e^{ip_i x_i} \quad \checkmark \quad (8.4)$$

but take note that this is 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

$$i\partial_0 \Phi = E\Phi \quad (8.5)$$

or angular momentum eigenfunctions⁴. Analogous to eigenvectors for matrices, these eigenfunctions are bases⁵. This means we can expand an arbitrary function Ψ in terms of eigenfunctions. For example, in terms of momentum eigenfunctions⁶ (for brevity in one dimension)

$$\Psi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dp \Psi_p e^{-ipx}, \quad (8.6)$$

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 energy eigenfunctions Φ_{E_n} :

$$\Psi = \sum_n c_n \Phi_{E_n}. \quad (8.7)$$

Take note that, in general, a set of eigenfunctions for one operator is not a set of eigenfunctions for another operator. Only for operators that commute $[A, B] = AB - BA = 0$, we can find a simultaneous

⁴ The discussion for angular momentum eigenfunctions is a bit more complicated, which can be seen, 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, which is introduced in appendix D.1 and the factor $\frac{1}{\sqrt{2\pi}}$ is a matter of convention.

set of eigenfunctions for both operators. To see this, assume we have $[C, D] \neq 0 \rightarrow CD \neq DC$. Then for some eigenfunction Ψ of C , we have $C\Psi = c\Psi$, where c is the corresponding eigenvalue. If Ψ could be at the same time an eigenfunction of D : $D\Psi = d\Psi$, we have:

$$\begin{aligned} CD\Psi &= C d\Psi = \underbrace{dC\Psi}_{\text{because } d \text{ is just a number}} = dc\Psi \\ DC\Psi &= Dc\Psi = cD\Psi = c d\Psi = \underbrace{cd\Psi}_{\text{because numbers commute}} = dc\Psi \\ \rightarrow DC &= CD \quad \text{which is contradictory to } [C, D] \neq 0 \end{aligned} \quad (8.8)$$

⁷ 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.

⁸ This means all other coefficients in the expansion $\Psi = \sum_n c_n \Phi_{E_n}$ are zero.

In general our operators act on something we call⁷ Ψ , which denotes the **state** of the physical system in question. We get this Ψ , by solving the corresponding equation of motion.

In general, 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 eigenstates⁸ $\Psi = c_1\Phi_{E_1} + c_2\Phi_{E_2}$. Acting with the energy operator on this state yields

$$\hat{E}\Psi = \hat{E}(c_1\Phi_{E_1} + c_2\Phi_{E_2}) = c_1E_1\Phi_{E_1} + c_2E_2\Phi_{E_2} \neq E(c_1\Phi_{E_1} + c_2\Phi_{E_2}). \quad (8.9)$$

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 relevant⁹.

Secondly, observe that a solution to any equation of motion we derived so far is a function of¹⁰ \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 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

$$\int dx \Psi^*(x, t) \Psi(x, t) \stackrel{!}{=} 1, \quad (8.10)$$

which is called the normalization prescription, because the probability for finding the particle anywhere in space must be 100% = 1.

⁹ 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?

¹⁰ This will be made explicit in the next sections.

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 in terms of energy eigenfunctions

$\Psi = c_1\Phi_{E_1} + c_2\Phi_{E_2} + \dots$. 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}

$$P(E_1) = \left| \int dx \Phi_{E_1}^*(x, t) \Psi(x, t) \right|^2$$

In the example above this means

$$\begin{aligned} P(E_1) &= \left| \int dx \Phi_{E_1}^*(x, t) \Psi(x, t) \right|^2 = \left| \int dx \Phi_{E_1}^*(x, t) (c_1\Phi_{E_1} + c_2\Phi_{E_2}) \right|^2 \\ &= \underbrace{\left| c_1 \int dx \Phi_{E_1}^*(x, t) \Phi_{E_1} \right|^2}_{=1 \text{ as explained above}} + \underbrace{\left| c_2 \int dx \Phi_{E_1}^*(x, t) \Phi_{E_2} \right|^2}_{=0 \text{ because eigenstates are orthogonal}} \\ &= |c_1|^2 \end{aligned} \quad (8.11)$$

Analogously, if we can 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 collecting 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

$$\langle x \rangle = \sum_i \rho_i x_i \quad (8.12)$$

where ρ_i denotes the probability. Equally for a continuous distribution we have

$$\langle x \rangle = \int dx \rho(x) x \quad (8.13)$$

In quantum mechanics the expectation value for a physical quantity \hat{O} is defined analogously

$$\langle \hat{O} \rangle = \int d^3x \Psi^* \hat{O} \Psi. \quad (8.14)$$

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, we have the expectation value for the location of some particle

$$\langle \hat{x} \rangle = \int d^3x \Psi^* \hat{x} \Psi = \int d^3x \Psi^* x \Psi = \int d^3x x \underbrace{\Psi^* \Psi}_{\text{probability density of its location}}. \quad (8.15)$$

We take now, for computational simplicity, the non-relativistic limit of the Klein-Gordon equation.

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

$$\begin{aligned} 0 &= (\partial_\mu \partial^\mu + m^2) \Phi \\ &= (\partial_\mu \partial^\mu + m^2) e^{\pm i p_\mu x^\mu} \\ &= (i^2 p_\mu p^\mu + m^2) e^{\pm i p_\mu x^\mu} = 0 \\ &= (-m^2 + m^2) e^{\pm i p_\mu x^\mu} = 0 \quad \checkmark \end{aligned} \quad (8.16)$$

We can write one of the solutions a little differently

$$\Phi = e^{-i p_\mu x^\mu} = \Phi = e^{i(-Et + \vec{x} \cdot \vec{p})}$$

and then we see that the dependence is given by e^{-iEt} , i.e. $\Phi \propto e^{-iEt}$. From Eq. 8.2 we know the energy is

$$E = \sqrt{\vec{p}^2 + m^2} = \sqrt{m^2 \left(\frac{\vec{p}^2}{m^2} + 1 \right)} = m \sqrt{\frac{\vec{p}^2}{m^2} + 1}.$$

In the non-relativistic limit $|\vec{p}| \ll m$, which means our object moves much slower than the speed of light and therefore its momentum is much smaller than its mass, we can approximate the energy using the Taylor-series as

$$E = m(1 + \frac{1}{2} \frac{\vec{p}^2}{m^2} + \dots)$$

$$\rightarrow E \approx \underbrace{m}_{\text{rest-mass}} + \underbrace{\frac{\vec{p}^2}{2m}}_{\text{kinetic energy}}.$$

We can now write

$$\Phi = e^{i(-Et + \vec{x} \cdot \vec{p})} \approx e^{-imt} \underbrace{e^{i\vec{p} \cdot \vec{x} - i(\vec{p}^2/2m)t}}_{\equiv \phi(\vec{x}, t)} = e^{-imt} \phi(\vec{x}, t). \quad (8.17)$$

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 this ansatz into the Klein-Gordon equation we get

$$(\partial_\mu \partial^\mu + m^2) e^{-imt} \phi(\vec{x}, t) = (\partial_0 \partial^0 + \partial_i \partial^i + m^2) e^{-imt} \phi(\vec{x}, t) = 0.$$

Then we use $\partial_t e^{-imt}(\dots) = e^{-imt}(-im + \partial_t)(\dots)$, 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

$$\begin{aligned} (\partial_t)^2 \phi(\vec{x}, t) &= (\partial_t)^2 \exp[i\vec{p} \cdot \vec{x} - i(\vec{p}^2/2m)t] \\ &= (\vec{p}^2/2m)^2 \exp[i\vec{p} \cdot \vec{x} - i(\vec{p}^2/2m)t] \propto \frac{p^4}{m^2} \end{aligned} \quad (8.18)$$

with the second term

$$\begin{aligned} im\partial_t \phi(\vec{x}, t) &= im\partial_t \exp[i\vec{p} \cdot \vec{x} - i(\vec{p}^2/2m)t] \\ &= m(\vec{p}^2/2m) \exp[i\vec{p} \cdot \vec{x} - i(\vec{p}^2/2m)t] \propto p^2 \end{aligned} \quad (8.19)$$

shows us that, because of $|\vec{p}| \ll m$, we can neglect the third term in this limit, thus

$$\begin{aligned} &(-2im\partial_t + \nabla^2) \phi(\vec{x}, t) = 0 \\ &\underbrace{\rightarrow}_{\text{dividing by } (-2m)} (i\partial_t - \frac{1}{2m} \nabla^2) \phi(\vec{x}, t) = 0 \end{aligned}$$

$$\rightarrow (i\partial_t - \frac{\nabla^2}{2m})\phi(\vec{x}, t) = 0, \quad (8.20)$$

which is the famous **Schrödinger equation**. If we now make the identifications we reiterated at the beginning of this chapter, the equation reads

$$\begin{aligned} \rightarrow (E + \underbrace{\frac{\vec{p}^2}{2m}}_{=\text{kinetic energy}})\phi(\vec{x}, t) &= 0 \\ \rightarrow E &= \frac{\vec{p}^2}{2m} \end{aligned} \quad (8.21)$$

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

$$i\partial_t\phi(\vec{x}, t) = \underbrace{\frac{\nabla^2}{2m}}_{\equiv \hat{H}}\phi(\vec{x}, t) \rightarrow i\partial_t\phi(\vec{x}, t) = \hat{H}\phi(\vec{x}, t). \quad (8.22)$$

Following a similar procedure, it is possible to derive the non-relativistic limit of the Dirac equation, which is known as Pauli equation.

8.4.1 Schrödinger Equation with External Field

In addition, we can follow the same route and derive the non-relativistic limit of the **interacting** Klein-Gordon equation (Eq. 7.43), i.e. the equation that describes the interaction between a massive spin 0 field and a massless spin 1 field (the photon field). The resulting equation is

$$\left(i\partial_t - \frac{1}{2m} \left(\nabla - ia\vec{A}\right)^2 + q\Phi\right)\phi(\vec{x}, t) = 0 \quad (8.23)$$

8.5 From Wave Equations to Particle Motion

Now, let's look at two examples of how the standard interpretation of quantum mechanics works.

8.5.1 Example: Free Particle

A solution for the free (without external potential) Schrödinger equation (Eq. 8.20) is given by

$$\Psi = e^{-i(Et - \vec{p}\vec{x})} \quad (8.24)$$

because

$$\begin{aligned} i\partial_t e^{-i(Et - \vec{p}\vec{x})} &= -\frac{\nabla^2}{2m} e^{-i(Et - \vec{p}\vec{x})} \\ \rightarrow E e^{-i(Et - \vec{p}\vec{x})} &= \frac{\vec{p}^2}{2m} e^{-i(Et - \vec{p}\vec{x})} \end{aligned} \quad (8.25)$$

where E is just the numerical value for the total energy of the free particle, which was derived in Eq. 8.21 as $E = \frac{\vec{p}^2}{2m}$. A more general solution can be composed by linear combination

$$\Psi = A e^{-i(Et - \vec{p}\vec{x})} + B e^{-i(E't + \vec{p}'\vec{x})} + \dots$$

We interpret the wave-function as a probability amplitude and therefore the wave-function, describing a particle, must be normalized, because the total probability for finding the particle must be 100% = 1. This is not possible for the wave-function above, spreading out over all of space. To describe an individual free particle we have to use a suited linear combination, called a wave-packet:

$$\Psi_{WP}(x, t) = \int dp A(p) e^{i(\vec{p}\vec{x} - Et)}, \quad (8.26)$$

where the complex numbers $A(p)$ have to be chosen in a way that makes the wave packet normalizable. One possibility is a Gaussian wave-packet, where $A(p)$ is a Gauss distribution.

$$\Psi_{GWP}(x, t) = \int dp A(p) e^{i(\vec{p}\vec{x} - Et)} = \int dp \psi_0 e^{i(\vec{p} - \vec{p}_0)^2 / 4\sigma^2} e^{i(\vec{p}\vec{x} - Et)}$$

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).

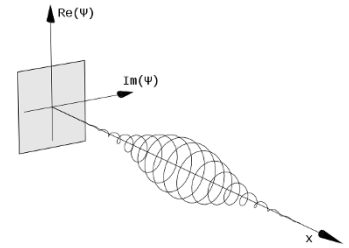


Fig. 8.1: Free wave-packet with Gaussian envelope. Figure by Inductiveload (Wikimedia Commons) released under a public domain licence. URL: http://commons.wikimedia.org/wiki/File:Travelling_Particle_Wavepacket.svg, Accessed: 4.5.2014

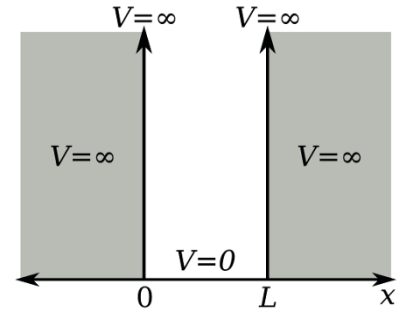


Fig. 8.2: Infinite potential well. Figure by Benjamin D. Esham (Wikimedia Commons) released under a public domain licence. URL: http://commons.wikimedia.org/wiki/File:Infinite_potential_well.svg, Accessed: 4.5.2014

The potential is defined piece-wise

$$V = \begin{cases} 0, & 0 < x < L \\ \infty, & \text{otherwise} \end{cases} \quad (8.27)$$

and therefore, we have to solve the one-dimensional Schrödinger equation piece-wise.

$$i\partial_t \Psi(\vec{x}, t) = -\frac{\partial_x^2}{2m} \Psi(x, t) + V(x) \Psi(x, t)$$

- Inside the box, the solution is equal to the free particle solution, because $V = 0$ for $0 < x < L$
- Outside, because $V = \infty$, the only possible, physical solution is $\Psi(x, t) = 0$.

¹¹ Using $\sin(x) = \frac{1}{2i}(e^{ix} - e^{-ix})$ and $\cos(x) = \frac{1}{2}(e^{ix} + e^{-ix})$, which follows directly from the series expansion of $\cos(x)$, $\sin(x)$ and e^{ix} as derived in appendix B.4.1.

We can rewrite the general free particle solution¹¹

$$\begin{aligned} \Psi(x, t) &= Ae^{-i(Et - \vec{p} \cdot \vec{x})} + Be^{-i(Et + \vec{p} \cdot \vec{x})} \\ &= (C \sin(\vec{p} \cdot \vec{x}) + D \cos(\vec{p} \cdot \vec{x}))e^{-iEt}, \end{aligned}$$

which we can rewrite again using the non-relativistic energy-momentum relation, which was derived in Eq. 8.21

$$E = \frac{\vec{p}^2}{2m} \quad \rightarrow \quad \vec{p} = \sqrt{2mE}$$

$$\Psi(x, t) = (C \sin(\sqrt{2mE}x) + D \cos(\sqrt{2mE}x))e^{-iEt} \quad (8.28)$$

Next we use that the wave-function must be a continuous function¹². Therefore, we have the boundary conditions $\Psi(0) = \Psi(L) \stackrel{!}{=} 0$. We see that, because $\cos(0) = 1$ we have $D \stackrel{!}{=} 0$. Furthermore, we see that these conditions impose

$$\sqrt{2mE} \stackrel{!}{=} \frac{n\pi}{L}, \quad (8.29)$$

with arbitrary integer n , because for¹³

$$\Phi_n(x, t) = C \sin\left(\frac{n\pi}{L}x\right)e^{-iE_n t} \quad (8.30)$$

both boundary conditions are satisfied

$$\rightarrow \Phi_n(L, t) = C \sin\left(\frac{n\pi}{L}L\right)e^{-iE_n t} = C \sin(n\pi)e^{-iE_n t} = 0 \quad \checkmark$$

$$\rightarrow \Phi_n(0, t) = C \sin\left(\frac{n\pi}{L}0\right)e^{-iE_n t} = C \sin(0)e^{-iE_n t} = 0 \quad \checkmark$$

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

¹² 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 jumping point would be infinite.

¹³ Take note that we put an index n to our wave-function, because we have a different solution for each n .

be 100% = 1 and the probability outside is zero, because there we have $\Psi = 0$. Therefore

$$\begin{aligned}
 P &= \int_0^L dx \Phi_n^*(x, t) \Phi_n(x, t) \stackrel{!}{=} 1 \\
 P &= \int_0^L dx C^2 \sin\left(\frac{n\pi}{L}x\right) e^{+iEt} \sin\left(\frac{n\pi}{L}x\right) e^{-iEt} \\
 &= C^2 \int_0^L dx \sin^2\left(\frac{n\pi}{L}x\right) = C^2 \left[\frac{x}{2} - \frac{\sin\left(\frac{2n\pi}{L}x\right)}{4\frac{n\pi}{L}} \right]_0^L \\
 &= C^2 \left(\frac{L}{2} - \frac{\sin\left(\frac{2n\pi}{L}L\right)}{4\frac{n\pi}{L}} \right) = C^2 \frac{L}{2} \stackrel{!}{=} 1 \\
 &\rightarrow C^2 \stackrel{!}{=} \frac{2}{L} \quad \checkmark
 \end{aligned}$$

We can now solve Eq. 8.29 for the energy E

$$E_n \stackrel{!}{=} \frac{n^2 \pi^2}{L^2 2m}. \quad (8.31)$$

The possible energies are quantized, which means that the corresponding quantity can only be integer multiplies of some constant, here $\frac{\pi^2}{L^2 2m}$. 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 interpretation¹⁴.

Next we can ask¹⁵, 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

$$\Psi(x, t) = \sqrt{\frac{3}{5}}\Phi_2(x, t) + \sqrt{\frac{2}{5}}\Phi_3(x, t).$$

The answer in the conventional interpretation of quantum mechanics is: It is the absolute value squared of the overlap between Ψ and Φ_2

$$P(E = \frac{2^2 \pi^2}{L^2 2m}) = \left| \int dx \Phi_2^*(x, t) \Psi(x, t) \right|^2$$

where the overlap can be seen as a scalar product¹⁶ of Φ_2 and Ψ :

$$(\Phi_2, \Psi) = \int dx \Phi_2^* \Psi = \underbrace{c}_{\text{complex number}}.$$

¹⁴ A probability of more than 1 = 100% doesn't make sense

¹⁵ Instead of what's the probability to find a particle at place x .

¹⁶ In fact, this is the scalar product of the Hilbert space in which our state vectors Ψ, Φ_n live.

The computation is easy, because the solutions we found are orthogonal, i.e.

$$\int dx \Phi_n^*(x, t) \Phi_{n'}(x, t) = \delta_{nn'}$$

¹⁷ You can check this easily using integration by parts or something like Wolframalpha.com

For example¹⁷,

$$\begin{aligned} \int_0^L dx \Phi_2^* \Phi_3(x, t) &= \int_0^L dx C \sin\left(\frac{2\pi}{L}x\right) e^{+iEt} C \sin\left(\frac{3\pi}{L}x\right) e^{-iEt} \\ &= C^2 \int_0^L dx \sin\left(\frac{2\pi}{L}x\right) \sin\left(\frac{3\pi}{L}x\right) = 0. \end{aligned}$$

Therefore, we get the probability for finding the energy $E = \frac{2^2\pi^2}{L^2 2m}$

$$\begin{aligned} P(E = \frac{2^2\pi^2}{L^2 2m}) &= \left| \int dx \Phi_2^*(x, t) \Psi(x, t) \right|^2 \\ &= \left| \int dx \Phi_2^*(x, t) \left(\sqrt{\frac{3}{5}} \Phi_2(x, t) + \sqrt{\frac{2}{5}} \Phi_3(x, t) \right) \right|^2 \\ &= \left| \int dx \left(\underbrace{\sqrt{\frac{3}{5}} \Phi_2^*(x, t) \Phi_2(x, t)}_{=1 \text{ if integrated}} + \underbrace{\sqrt{\frac{2}{5}} \Phi_2^*(x, t) \Phi_3(x, t)}_{=0 \text{ if integrated}} \right) \right|^2 \\ &= \left(\sqrt{\frac{3}{5}} \right)^2 \end{aligned} \quad (8.32)$$

Take note that we are able to call the functions we just found in Eq. 8.30 eigenstates of the energy operator $i\partial_t$ or equivalently of the Hamiltonian¹⁸ $\hat{H} \equiv -\frac{\partial_x^2}{2m}$, because

¹⁸ This follows directly from the Schrödinger equation

$$i\partial_t \Phi = -\frac{\partial_x^2}{2m} \Phi \equiv \hat{H} \Phi$$

$$\hat{H} \Phi_n = E_n \Phi_n. \quad (8.33)$$

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

$$\hat{H} \Psi = \hat{H} \left(\sqrt{\frac{3}{5}} \Phi_2 + \sqrt{\frac{2}{5}} \Phi_3 \right) \underset{\text{Eq. 8.33}}{=} \sqrt{\frac{3}{5}} E_2 \Phi_2 + \sqrt{\frac{2}{5}} E_3 \Phi_3$$

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 is of great benefit in understanding the structure of quantum mechanics.

8.5.3 Dirac Notation

In the Dirac notation the state of a physical system is denoted abstractly by

$$|\Psi\rangle, \quad (8.34)$$

which is called a **ket**¹⁹. 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

$$\langle\Psi|^\dagger = |\Psi\rangle, \quad (8.35)$$

where the \dagger (called dagger) symbol 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

$$(|\Phi\rangle, |\Psi\rangle) \equiv \langle\Phi|\Psi\rangle$$

If a ket is multiplied with a **bra** from the left-hand side, the result is a complex number

$$\langle\Phi|\Psi\rangle = c \quad (8.36)$$

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, because we are using a probabilistic interpretation, our states must fulfil a normalization condition²⁰. For example, we have

$$\int dx |\Psi(x, t)|^2 = \int dx \Psi(x, t)^\dagger \Psi(x, t) \stackrel{!}{=} 1 \quad (8.37)$$

¹⁹ The pun will become clear in a second.

²⁰ The probability for finding the particle anywhere must be $100\% = 1$ or equally the probability for finding the particle with any momentum must be $100\% = 1$. In other words: The sum of probabilities for all possible outcomes must add up to 1.

and equally

$$\int dp |\Phi(p, t)|^2 = \int dp \Phi(p, t)^\dagger \Phi(p, t) \stackrel{!}{=} 1 \quad (8.38)$$

or written in our new notation

$$\int dx |\langle x | \Psi \rangle|^2 = \int dx (\langle x | \Psi \rangle)^\dagger \langle x | \Psi \rangle = \int dx \langle \Psi | x \rangle \langle x | \Psi \rangle \stackrel{!}{=} 1 \quad (8.39)$$

and

$$\int dp |\langle p | \Phi \rangle|^2 = \int dp (\langle p | \Phi \rangle)^\dagger \langle p | \Phi \rangle = \int dp \langle \Phi | p \rangle \langle p | \Phi \rangle \stackrel{!}{=} 1 \quad (8.40)$$

where we can see a new kind of operator: $|p\rangle \langle p|$ and $|x\rangle \langle x|$, which are called **projection operators**²¹. They are operators because they transform a ket into another ket. For example,

$$|x\rangle \langle x | \Psi \rangle = |x\rangle \Psi(x),$$

=some complex number we name $\Psi(x)$

which is again a ket, because the product of a complex number with a ket is again a ket. In general, an operator is any object that acts on a ket to generate another ket. We can now, by looking at Eq. 8.39, introduce another operator

$$\int dx \langle \Psi | x \rangle \langle x | \Psi \rangle = \langle \Psi | \underbrace{\left(\int dx |x\rangle \langle x| \right)}_{\equiv \hat{I}} | \Psi \rangle = \langle \Psi | \hat{I} | \Psi \rangle \stackrel{!}{=} 1. \quad (8.41)$$

From this we can conclude

$$\hat{I} | \Psi \rangle \stackrel{!}{=} | \Psi \rangle \quad (8.42)$$

for an arbitrary ket $| \Psi \rangle$, because $\langle \Psi | \Psi \rangle = 1$. This follows, because the probability for a system prepared in state $| \Psi \rangle$ to be found in $| \Psi \rangle$ 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 results

$$\int dx |x\rangle \langle x| = \hat{I} \quad (8.43)$$

or for a discrete basis

$$\sum_i |i\rangle \langle i| = \hat{I} \quad (8.44)$$

are called **completeness relations**. In general, we say the component of a ket $|a\rangle$ in the basis $|i\rangle$ is

$$|i\rangle^\dagger |a\rangle \equiv \langle i | a \rangle \equiv a_i,$$

²¹ Exactly like the projection operators for left-chiral and right-chiral spinors we introduced earlier, these projection operators fulfil the defining condition $p^2 = p$.

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 expansion²² 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

$$|\Psi\rangle = \int dx |x\rangle \underbrace{\langle x|\Psi\rangle}_{\equiv \Psi(x) \text{ complex number}} = \int dx |x\rangle \Psi(x).$$

The expectation value we introduced in Sec. 8.3.1 is in the Dirac notation given by

$$\langle \hat{O} \rangle = \langle \Psi | \hat{O} | \Psi \rangle$$

Let's return to the example of the particle in a box, which we can now solve 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

$$\begin{aligned} P(E_2) &= \langle E = \frac{n^2 \pi^2}{2mL^2} | \Psi \rangle = \langle E = \frac{n^2 \pi^2}{2mL^2} | \underbrace{\left(\int dx |x\rangle \langle x| \right)}_{=\hat{I}} | \Psi \rangle \\ &= \int dx \langle E = \frac{n^2 \pi^2}{2mL^2} | x \rangle \langle x | \Psi \rangle = \int dx \Phi_2^*(x) \Psi(x) \end{aligned}$$

which is exactly the result we derived using the standard wave mechanics in Sec. 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 Sec. 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 one part of the conserved quantity that we called spin, too²³. The operator we derived

²² 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.

²³ See 4.5.4

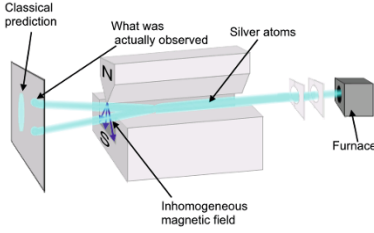


Fig. 8.3: Illustration of the Stern-Gerlach experiment. The original experiment was performed with silver atoms, whose spin behaviour is dominated by the one electron in the outermost atomic orbital. The experimental result is the same as with electrons. A beam of particles is affected by an inhomogeneous 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.

²⁴ 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

in Sec. 5.1.1 gives us, when acting on a state, the spin of the particle. For the scalar representation, this operator is given by $\hat{S} = 0$, which of course always yields 0 when acting 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 Sec. 3.7.5

$$\hat{S}_i = \frac{\sigma_i}{2} \quad (8.45)$$

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

$$\hat{S}_3 = \frac{\sigma_3}{2} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \quad (8.46)$$

The corresponding eigenstates are

$$v_{\frac{1}{2}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad v_{-\frac{1}{2}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (8.47)$$

with eigenvalues $\frac{1}{2}$ and $-\frac{1}{2}$, respectively. This means a particle described by a spinor²⁴ 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. In the quantum framework this is interpreted that a measurement of spin can only result in $\frac{1}{2}$ and $-\frac{1}{2}$. In Sec. 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

$$\hat{S}_z = \frac{1}{2}\sigma_3 = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}. \quad (8.48)$$

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. If we did no such filtering, the coefficients are $a = b = \frac{1}{\sqrt{2}}$, which means probability²⁵ $\frac{1}{2}$ for each possibility. Things get really 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 v 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

$${}_z\langle -\frac{1}{2} || X \rangle = a {}_z\langle -\frac{1}{2} || \frac{1}{2} \rangle_z + b {}_z\langle -\frac{1}{2} || \frac{-1}{2} \rangle_z = b \quad (8.49)$$

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)

$$S_x = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \quad (8.50)$$

and the corresponding normalized eigenstates are $|\frac{1}{2}\rangle_x \triangleq \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $|\frac{-1}{2}\rangle_x \triangleq \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. 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 $|\frac{-1}{2}\rangle_x$:

$$\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_z = \frac{1}{\sqrt{2}} \left(\underbrace{\begin{pmatrix} 1 \\ 2 \end{pmatrix}_x}_{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}} + \underbrace{\begin{pmatrix} -1 \\ 2 \end{pmatrix}_x}_{\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}} \right) \quad (8.51)$$

²⁵ 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.

$$\underbrace{\left|-\frac{1}{2}\right\rangle_z}_{\begin{pmatrix} 0 \\ 1 \end{pmatrix}} = \frac{1}{\sqrt{2}} \left(\underbrace{\left|\frac{1}{2}\right\rangle_x}_{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}} - \underbrace{\left|-\frac{1}{2}\right\rangle_x}_{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}} \right) \quad (8.52)$$

And therefore

$$|X\rangle = a\left|\frac{1}{2}\right\rangle_z + b\left|-\frac{1}{2}\right\rangle_z = a\frac{1}{\sqrt{2}}\left(\left|\frac{1}{2}\right\rangle_x + \left|-\frac{1}{2}\right\rangle_x\right) + b\frac{1}{\sqrt{2}}\left(\left|\frac{1}{2}\right\rangle_x - \left|-\frac{1}{2}\right\rangle_x\right). \quad (8.53)$$

The probability amplitude for measuring $-\frac{1}{2}$ along the x -axis is then

$$\begin{aligned} {}_x\langle -\frac{1}{2}|X\rangle &= {}_x\langle -\frac{1}{2}| \left(a\frac{1}{\sqrt{2}}\left(\left|\frac{1}{2}\right\rangle_x + \left|-\frac{1}{2}\right\rangle_x\right) + b\frac{1}{\sqrt{2}}\left(\left|\frac{1}{2}\right\rangle_x - \left|-\frac{1}{2}\right\rangle_x\right) \right) \\ &= \frac{a}{\sqrt{2}} - \frac{b}{\sqrt{2}} \end{aligned} \quad (8.54)$$

and the probability is $P_{x=-\frac{1}{2}} = \left|\frac{a}{\sqrt{2}} - \frac{b}{\sqrt{2}}\right|^2$.

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

$$|X\rangle_{\text{after } z\text{-axis filtering}} = \left|\frac{1}{2}\right\rangle_z \quad (8.55)$$

This means $a = 1$ and $b = 0$ and we get a non-zero probability for measuring spin $-\frac{1}{2}$ along the x -axis $P_{x=-\frac{1}{2}} = \left|\frac{1}{\sqrt{2}} - \frac{0}{\sqrt{2}}\right|^2 = \frac{1}{2}$. 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

$$|X\rangle_{\text{after } x\text{-axis filtering}} = \left|\frac{1}{2}\right\rangle_x. \quad (8.56)$$

If we want to know the probability for measuring spin $-\frac{1}{2}$ along the z -axis, we need to write $\left|\frac{1}{2}\right\rangle_x$ in terms of $\left|\frac{1}{2}\right\rangle_z$ and $\left|-\frac{1}{2}\right\rangle_z$:

$$\underbrace{|X\rangle_{\text{after } x\text{-axis filtering}}}_{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}} = \underbrace{\left|\frac{1}{2}\right\rangle_x}_{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}} = \frac{1}{\sqrt{2}} \left(\underbrace{\left|\frac{1}{2}\right\rangle_z}_{\begin{pmatrix} 1 \\ 0 \end{pmatrix}} + \underbrace{\left|-\frac{1}{2}\right\rangle_z}_{\begin{pmatrix} 0 \\ 1 \end{pmatrix}} \right) \quad (8.57)$$

and we get the probability $P_{z=-\frac{1}{2}} = |\langle -\frac{1}{2}|_z|X\rangle|^2 = \frac{1}{2}$. To summarize, this means:

- We start with a measurement of spin along the z-axis and filter out all particles with $-\frac{1}{2}$. This leaves us with a state

$$|X\rangle_{\text{after z-axis filtering}} = \left|\frac{1}{2}\right\rangle_z \quad (8.58)$$

- 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%.

$$\left\langle \frac{1}{2} \right|_z |X\rangle_{\text{after z-axis filtering}} = 1 \quad (8.59)$$

$$\left\langle -\frac{1}{2} \right|_z |X\rangle_{\text{after z-axis filtering}} = 0 \quad (8.60)$$

- If we now 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 $|X\rangle_{\text{after x-axis filtering}} = \left|\frac{1}{2}\right\rangle_x$.
- 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 Lectures²⁶ Vol. 3.

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 behaviour back to the fact that²⁷ $\hat{S}_x \hat{S}_z \neq \hat{S}_z \hat{S}_x$. 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.

²⁶ Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics, Volume 3*. Addison Wesley, 1st edition, 1971. ISBN 9780201021189

²⁷ Recall that we identify the spin operators with the corresponding finite-dimensional representations for the rotation generators. These fulfil the commutator relation $[\hat{J}_i, \hat{J}_j] = \hat{J}_i \hat{J}_j - \hat{J}_j \hat{J}_i = i\epsilon_{ijk} \hat{J}_k \neq 0 \rightarrow \hat{J}_i \hat{J}_j \neq \hat{J}_j \hat{J}_i$. For example, if we describe spin $\frac{1}{2}$ particles, we must use the two-dimensional representation $\hat{J}_i = \frac{\sigma_i}{2}$.

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 behaviour for measurements of position and momentum. Take a look again at Eq. 5.3, which we recite here for convenience:

$$[\hat{p}_i, \hat{x}_j] = \hat{p}_i \hat{x}_j - \hat{x}_j \hat{p}_i = i\delta_{ij}. \quad (8.61)$$

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. Take note that only for measurements along the same axis is the commutator non-zero²⁸. A measurement of momentum in the y-direction has no influence on what we can expect for the position on the x-axis. In other words this means that we can't know momentum and position in the same direction at the same time with arbitrary precision.

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 symmetry as measurement operators. For instance, this has the consequence that a measurement of momentum is equivalent to the action of the translation generator²⁹. 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.

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 formalism³⁰ is, in terms of results, completely equivalent to the wave mechanics, we described in this chapter. But computations in this formalism

²⁸ The Kronecker delta δ_{ij} is zero for $i \neq j$ and one for $i = j$ as defined in appendix B.5.5.

²⁹ Recall: Invariance under translations leads us to conservation of momentum.

³⁰ 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 edition, 7 2010. ISBN 9780486477220

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 story³¹

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 putting the ansatz: $Re^{S/\hbar}$ 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.

Further Reading Tips

- **Richard P. Feynman - The Feynman Lectures on Physics, Vol. 3**³² 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**³³ is a very readable and enlightening book.
- **J. J. Sakurai - Modern Quantum Mechanics**³⁴ is a brilliant book, which often offers a unique perspective on the concepts of quantum mechanics.
- **Paul A. M. Dirac - Lectures on Quantum Mechanics**³⁵ 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 very enlightening.

³¹ Harry Woolf, editor. *Some Strangeness in the Proportion*. Addison-Wesley, 1st edition, 2 1981. ISBN 9780201099249

³² Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *The Feynman Lectures on Physics, Volume 3*. Addison Wesley, 1st edition, 1 1971. ISBN 9780201021189

³³ David J. Griffiths. *Introduction to Quantum Mechanics*. Pearson Prentice Hall, 2nd edition, 4 2004. ISBN 9780131118928

³⁴ J. J. Sakurai. *Modern Quantum Mechanics*. Addison Wesley, 1st edition, 9 1993. ISBN 9780201539295

³⁵ Paul A. M. Dirac and Physics. *Lectures on Quantum Mechanics*. Dover Publications, 1st edition, 3 2001. ISBN 9780486417134

- 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. $u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$.

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,

³⁶ They are labelled by different quantum numbers and therefore behave differently in experiments!

but the crucial point is that these are really distinct particles/fields³⁶, 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.

- The two components of each Weyl spinor describe different spin configurations³⁷ 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**

³⁷ This was discussed in Sec. 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**.

- Any other Weyl spinor is simply a mixture of 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 Sec. 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 Sec. 8.9. Here we just use the results. There are four independent solutions of the Dirac equation and two are of the form

$$\psi_i = \begin{pmatrix} u_i \\ u_i \end{pmatrix} \quad (8.62)$$

with for example³⁸ $u_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} e^{-imt}$ and $u_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-imt}$ and two solutions are of the form

$$\tilde{\psi}_i = \begin{pmatrix} -v_i \\ v_i \end{pmatrix}, \quad (8.63)$$

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 configurations³⁹.

- 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

³⁸ This is a basis choice. The only requirement is that they are linearly independent.

³⁹ 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: $\psi_1^c = i\gamma_2 \psi_1^*$. Therefore: $(\psi_1)^c = \tilde{\psi}_2$ and $(\psi_2)^c = \tilde{\psi}_1$.

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

$$\text{physical electron} = \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix} \propto \begin{pmatrix} u \\ u \end{pmatrix}. \quad (8.64)$$

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

$$\text{physical positron} = \begin{pmatrix} -\xi_L \\ \chi_R \end{pmatrix} \propto \begin{pmatrix} -v \\ v \end{pmatrix} \quad (8.65)$$

which we can see through charge conjugation⁴⁰

$$\begin{aligned} (\text{physical electron})^c &= i\gamma_2 \begin{pmatrix} \chi_L \\ \xi_R \end{pmatrix}^* = \begin{pmatrix} -\xi_L \\ \chi_R \end{pmatrix} \propto i\gamma_2 \begin{pmatrix} u \\ u \end{pmatrix}^* = \begin{pmatrix} -u^c \\ u^c \end{pmatrix} \\ &= \text{physical positron} \end{aligned} \quad (8.66)$$

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

$$e_L^\uparrow = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (8.67)$$

⁴⁰ See Sec. 7.1.5 and use the explicit form of the matrix as defined in Eq. 6.13: $\gamma_2 = \begin{pmatrix} 0 & \bar{\sigma}_2 \\ \sigma_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix}$.

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.

$$e_L^\uparrow = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{2} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right) = \Psi_1(t=0) - \tilde{\Psi}_1(t=0) \quad (8.68)$$

We know how Ψ_1 and $\tilde{\Psi}_1$ evolve in time

$$\Psi_1(t) - \tilde{\Psi}_1(t) \Rightarrow \frac{1}{2} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{-imt} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{imt} \right) \quad (8.69)$$

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

$$\rightarrow \frac{1}{2} \left(\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \underbrace{e^{-i\frac{\pi}{2}}}_{=-i} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \underbrace{e^{i\frac{\pi}{2}}}_{=i} \right) = \frac{i}{2} \left(\begin{pmatrix} -1 \\ 0 \\ -1 \\ 0 \end{pmatrix} - \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \end{pmatrix} \right) = -i \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = -ie_R^\uparrow \quad (8.70)$$

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 \mathbf{u} and \mathbf{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 Chap. 6, when we talk about quantum field theory. The Dirac equation is

$$(i\partial_\mu \gamma^\mu - m)\psi = 0. \quad (8.71)$$

Anticipating plane wave solutions, we make the ansatz

$\Psi = ue^{-ipx}$, 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 $-ipx = -i(p_0x_0 - \vec{p}\vec{x}) = -ip_0x_0$. Now using the relativistic energy-momentum relation $E = \sqrt{\vec{p}^2 + m^2}$, which we derived at the beginning of this chapter, and using that $p_0 = E$ and $x_0 = t$, we have $-ipx = -iEt = -i\sqrt{\underbrace{\vec{p}^2}_{=0} + m^2}t = -imt$. Putting this ansatz into the Dirac equation yields

$$\begin{aligned} (i\partial_\mu \gamma^\mu - m)ue^{-imt} &= 0 \\ \rightarrow (i(\partial_0 \gamma^0 - \partial_i \gamma_i) - m)ue^{-imt} &= 0 \\ \rightarrow i((-im)\gamma^0 - m)ue^{-imt} &= 0 \\ \rightarrow (m\gamma^0 - m)u &= 0 \\ \xrightarrow{\text{dividing by } m} \left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) u &= 0 \\ \rightarrow \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= 0 \\ \rightarrow \begin{pmatrix} -u_1 + u_2 \\ u_1 - u_2 \end{pmatrix} &= 0 \end{aligned} \quad (8.72)$$

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

$$\Psi_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{-imt} \quad \Psi_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} e^{-imt} \quad (8.73)$$

We can find two other solutions by making the ansatz $\tilde{\Psi} = v e^{ipx}$, which analogously reduces in the rest frame to $\tilde{\Psi} = v e^{imt}$. This ansatz yields

$$\begin{aligned} (i\partial_\mu \gamma^\mu - m) v e^{imt} &= 0 \\ \rightarrow (-m\gamma^0 - m)v &= 0 \\ \rightarrow \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= 0 \\ \rightarrow \begin{pmatrix} -v_1 - v_2 \\ -v_1 - v_2 \end{pmatrix} &= 0. \end{aligned} \quad (8.74)$$

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

$$\tilde{\Psi}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix} e^{imt} \quad \tilde{\Psi}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} e^{imt}. \quad (8.75)$$

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

$$\partial_\mu \bar{\psi} \gamma_\mu \psi = \partial_\mu \bar{\psi} \underbrace{N^{-1}N}_{=1} \gamma_\mu \underbrace{N^{-1}N}_{=1} \psi = \partial_\mu \underbrace{\bar{\psi} N^{-1}}_{\equiv \bar{\psi}'} \underbrace{N \gamma_\mu N^{-1}}_{\equiv \gamma'_\mu} \underbrace{N \psi}_{\equiv \psi'} = \partial_\mu \bar{\psi}' \gamma'_\mu \psi'. \quad (8.76)$$

The basis we worked with in this text so far is called the **chiral/ Weyl basis**. Conventionally the Dirac equation is solved in another basis, called **mass/Dirac basis**. In the chiral/Weyl basis we worked with so far, the Dirac Lagrangian

$$\mathcal{L}_D = i\chi_L^\dagger \sigma^\mu \partial_\mu \chi_L + i\bar{\chi}_R^\dagger \bar{\sigma}^\mu \partial_\mu \bar{\chi}_R - m\chi_L^\dagger \bar{\chi}_R - m\bar{\chi}_R^\dagger \chi_L \quad (8.77)$$

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

$$\bar{\psi}' M' \psi' = \psi^\dagger \gamma_0' M' \psi' = \begin{pmatrix} u' \\ v' \end{pmatrix}^\dagger \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix} = (u')^\dagger m_1 u' + (v')^\dagger m_2 v', \quad (8.78)$$

whereas at the moment we are dealing with

$$\bar{\psi} M \psi = \psi^\dagger \gamma_0 M \psi = \begin{pmatrix} \chi_L \\ \bar{\chi}_R \end{pmatrix}^\dagger \begin{pmatrix} 0 & m \\ m & 0 \end{pmatrix} \begin{pmatrix} \chi_L \\ \bar{\chi}_R \end{pmatrix} = m\chi_L^\dagger \bar{\chi}_R + m\bar{\chi}_R^\dagger \chi_L. \quad (8.79)$$

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 $N = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$, which means

$$N^{-1} \underbrace{\begin{pmatrix} -m & 0 \\ 0 & m \end{pmatrix}}_{\equiv M'} N = M \quad (8.80)$$

$$\rightarrow m \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} = m \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (8.81)$$

and therefore we redefine the Dirac spinors accordingly

$$\begin{aligned}
\bar{\psi} M \psi &= \bar{\psi} \underbrace{N N^{-1}}_{=1} M \underbrace{N N^{-1}}_{=1} \psi \\
&= \underbrace{\bar{\psi} N}_{\equiv \bar{\psi}'} \underbrace{N^{-1} M N}_{\equiv M'} \underbrace{N^{-1} \psi}_{\equiv \psi'} \\
&= \bar{\psi}' M' \psi'.
\end{aligned} \tag{8.82}$$

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 find the corresponding γ_5 matrix, which is

$$\begin{aligned}
\tilde{\gamma}_5 &= N^{-1} \gamma_5 N = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\
&= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \\
&= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\end{aligned} \tag{8.83}$$

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

$$P_L = \frac{1-\gamma_5}{2} = \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{8.84}$$

8.10.1 Solutions of the Dirac Equation in the Mass Basis

We can solve the Dirac equation in the mass basis

$$(i\gamma_\mu \partial^\mu - m)\Psi = 0 \tag{8.85}$$

by making the ansatz $\psi = u e^{-ipx}$, which yields

$$(\gamma_\mu p^\mu - m)u e^{-ipx} = 0$$

$$\rightarrow (\gamma_\mu p^\mu - m)u = 0$$

Equivalently we can make the ansatz $\psi = ve^{ipx}$, which yields

$$(-\gamma_\mu p^\mu - m)ve^{ipx} = 0$$

$$\rightarrow (-\gamma_\mu p^\mu - m)v = 0.$$

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 we have, because $p_i = 0$

$$\rightarrow (\gamma_0 p^0 - m)u = 0$$

$$\rightarrow (-\gamma_0 p^0 - m)v = 0.$$

In addition, $p_0 = E$ 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/Dirac basis, which can be computed using the matrix N from above and $\gamma'_0 = N^{-1}\gamma_0 N$. Remember that we have an implicit unit matrix behind m and therefore

$$\begin{aligned} &\rightarrow \left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} m - m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) u = 0 \\ &\rightarrow \left(- \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} m - m \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right) v = 0. \\ &\rightarrow \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} u = 0 \\ &\rightarrow \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} v = 0. \end{aligned}$$

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:

$$\begin{aligned} &\rightarrow \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 0 \\ -2u_2 \end{pmatrix} = 0 \rightarrow u_2 = 0 \\ &\rightarrow \begin{pmatrix} -2 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -2v_1 \\ 0 \end{pmatrix} = 0 \rightarrow v_1 = 0. \end{aligned}$$

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

$$\Psi'_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} e^{-imt} \quad \Psi'_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} e^{-imt} \quad (8.86)$$

and

$$\tilde{\Psi}'_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} e^{imt} \quad \tilde{\Psi}'_2 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} e^{imt}. \quad (8.87)$$

A general solution in this frame, in this basis, is a linear combination

$$\psi = u e^{-ipx} + v e^{ipx} = \begin{pmatrix} u_1 \\ 0 \end{pmatrix} e^{-ipx} + \begin{pmatrix} 0 \\ v_1 \end{pmatrix} e^{ipx} \quad (8.88)$$

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 configurations⁴¹

$$\Psi = \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{E_p}} \left(c_r(p) u_r(p) e^{-ipx} + d_r^\dagger(p) v_r(p) e^{+ipx} \right) \quad (8.89)$$

⁴¹ Recall that the two components of a Weyl spinor represent different spin states.

9

Quantum Field Theory

Summary

In this chapter the framework of quantum field theory is introduced. Starting with the equation derived in Chap. 5

$$[\Phi(x), \pi(y)] = i\delta(x - y),$$

we are able to see that the fields itself are operators. The solutions to 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.

¹ The idea behind the Fourier transform is explained in appendix D.1.

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, which are denoted in the Dirac notation by

$$\langle f | \hat{S} | i \rangle ,$$

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

$$\hat{S}(t, t_i) = e^{-i \int_{t_i}^t dt' H_I}.$$

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 fields 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 $e^-e^+ \rightarrow \gamma$. This means we start with an initial state consisting of an electron and a positron $|e^-e^+\rangle$, which are annihilated by the field operators of the spin $\frac{1}{2}$ fields and afterwards a photon $|\gamma\rangle$ 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 $e^-e^+ \rightarrow \gamma \rightarrow e^-e^+$, 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 we evaluated above.

9.1 Field Theory Identifications

In this section, we want to understand how the Lagrangians 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 Chap. 5, specifically Eq. 5.5, which we recite here for convenience:

$$[\Phi(x), \pi(y)] = i\delta(x - y), \quad (9.1)$$

where the conjugate momentum density $\pi(y)$ is given by

$$\pi(y) = \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi(y))}. \quad (9.2)$$

9.2 Free Spin 0 Field Theory

"Every act of creation is first of all an act of destruction."

- Pablo Picasso²

Again, let's start with the simplest possible case: free spin 0 fields, described by scalars, which are objects that do not transform at all under Lorentz transformations, as derived in Sec. 3.7.4. We already derived in Chap. 6.2 the corresponding Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \quad (9.3)$$

and the equation of motion, called Klein-Gordon equation

$$(\partial_\mu \partial^\mu + m^2)\Phi = 0. \quad (9.4)$$

Computing the conjugate momentum is straightforward

$$\pi(x) = \frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi(x))} = \frac{\partial}{\partial(\partial_0 \Phi(x))} \frac{1}{2}(\partial_\mu \Phi(x) \partial^\mu \Phi(x) - m^2 \Phi^2(x)) = \partial_0 \Phi(x)$$

The most general solution of the Klein-Gordon equation can be written in terms of its Fourier-expansion³

$$\Phi(x) = \int dk^3 \frac{1}{(2\pi)^3 2\omega_k} \left(a(k) e^{-i(kx)} + b(k) e^{i(kx)} \right), \quad (9.5)$$

with $(\omega_k)^2 \equiv \vec{k}^2 - m^2$, which we can write, if we restrict to real scalar fields as

$$\Phi(x) = \int dk^3 \frac{1}{(2\pi)^3 2\omega_k} \left(a(k) e^{-i(kx)} + a^\dagger(k) e^{i(kx)} \right) \quad (9.6)$$

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 taking 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 compute⁴

$$[a(k), a^\dagger(k')] = (2\pi)^3 \delta^3(\vec{k} - \vec{k}') \quad (9.7)$$

and

$$[a(k), a(k')] = 0 \quad (9.8)$$

² As quoted in Rollo May. *The Courage to Create*. W. W. Norton and Company, reprint edition, 3 1994. ISBN 9780393311068

³ See the appendix in Sec. 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.

⁴ See for example chapter 4.1 in Lewis H. Ryder. *Quantum Field Theory*. Cambridge University Press, 2nd edition, 6 1996. ISBN 9780521478144

$$[a^\dagger(k), a^\dagger(k')] = 0 \quad (9.9)$$

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 coefficients $a(k)$ and $a^\dagger(k)$ act on something abstract and by doing this we, of course, learn how the fields act on something abstract. To build 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, which we derived in Eq. 4.40 from time-translation invariance

$$\begin{aligned} E &= \int d^3x T^{00} \\ &= \int d^3x \left(\frac{\partial \mathcal{L}}{\partial(\partial_0 \Phi)} \underbrace{\frac{\partial \Phi}{\partial x_0}}_{=\partial_0 \Phi} - \mathcal{L} \right) \\ &= \int d^3x \left((\partial_0 \Phi)^2 - \frac{1}{2} (\partial_\mu \Phi \partial^\mu \Phi - m^2 \Phi^2) \right) \\ &\stackrel{\substack{= \\ \partial_\mu \partial^\mu = \partial_0^2 - \partial_i^2}}{=} \frac{1}{2} \int d^3x \left((\partial_0 \Phi)^2 + (\partial_i \Phi)^2 + m^2 \Phi^2 \right) \end{aligned} \quad (9.10)$$

By substituting Eq. 9.6 into Eq. 9.10 and using the commutation relations (Eq. 9.7-9.9), we can write

$$\begin{aligned} E &= \frac{1}{2} \int d^3k \frac{1}{(2\pi)^3} \omega_k \left(a^\dagger(k) a(k) + a(k) a^\dagger(k) \right) \\ &\stackrel{\text{Eq. 9.7}}{=} \int d^3k \frac{1}{(2\pi)^3} \omega_k \left(a^\dagger(k) a(k) + \frac{1}{2} (2\pi)^3 \delta^3(0) \right) \end{aligned} \quad (9.11)$$

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 man 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 commutator⁵ of \hat{H} and the Fourier coefficients $a(k)$ and $a^\dagger(k)$. We get

$$\begin{aligned}
 [\hat{H}, a^\dagger(k')] &= \int d^3k \frac{1}{(2\pi)^3} \omega_k [a^\dagger(k) a(k), a^\dagger(k')] \\
 &\stackrel{[a^\dagger(k), a^\dagger(k')]=0}{=} \int d^3k \frac{1}{(2\pi)^3} \omega_k a^\dagger(k) [a(k), a^\dagger(k')] \\
 &= \int d^3k \omega_k a^\dagger(k) \delta^3(k - k') \\
 &\stackrel{\text{See appendix D.2}}{=} \omega_{k'} a^\dagger(k')
 \end{aligned} \tag{9.12}$$

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, which was explained in Sec. 8.3. In this case, 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 field⁶ act on our system?** Let's have a look⁷ at the effect of the first Fourier coefficient, now an operator, on the energy E of the system.

$$\begin{aligned}
 \hat{H} (a(k') |\rangle) &= \left(a(k') \hat{H} + \underbrace{\hat{H} a(k') - a(k') \hat{H}}_{[\hat{H}, a(k')]} \right) |\rangle \\
 &= a(k') \underbrace{\hat{H} |\rangle}_{=E|\rangle} + [\hat{H}, a(k')] |\rangle \\
 &= \left(a(k') E + [\hat{H}, a(k')] \right) |\rangle \\
 &\stackrel{\text{Eq. 9.13}}{=} \left(a(k') E - \omega_{k'} a(k') \right) |\rangle \\
 &= (E - \omega_{k'}) (a(k') |\rangle)
 \end{aligned} \tag{9.15}$$

and equally for the second Fourier coefficient

$$\hat{H} a^\dagger(k') |\rangle = (E + \omega_{k'}) a^\dagger(k') |\rangle \tag{9.16}$$

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

$$|\rangle_2 \equiv a(k') |\rangle$$

⁵ It will become clear in a moment why this is useful.

⁶ Remember: Field=Operator!

⁷ We will do something very clever here, which was first discovered by Dirac while solving the problem of the harmonic oscillator in quantum mechanics.

with

$$\hat{H} |?_2\rangle \underbrace{=} (E - \omega_{k'}) |?_2\rangle.$$

Using Eq. 9.15

This suggests how we should interpret what the field does. Imagine a completely empty system $|0\rangle$, with by definition $H|0\rangle = 0|0\rangle$. If we now act with $a^\dagger(k')$ on $|0\rangle$ we know that this transforms our empty system into a system having energy $\omega_{k'}$

$$\hat{H} a^\dagger(k') |0\rangle \underbrace{=} \omega_{k'} a^\dagger(k') |0\rangle \quad (9.17)$$

Using Eq. 9.16

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

$$a^\dagger(k) |0\rangle \equiv |1_k\rangle \quad (9.18)$$

$$a^\dagger(k) |1_k\rangle \equiv |2_k\rangle \quad (9.19)$$

$$a^\dagger(k') |2_k\rangle \equiv |2_k, 1_{k'}\rangle \quad (9.20)$$

Take a look at the energy again

$$E = \int d^3k \frac{1}{(2\pi)^3} \omega_k a^\dagger(k) a(k).$$

What happens if this operator acts on a state like $|2k_1, k_2\rangle$? 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

$$N(k) \equiv a^\dagger(k) a(k) \quad (9.21)$$

appearing here is a number operator, denoted $N(k)$ that extracts the number of particles with momentum k from a state:

$$N(k) |n_k, n'_{k'}, \dots\rangle = n_k |n_k, n'_{k'}, \dots\rangle \quad (9.22)$$

The energy operator can then be written as

$$E = \int d^3k \frac{1}{(2\pi)^3} \omega_k N(k).$$

Furthermore, take note that there are physical systems, where the momentum spectrum is not continuous, but discrete⁸. 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

$$[a(k), a^\dagger(k')] = \delta_{k,k'}. \quad (9.23)$$

Take note that quantum field theory is, like quantum mechanics, a theory making probabilistic predictions. Therefore, our states need to be normalized $\langle k, k', \dots | k, k', \dots \rangle \stackrel{!}{=} 1$, 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. Therefore, we write

$$a^\dagger(k) |n_k\rangle = C |n_k + 1\rangle, \quad (9.24)$$

where n_k denotes the number of particles with momentum k , and C is some number. From this we get⁹

$$\begin{aligned} (a^\dagger(k) |n_k\rangle)^\dagger &= (C |n_k + 1\rangle)^\dagger \\ &\rightarrow \langle n_k | a(k) = \langle n_k + 1 | C^\dagger. \end{aligned} \quad (9.25)$$

We can therefore write

$$\underbrace{\langle n_k | a(k)}_{9.25} \underbrace{a^\dagger(k) | n_k \rangle}_{9.24} = \underbrace{\langle n_k + 1 | C^\dagger C | n_k + 1 \rangle}_{\text{a number and no operator}} = C^\dagger C \underbrace{\langle n_k + 1 | n_k + 1 \rangle}_{=1} \quad (9.26)$$

or using the discrete commutation relation (Eq. 9.23)

$$\begin{aligned} \langle n_k | a(k) a^\dagger(k) | n_k \rangle &= \langle n_k | \left(\underbrace{a^\dagger(k) a(k)}_{=N(k) \text{ Eq. 9.21}} + \underbrace{\delta_{k,k}}_{=1} \right) | n_k \rangle \\ &\stackrel{\text{Eq. 9.22}}{=} \underbrace{\langle n_k | (n_k + 1) | n_k \rangle}_{\text{a number and no operator}} = (n_k + 1) \underbrace{\langle n_k | n_k \rangle}_{=1} \end{aligned} \quad (9.27)$$

Putting Eq. 9.26 and Eq. 9.27 together yields

$$C^\dagger C = n_k + 1 \rightarrow C = \sqrt{n_k + 1} \quad (9.28)$$

⁸ 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 $\lim_{V \rightarrow \infty}$ is taken.

⁹ Recall that $|n_k\rangle^\dagger = \langle n_k|$ and we have of course $(a^\dagger)^\dagger = a$.

and we therefore have

$$a^\dagger(k) |n_k\rangle = \sqrt{n_k + 1} |n_k + 1\rangle. \quad (9.29)$$

Following the same steps we can derive

$$a |n_k\rangle = \sqrt{n_k} |n_k - 1\rangle. \quad (9.30)$$

Two questions may pop up at this point. Firstly: What happens if we want to annihilate a particle in a completely empty system? And secondly: What about energy or charge conservation? How can we create something from nothing without violating conservation laws? Firstly, the conservation laws are never violated, but this will only become clear when we develop the theory further. Maybe it helps to see that at this point Richard Feynman had the same problem¹⁰

¹⁰ Feynman's Nobel Lecture (December 11, 1965)

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 of course 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 set $|0\rangle$ we get from Eq. 9.30

$$a(k) |0_k\rangle = \sqrt{0} |0_k - 1_k\rangle = 0 \quad (9.31)$$

or equally

$$a(k') |1_k\rangle = \sqrt{0} |1_k, 0_{k'} - 1_{k'}\rangle = 0. \quad (9.32)$$

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 operator 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 $\frac{1}{2}$ Theory

The equation of motion for free spin $\frac{1}{2}$ fields is the Dirac equation, which was derived in Sec. 6.3

$$(i\gamma_\mu \partial^\mu - m)\Psi = 0$$

The general solution of the Dirac equation can be written in the form¹¹

$$\begin{aligned}\Psi &= \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{w_p}} \left(c_r(p) u_r(p) e^{-ipx} + d_r(p) v_r(p) e^{+ipx} \right) \\ &= \Psi^+ + \Psi^-, \end{aligned} \quad (9.33)$$

but this time we do not restrict to real fields, because we saw in Sec. 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

$$\Psi = \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{w_p}} \left(c_r(p) u_r(p) e^{-ipx} + d_r^\dagger(p) v_r(p) e^{+ipx} \right), \quad (9.34)$$

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 Sec. 6.3, the solution

$$\begin{aligned}\bar{\Psi} &= \sum_r \sqrt{\frac{m}{(2\pi)^3}} \int \frac{d^3p}{\sqrt{w_p}} \left(c_r^\dagger(p) \bar{u}_r(p) e^{+ipx} + d_r(p) \bar{v}_r(p) e^{-ipx} \right) \\ &= \bar{\Psi}^+ + \bar{\Psi}^- \end{aligned} \quad (9.35)$$

where u_1, u_2, v_1, v_2 are the "basis spinors" in an arbitrary frame¹²

$$u_1 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} 1 \\ 0 \\ \frac{p_3}{E+m} \\ \frac{p_1+ip_2}{E+m} \end{pmatrix} \quad u_2 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} 0 \\ 1 \\ \frac{p_1-ip_2}{E+m} \\ \frac{-p_3}{E+m} \end{pmatrix} \quad (9.36)$$

$$v_1 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \frac{p_1-ip_2}{E+m} \\ \frac{-p_3}{E+m} \\ 0 \\ 1 \end{pmatrix} \quad v_2 = \sqrt{\frac{E+m}{2m}} \begin{pmatrix} \frac{p_3}{E+m} \\ \frac{p_1+ip_2}{E+m} \\ 1 \\ 0 \end{pmatrix}. \quad (9.37)$$

The rest of the theory for free spin $\frac{1}{2}$ fields can be developed similarly to the scalar theory, but there is one small¹³ difference. It can

¹¹ The Dirac equation is solved in Sec. 8.9. The general solution is then written analogous to the solution of the Klein-Gordon equation, discussed in the last section.

¹² We derived these spinors in the rest frame ($p_i = 0$) in Sec. 8.10. The basis spinors in an arbitrary frame can be computed from the basis spinors in the rest frame by a boost transformation.

¹³ With incredible huge consequences! In fact, nothing in this universe would be stable if the spin $\frac{1}{2}$ would work exactly like the scalar theory.

$$\begin{aligned} &^{14} [\Phi(x), \pi(y)] \\ &= \Phi(x)\pi(y) - \pi(y)\Phi(x) = i\delta(x-y) \end{aligned}$$

be seen that the commutation relation we used for scalar fields does not work for spin $\frac{1}{2}$ fields. 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. If we are now computing, using the **commutation** relation¹⁴, the Hamiltonian for a spin $\frac{1}{2}$ field, we get something of the form

$$H \sim \int c^\dagger c - d^\dagger d$$

which 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 as in the scalar case. But if we use instead the **anticommutation** relation

$$\{\Phi(x), \pi(y)\} = \Phi(x)\pi(y) + \pi(y)\Phi(x) = i\delta(x-y),$$

the term involving the anti-particle creation and annihilation operators comes out positive and the problem is therefore solved.

¹⁵ Analogous to Eq. 9.9 for scalars, but now with the anticommutator instead of the commutator $[\cdot, \cdot] \rightarrow \{\cdot, \cdot\}$.

What may seem like an awkward trick, has very interesting consequences. For example, we have then¹⁵

$$\{c^\dagger(k), c^\dagger(k')\} = 0$$

from which we can conclude

$$\{c^\dagger(k), c^\dagger(k)\} = c^\dagger(k)c^\dagger(k) + c^\dagger(k)c^\dagger(k) = 2c^\dagger(k)c^\dagger(k) = 0.$$

$$\Rightarrow c^\dagger(k)c^\dagger(k) = 0 \quad (9.38)$$

¹⁶ For spin 0 fields the corresponding equation isn't very surprising, because there we use the commutator: $[c^\dagger(k), c^\dagger(k)] = c^\dagger(k)c^\dagger(k) - c^\dagger(k)c^\dagger(k) = 0$.

The action of two equal creation operators¹⁶ 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:

$$\{c_r(p), c_s^\dagger(p')\} = \delta_{rs}\delta(p-p') \quad \{d_r(p), d_s^\dagger(p')\} = \delta_{rs}\delta(p-p') \quad (9.39)$$

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 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 analogously to the Hamiltonian for a free spin 0 field

$$H_{\text{free}}^{\frac{1}{2}} = \int d^3x (-i\bar{\Psi}\gamma_i\partial^i\Psi - m\bar{\Psi}\Psi) \quad (9.40)$$

or expressed in terms of the Fourier coefficients

$$\mathcal{H}_{\text{free}}^{\frac{1}{2}} = \sum_r \int d^3p \, w_p \left(c_r^\dagger(p) c_r(p) + d_r^\dagger(p) d_r(p) + \text{const} \right) \quad (9.41)$$

where again the constant term leads to an infinite term and we choose to ignore this.

9.4 Free Spin 1 Theory

The solution for the equation of motion for free spin 1 fields, the Proca equation, which was derived in Sec. 6.4

$$m^2 A^\rho = \frac{1}{2} \partial_\sigma (\partial^\sigma A^\rho - \partial^\rho A^\sigma) \quad (9.42)$$

is, analogously to the spin 0 field solution, of the form

$$A_\mu = \int \frac{d^3k}{\sqrt{(2\pi)^3 2\omega_k}} \left(\epsilon_{r,\mu}(k) a_r(k) e^{-ikx} + \epsilon_{r,\mu}(k) a_r^\dagger(k) e^{ikx} \right) \quad (9.43)$$

$$A_\mu = A_\mu^+ + A_\mu^- \quad (9.44)$$

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 and 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)

$$\mathcal{L}_{\text{Dirac+Extra-Term}} = -m \bar{\Psi} \Psi + i \bar{\Psi} \gamma_\mu \partial^\mu \Psi + g A_\mu \bar{\Psi} \gamma^\mu \Psi \quad (9.45)$$

from which we derive the corresponding Hamiltonian¹⁷

¹⁷ We will use this Hamiltonian in a moment!

$$\begin{aligned}
H &= \int d^3x T^{00} \\
&= \int d^3x \left(\underbrace{\frac{\partial \mathcal{L}}{\partial(\partial_0 \Psi)}}_{=i\bar{\Psi}\gamma_0} \partial_0 \Psi - \mathcal{L} \right) \\
&= \int d^3x \left(i\bar{\Psi}\gamma_0 \partial_0 \Psi + m\bar{\Psi}\Psi - i\bar{\Psi} \underbrace{\gamma_\mu \partial^\mu \Psi}_{=\gamma_0 \partial_0 - \gamma_i \partial_i} - g A_\mu \bar{\Psi} \gamma^\mu \Psi \right) \\
&= \underbrace{\int d^3x (m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_i \partial_i \Psi)}_{=H_{\text{free}}^{\frac{1}{2}}} - \underbrace{\int d^3x (g A_\mu \bar{\Psi} \gamma^\mu \Psi)}_{\equiv -H_I} \\
&= H_{\text{free}}^{\frac{1}{2}} + H_I
\end{aligned} \tag{9.46}$$

9.5.1 Scatter Amplitudes

¹⁸ This is what physicists prepare in collider experiments.

One of the main goals in quantum field theory is to compute the probability for a fixed number n of particles with defined¹⁸ momenta p_1, p_2, \dots, p_n to transform into a (possibly different) number of (possibly different) particles n' with momenta $q_1, q_2, \dots, q_{n'}$. We can write this using Dirac's notation

$$\langle q_1, q_2, \dots, q_{n'} | \hat{S} | p_1, p_2, \dots, p_n \rangle, \tag{9.47}$$

where \hat{S} is the operator describing the scattering. We will derive how this operator looks concretely in the next section. This means 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, denoted by $|\rangle$, 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)

$$H_{\text{free}}^0 = \frac{1}{2} \int d^3x \left((\partial_0 \Phi)^2 + (\partial_i \Phi)^2 + m\Phi^2 \right) \tag{9.48}$$

while for the free spin $\frac{1}{2}$ theory we have (Eq. 9.40)

$$H_{\text{free}}^{\frac{1}{2}} = \int d^3x (m\bar{\Psi}\Psi + i\bar{\Psi}\gamma_i \partial_i \Psi). \tag{9.49}$$

Both identifications are operators in a field theory, representing energy, and we therefore write

$$i\partial_0 |?(t)\rangle = H |?(t)\rangle, \quad (9.50)$$

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 to be $t = 0$ we have¹⁹

$$|?(t)\rangle = U(t) |?(0)\rangle. \quad (9.51)$$

Putting this ansatz into Eq. 9.50 yields

$$i\partial_0 U(t) |?(0)\rangle = HU(t) |?(0)\rangle. \quad (9.52)$$

This equation holds for arbitrary $|?(0)\rangle$ and therefore²⁰

$$i\partial_0 U(t) = HU(t). \quad (9.53)$$

The general solution of this equation is

$$U(t) = e^{-i \int_0^t dx_0 H} \quad (9.54)$$

as we can check

$$\begin{aligned} i\partial_0 U(t) = HU(t) &\rightarrow i\partial_0 e^{-i \int_0^t dx_0 H} = H e^{-i \int_0^t dx_0 H} \\ &\rightarrow i(-iH) e^{-i \int_0^t dx_0 H} = H e^{-i \int_0^t dx_0 H} \\ &\rightarrow H e^{-i \int_0^t dx_0 H} = H e^{-i \int_0^t dx_0 H} \quad \checkmark \end{aligned} \quad (9.55)$$

In experiments we never measure a ket $|?(t)\rangle$, but always the combinations of a bra with a ket. In general, we have objects of the form

$$\langle f(t) | \hat{O} | i(t) \rangle, \quad (9.56)$$

with some operator \hat{O} , initial state $|i(t)\rangle$ and final state $\langle f(t)|$. 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 this in terms of the time-evolution operator leads us to²¹

$$\langle f(0) | U^\dagger(t) \hat{O} U(t) | i(0) \rangle. \quad (9.57)$$

An important idea is that we can switch our perspective and say the operator \hat{O} evolves in time, following the rule $U^\dagger(t) \hat{O} U(t)$ and the bra and ket are time-independent. This new perspective is called **Heisenberg picture**.

¹⁹ In general $|?(t)\rangle = U(t - t') |?(t')\rangle$.

²⁰ We discussed this line of thought already in a sidenote in Sec. 5.1.

²¹ Remember $\langle f | = |f\rangle^\dagger$

There is a similar, very useful trick that is conventionally used in interaction theory. The Hamiltonian is always the sum of the free Hamiltonian and the interaction Hamiltonian (Eq. 9.46)

$$H = H_{\text{free}} + H_I. \quad (9.58)$$

The trick is to use a mixture of the two perspectives introduced above. We let the states evolve according to H_I and the operators²² 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

$$|i(t)\rangle_I \equiv U_{\text{free}}^\dagger |i(t)\rangle_S, \quad (9.59)$$

where $U_{\text{free}} = e^{-i \int_0^t dx_0 H_{\text{free}}}$ 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 get²³ the time-evolution equation in the interaction picture:

$$i\partial_0 |i(t)\rangle_S = H |i(t)\rangle_S$$

$$\rightarrow i\partial_0 U_{\text{free}} |i(t)\rangle_I = H U_{\text{free}} |i(t)\rangle_I$$

$$\rightarrow i\partial_0 e^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I = (H_{\text{free}} + H_I) e^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I$$

$$\underbrace{\rightarrow H_{\text{free}} e^{-i \int_0^t dx_0 H_{\text{free}}}}_{\text{product rule}} |i(t)\rangle_I + i e^{-i \int_0^t dx_0 H_{\text{free}}} \partial_0 |i(t)\rangle_I = (H_{\text{free}} + H_I) e^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I$$

$$\rightarrow i e^{-i \int_0^t dx_0 H_{\text{free}}} \partial_0 |i(t)\rangle_I = H_I e^{-i \int_0^t dx_0 H_{\text{free}}} |i(t)\rangle_I$$

$$\rightarrow i\partial_0 |i(t)\rangle_I = \underbrace{e^{i \int_0^t dx_0 H_{\text{free}}} H_I e^{-i \int_0^t dx_0 H_{\text{free}}}}_{=H_I^{\text{int}}} |i(t)\rangle_I$$

$=H_I^{\text{int}}$ The interaction Hamiltonian in the interaction picture.

$$\rightarrow i\partial_0 |i(t)\rangle_I = H_I^{\text{int}} |i(t)\rangle_I \quad (9.60)$$

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

$$\langle f(t_f) | \hat{S}(t_f, t_i) | i(t_i) \rangle \quad ? \quad (9.61)$$

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, this is not one specific

²² Recall that in quantum field theory the fields are our operators.

²³ 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.

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 have²⁴

$$\hat{S}(t, t_i) |i(t_i)\rangle = |\Psi(t)\rangle = \sum_f \underbrace{S_{fi}}_{\text{complex numbers}} |f(t)\rangle. \quad (9.62)$$

The multiplication with one specific $\langle f'(t)|$ from the left-hand side terminates all but one term of the sum:

$$\begin{aligned} \langle f'(t)| \sum_f S_{fi} |f(t)\rangle &= \sum_f S_{fi} \underbrace{\langle f'(t)| |f(t)\rangle}_{=\delta_{ff'} \text{ because basis states are orthogonal}} = \sum_f S_{fi} \delta_{ff'} = S_{f'i}. \end{aligned} \quad (9.63)$$

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 we derived above²⁵

$$i\partial_t |\Psi(t)\rangle_I = H_I |\Psi(t)\rangle_I. \quad (9.64)$$

We can rewrite this in terms of our initial state and the operator \hat{S} using Eq. 9.62,

$$\begin{aligned} \rightarrow i\partial_t \hat{S}(t, t_i) |i(t_i)\rangle_I &= H_I \hat{S} |i(t_i)\rangle_I \\ \underbrace{\rightarrow}_{\text{product rule}} i(\partial_t \hat{S}(t, t_i)) |i(t_i)\rangle_I + i\hat{S}(t, t_i) \underbrace{\partial_t |i(t_i)\rangle_I}_{=0 \text{ because } |i(t_i)\rangle_I \text{ does not depend on } t} &= H_I \hat{S} |i(t_i)\rangle_I. \end{aligned}$$

Now using that this equation holds for arbitrary initial states we can write

$$i\partial_t \hat{S}(t, t_i) = H_I \hat{S} \quad (9.65)$$

with the general solution²⁶

$$\hat{S}(t, t_i) = e^{-i \int_{t_i}^t dt' H_I} \quad (9.66)$$

At a first glance the problem is easy now. We know H_I and just have to solve the integral, which gives the operator $\hat{S}(t, t_i)$ and by acting with it on our initial states gives us the probability amplitudes for the different processes. Unfortunately, the integral cannot be solved. 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

²⁴ This can be seen as a series expansion of the state $|\Psi(t)\rangle$ in terms of our particle states $|f(t)\rangle$.

²⁵ Eq. 9.60 and in order to avoid notational clutter we suppress the superscript "int", which denotes that we are working in the interaction picture here.

²⁶ We omitted something very important here, called time ordering, which we will discuss in the next section.

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, because what we measure is always a result after all the interactions have happened.

9.5.3 Dyson Series

Because there is no analytic solution for the exponentiated integral we expand it in a Taylor series²⁷

²⁷ This is derived in appendix B.4.1:

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \dots$$

$$\begin{aligned}\hat{S}(\infty, -\infty) &= e^{-i \int_{-\infty}^{\infty} dt' H_I(t')} \\ &= 1 - i \int_{-\infty}^{\infty} dt_1 H_I(t_1) \\ &\quad - \frac{1}{2!} \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) + \dots \quad (9.67)\end{aligned}$$

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

$$T \{ A(x) B(y) \} := \begin{cases} A(t_1) B(t_2) & \text{if } t_1 > t_2 \\ B(t_2) A(t_1) & \text{if } t_1 < t_2. \end{cases} \quad (9.68)$$

Therefore we write, giving a name to each term of the series expansion,

$$\begin{aligned}\hat{S}(\infty, -\infty) &= \underbrace{1}_{S^{(0)}} - i \underbrace{\int_{-\infty}^{\infty} dt_1 H_I(t_1)}_{S^{(1)}} \\ &\quad - \underbrace{\frac{1}{2!} T \left\{ \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) \right\}}_{S^{(2)}} + \dots \quad (9.69)\end{aligned}$$

or written with a sum

$$\begin{aligned}\hat{S}(\infty, -\infty) &= \sum_{n=0}^{\infty} \frac{(-i)^n}{n!} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} T \left\{ \left(\int_{-\infty}^{\infty} dt_1 H_I(t_1) \right) \left(\int_{-\infty}^{\infty} dt_2 H_I(t_2) \right) \dots \left(\int_{-\infty}^{\infty} dt_n H_I(t_n) \right) \right\} \\ &= \sum_{n=0}^{\infty} S^{(n)}.\end{aligned}\quad (9.70)$$

This is useful because H_I has a numerical factor²⁸ in it, the coupling constant of the corresponding interaction, i.e. $H_I \propto g$. This coupling constant is, for example, for the 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. 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.

²⁸ See Eq. 9.46 where the coupling constant is denoted e .

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 further²⁹, so we are going to use these results without proofs. The interested reader is referred to the standard texts on quantum field theory³⁰.

²⁹ Because now, the fun is about to begin!

³⁰ 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 section: 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 individually.

The first term of the series is trivial as it is simply the identity operator

$$S^{(0)} = I \quad (9.71)$$

The second term is more exciting.

$$S^{(1)} = -i \int_{-\infty}^{\infty} dt H_I = ig \int_{-\infty}^{\infty} d^4x A_\mu \bar{\Psi} \gamma^\mu \Psi \quad (9.72)$$

which we rewrite, recalling Eq. 9.33 and Eq. 9.44

$$S^{(1)} = ig \int_{-\infty}^{\infty} d^4x (A_\mu^+ + A_\mu^-) (\bar{\Psi}^+ + \bar{\Psi}^-) \gamma^\mu (\Psi^+ + \Psi^-). \quad (9.73)$$

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 $|e^+(p_1), e^-(p_2)\rangle$

$$ig \int_{-\infty}^{\infty} d^4x A_{\mu}^+ \bar{\Psi}^+ \gamma^{\mu} \Psi^+ |e^+(p_1), e^-(p_2)\rangle.$$

³¹ Remember the operators with the † are those who create and the operators without are those who destroy. Ψ^+ is defined in Eq. 9.33.

³² We integrate over all possible momenta!

Ψ^+ consists of destruction³¹ operators for particles for all³² 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, if we want the contribution of this one term to the probability amplitude for the creation of a photon $\langle\gamma|$, with momentum k'

$$\begin{aligned} \langle f | S_1^{(1)} | i \rangle &= \langle \gamma_{k'} | S_1^{(1)} | e^+(p_1), e^-(p_2) \rangle \\ &= \int_{-\infty}^{\infty} d^4x \langle \gamma_{k'} | \sum_k \text{constant}(k) | \gamma_k \rangle e^{-ix(p_1+p_2-k)} \\ &= \int_{-\infty}^{\infty} d^4x \sum_k \text{constant}(k) \underbrace{\langle \gamma_{k'} | \gamma_k \rangle}_{=\delta_{kk'}} e^{-ix(p_1+p_2-k)} \\ &= \int_{-\infty}^{\infty} d^4x \text{constant}(k') e^{-ix(p_1+p_2-k')}. \end{aligned} \tag{9.74}$$

The integration over x results in a delta function $\delta(p_1 + p_2 - k)$ that represents 4-momentum conservation³³. 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

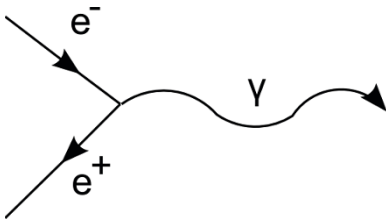


Fig. 9.1: Feynman graph for the process $e^+e^- \rightarrow \gamma$

³³ 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 : $p_1 + p_2 - k \stackrel{!}{=} 0$. Otherwise $\delta(p_1 + p_2 - k) = 0$ as explained in appendix D.2.

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} .

$$\begin{aligned} S^{(2)} &= -\frac{1}{2!} T \left\{ \left(\int_{-\infty}^{\infty} d^4 x_1 \mathcal{H}_I(x_1) \right) \left(\int_{-\infty}^{\infty} d^4 x_2 \mathcal{H}_I(x_2) \right) \right\} \\ &= -\frac{1}{2!} T \left\{ \left(\int_{-\infty}^{\infty} d^4 x_1 g A_\mu(x_1) \bar{\Psi}(x_1) \gamma^\mu \Psi(x_1) \right) \left(\int_{-\infty}^{\infty} d^4 x_2 g A_\mu(x_2) \bar{\Psi}(x_2) \gamma^\mu \Psi(x_2) \right) \right\}, \end{aligned} \quad (9.75)$$

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

$$-\frac{1}{2!} g^2 \int \int d^4 x_1 d^4 x_2 N \left\{ \bar{\Psi}(x_1) \gamma^\mu \Psi(x_1) [A_\mu(x_1), A_\mu(x_2)] \bar{\Psi}(x_2) \gamma^\mu \Psi(x_2) \right\}$$

where the computation of the commutator can be done using the explicit solution for A_μ and the result is called Photon propagator³⁴ $[A_\mu(x_1), A_\mu(x_2)] \equiv iD_\mu(x_1 - x_2)$.

From this term we get again many, many terms, because every $\Psi, \bar{\Psi}$ etc. is actually a sum of two terms, and we will take a look at just one of them. Therefore, qualitatively we have for one³⁵ of these many many terms, if we start again with an electron and a positron, something of the form

$$-\frac{1}{2!} g^2 \int \int d^4 x_1 d^4 x_2 \bar{\Psi}^-(x_1) \Psi^-(x_1) D_\mu(x_1 - x_2) \bar{\Psi}^+(x_2) \Psi^+(x_2) |e^+, e^-\rangle \quad (9.76)$$

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. In the same way, all the other terms can be interpreted as some reaction between massive³⁶ spin $\frac{1}{2}$ and mass-

³⁴ 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 $i\Delta = \langle 0 | [\Phi^{++}(y), \Phi^-(x)] | 0 \rangle$ and after many pages of math we get an expression of the form $i\Delta = \frac{-i}{2\pi^3} \int \frac{d^3 k}{\omega_k} e^{-ik(x-y)}$. 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.

³⁵ We pick again the term not resulting in a zero for the collision of an electron with a positron.

³⁶ Here we looked at electrons and positrons. Other possibilities are the quarks or the other two leptons μ and τ .

³⁷ Here photons

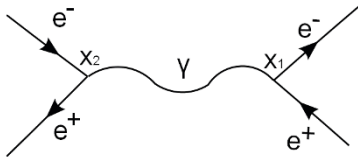


Fig. 9.2: Feynman graph for the process $e^+e^- \rightarrow \gamma \rightarrow e^+e^-$

less³⁷ 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. Unfortunately, this does not always work for the interaction term that follows from $SU(3)$ gauge symmetry, because the coupling constant of strong interactions is too big to allow the usage of a series expansion. If the coupling constant is bigger than 1, higher order terms 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. For quantum chromodynamics, as this branch of physics is called, other computational schemes are needed, but this goes beyond the scope of this book.

Further Reading Tips

³⁸ Robert D. Klauber. *Student Friendly Quantum Field Theory*. Sandtrove Press, 2nd edition, 12 2013. ISBN 9780984513956

³⁹ Francis Halzen and Alan D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1st edition, 1 1984. ISBN 9780471887416

⁴⁰ Anthony Zee. *Quantum Field Theory in a Nutshell*. Princeton University Press, 1st edition, 3 2003. ISBN 9780691010199

⁴¹ Franz Mandl and Graham Shaw. *Quantum Field Theory*. Wiley, 2nd edition, 5 2010. ISBN 9780471496847

⁴² Michele Maggiore. *A Modern Introduction to Quantum Field Theory*. Oxford University Press, 1st edition, 2 2005. ISBN 9780198520740

⁴³ Ian J.R. Aitchison and Anthony J.G. Hey. *Gauge Theories in Particle Physics*. CRC Press, 4th edition, 12 2012. ISBN 9781466513174

- **Robert D. Klauber - Student Friendly Quantum Field Theory**³⁸ 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**³⁹ 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**⁴⁰ 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.
- **Franz Mandl, Graham Shaw - Quantum Field Theory**⁴¹ is a very good starting point regarding weak and strong interaction theory
- **Michele Maggiore - A Modern Introduction to Quantum Field Theory**⁴² is a great introduction with a strong focus on group theoretical concepts.
- **Ian J. R. Aitchison and Anthony J. G. Hey - Gauge Theories in Particle Physics**⁴³ offers very illuminating explanations regarding the ideas of gauge theories.

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 do the job

$$\Phi(x) = ae^{i(px-Et)} = e^{-i(p_\mu x^\mu)}$$

because of the energy-momentum relation of special relativity (Eq. 8.2)

$$E^2 = \vec{p}^2 + m^2 \rightarrow p_\mu p^\mu = m^2$$

we see

$$\begin{aligned} (\partial_\mu \partial^\mu + m^2)\Phi &= (\partial_\mu \partial^\mu + m^2)e^{-i(p_\mu x^\mu)} \\ &= (-p_\mu p^\mu + m^2)e^{-i(p_\mu x^\mu)} \\ &= (-m^2 + m^2)e^{-i(p_\mu x^\mu)} \\ &= 0 \quad \checkmark \end{aligned} \tag{9.77}$$

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 by linear combinations.

A general solution is given by superposition of all possible solutions, which can be seen as the Fourier expansion⁴⁴ of the solution

$$\Phi(x) = \int \frac{d^4p}{(2\pi)^4} (a(p)e^{-i(p_\mu x^\mu)} + a^\dagger(p)e^{i(p_\mu x^\mu)})$$

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 $k_i = \frac{p_i}{\hbar}$ and the frequency $k_0 = \omega = \frac{E}{\hbar}$ instead of the energy and 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

$$\Phi(x) = \int \frac{d^4k}{(2\pi)^4} (a(k)e^{-i(kx)} + a^\dagger(k)e^{i(kx)})$$

Take note that not all solutions of the Klein-Gordon equation are suited to describe nature, because we have the **"mass-shell" condition**

$$p_\mu p^\mu = k_\mu k^\mu = m^2 \rightarrow k_0^2 - k_i^2 \stackrel{!}{=} m^2 \rightarrow k^2 \stackrel{!}{=} m^2. \tag{9.78}$$

⁴⁴ That's where the factors of 2π come from. Another way of seeing this is demanding the solutions to form an orthonormal set: $\int d^4k e^{ikx} e^{-ikx'} = \int d^4k e^{ik(x-x')} = 2\pi\delta(x-x')$. Therefore, the factors of 2π are normalisation constants.

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 distribution⁴⁵

⁴⁵ This is explained in appendix D.2.

$$\Phi(x)_{\text{physical}} = \int \frac{d^4k}{(2\pi)^4} 2\pi \delta(k^2 - m^2) \left(a(k) e^{-i(kx)} + a^\dagger(k) e^{i(kx)} \right)$$

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 build 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

$$\begin{aligned} dk^4 \delta(k^2 - m^2) \theta(k_0) &= dk^4 \delta(k_0^2 - \underbrace{\vec{k}^2 - m^2}_{\equiv -\omega_k^2 \text{ (definition)}}) \theta(k_0) \\ &= dk^4 \delta(k_0^2 - \omega_k^2) \theta(k_0) \\ &= dk^4 \delta((k_0 - \omega_k)(k_0 + \omega_k)) \theta(k_0) \\ &\underbrace{=}_{\text{using } \delta(f(x)) = \sum_i \frac{\delta(x - a_i)}{\left| \frac{df}{dx}(a_i) \right|} \text{ where } a_i \text{ denotes the roots, i.e. } f(a_i) = 0, \text{ of the function } f(x)} dk^4 \frac{1}{2k_0} (\delta(k_0 - \omega_k) + \delta(k_0 + \omega_k)) \theta(k_0) \\ &\underbrace{=}_{\text{because the argument of } \delta(k_0 + \omega_k) \text{ never becomes zero with } k_0 \geq 0} dk^4 \frac{1}{2k_0} \delta(k_0 - \omega_k) \\ &= dk^3 dk_0 \frac{1}{2k_0} \delta(k_0 - \omega_k) \\ &\underbrace{=}_{\text{Integrating over } k_0} dk^3 \frac{1}{2\omega_k} \end{aligned} \tag{9.79}$$

So finally the general and physical solution to the Klein-Gordon equation reads

$$\Phi(x) = \int dk^3 \frac{1}{(2\pi)^3 2\omega_k} \left(a(k) e^{-i(kx)} + a^\dagger(k) e^{i(kx)} \right) \tag{9.80}$$

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)

$$\langle \hat{O} \rangle = \int d^3x \Psi^* \hat{O} \Psi$$

and the Schrödinger equation for a particle in an external potential (Eq. 8.22)

$$\left(i \frac{d}{dt} - \frac{\nabla^2}{2m}\right) \Psi - V \Psi = 0$$

$$\rightarrow i \frac{d}{dt} \Psi = \underbrace{\left(\frac{\nabla^2}{2m} + V\right)}_{=:H} \Psi$$

$$\rightarrow \frac{d}{dt} \Psi = \frac{1}{i} H \Psi$$

$$\underbrace{\frac{d}{dt} \Psi^*}_{\text{Because } H^\dagger = H} = -\frac{1}{i} \underbrace{\Psi^* H}_{\text{Here } \Psi^\dagger = \Psi^*}$$

Taking the time derivative of the expectation value yields

$$\frac{d}{dt} \langle \hat{O} \rangle = \int d^3x \left(\left(\frac{d}{dt} \Psi^* \right) \hat{O} \Psi + \Psi^* \left(\frac{d}{dt} \hat{O} \right) \Psi + \Psi^* \hat{O} \left(\frac{d}{dt} \Psi \right) \right)$$

We use now $\frac{d}{dt} \hat{O} = 0$, which is true for most operators. For example, for $\hat{O} = \hat{p} = -i\vec{\nabla} \neq \hat{O}(t)$. In addition, we use the Schrödinger equation to rewrite the time derivatives of the wave function and its

complex conjugate. This yields

$$\begin{aligned}
 \frac{d}{dt} \langle \hat{O} \rangle &= \int d^3x \left(\left(-\frac{1}{i} \Psi^* H \right) \hat{O} \Psi + \Psi^* \hat{O} \left(\frac{1}{i} H \Psi \right) \right) \\
 &= \frac{1}{i} \int d^3x \left((-\Psi^* H) \hat{O} \Psi + \Psi^* \hat{O} (H \Psi) \right) \\
 &= \frac{1}{i} \int d^3x \Psi^* [\hat{O}, H] \Psi \\
 &= \frac{1}{i} \langle [\hat{O}, H] \rangle,
 \end{aligned} \tag{10.1}$$

which is known as **Ehrenfest theorem**. If we specify $\hat{O} = \hat{p}$ and use $H = \frac{\hat{p}^2}{2m} + V$ we get

$$\begin{aligned}
 \frac{d}{dt} \langle \hat{p} \rangle &= \frac{1}{i} \langle [\hat{p}, H] \rangle \\
 &= \frac{1}{i} \langle [\hat{p}, \frac{\hat{p}^2}{2m} + V] \rangle \\
 &= \frac{1}{i} \langle \underbrace{[\hat{p}, \frac{\hat{p}^2}{2m}]}_{=0} + [\hat{p}, V] \rangle \\
 &= \frac{1}{i} \langle [\hat{p}, V] \rangle \\
 &= \frac{1}{i} \int d^3x \Psi^* [\hat{p}, V] \Psi \\
 &= \frac{1}{i} \int d^3x \Psi^* \hat{p} V \Psi - \frac{1}{i} \int d^3x \Psi^* V \hat{p} \Psi \\
 &= \frac{1}{i} \int d^3x \Psi^* (-i \nabla) V \Psi - \frac{1}{i} \int d^3x \Psi^* V (-i \nabla) \Psi \\
 &\quad \underbrace{\quad}_{\text{Product rule}} - \int d^3x \Psi^* (\nabla V) \Psi - \int d^3x \Psi^* V \nabla \Psi + \int d^3x \Psi^* V \nabla \Psi \\
 &= - \int d^3x \Psi^* (\nabla V) \Psi \\
 &= \langle -\nabla V \rangle = \langle F \rangle.
 \end{aligned} \tag{10.2}$$

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 location² and we therefore have

$$m \frac{d^2}{dt^2} x = F_1 + F_2 + \dots \tag{10.3}$$

¹ Recall that we used this equation, without a derivation, to illustrate the conserved quantities following from Noether's theorem in Sec. 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 Chap. 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.

Happily, we already know something that is invariant under all Lorentz transformations: The invariant of special relativity, which was derived in Sec. 2.1

$$(ds)^2 = (cd\tau)^2 = (cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2, \quad (10.4)$$

where τ is the proper time as explained in Sec. 2.2. Equally the square root of this is invariant and therefore the simplest possible thing we can minimize is

$$S = \int C d\tau \quad \text{with some constant } C \text{ and } d\tau = \frac{1}{c} \sqrt{(cdt)^2 - (dx)^2 - (dy)^2 - (dz)^2} \quad (10.5)$$

The correct constant turns out to be $C = -mc^2$ and therefore we need to minimize

$$S = -mc^2 \int d\tau. \quad (10.6)$$

For brevity we will restrict the following discussion to one dimension. Then we can write

$$\begin{aligned} d\tau &= \frac{1}{c} \sqrt{(cdt)^2 - (dx)^2} = \frac{1}{c} \sqrt{(cdt)^2 \left(1 - \frac{(dx)^2}{c^2(dt)^2}\right)} \\ &= \frac{1}{c} (cdt) \sqrt{1 - \frac{1}{c^2} \left(\frac{dx}{dt}\right)^2} \underbrace{=}_{\substack{\frac{dx}{dt} = \dot{x} \text{ is the velocity of the particle in question}}} dt \sqrt{1 - \frac{\dot{x}^2}{c^2}}. \end{aligned} \quad (10.7)$$

Putting this into Eq. 10.6 yields

$$S = \int \underbrace{-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}}}_{\equiv \mathcal{L}} dt. \quad (10.8)$$

As always, we can find the minimum by putting this into the Euler-Lagrange equation (Eq. 4.7)

$$\begin{aligned}
 \frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}} \right) &= 0 \\
 \rightarrow \underbrace{\frac{\partial}{\partial x} \left(-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}} \right)}_{=0} - \frac{d}{dt} \left(\frac{\partial}{\partial \dot{x}} \left(-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}} \right) \right) &= 0 \\
 \rightarrow c^2 \frac{d}{dt} \left(\frac{-m \frac{\dot{x}}{c^2}}{\sqrt{1 - \frac{\dot{x}^2}{c^2}}} \right) = 0 \rightarrow \frac{d}{dt} \left(\frac{m \dot{x}}{\sqrt{1 - \frac{\dot{x}^2}{c^2}}} \right) &= 0 \quad (10.9)
 \end{aligned}$$

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 Lagrangian³

³ We add $-V(x)$ instead of $V(x)$, because then we get in the energy of the system, which we can compute using Noether's theorem, the potential energy term $+V(x)$.

$$\mathcal{L} = -mc^2 \sqrt{1 - \frac{(\dot{x})^2}{c^2}} - V(x) \quad (10.10)$$

For this Lagrangian the Euler-Lagrange equation yields

$$\frac{d}{dt} \left(\frac{m \dot{x}}{\sqrt{1 - \frac{\dot{x}^2}{c^2}}} \right) = -\frac{dV}{dx} \equiv F. \quad (10.11)$$

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 Lagrangian we derived in the last section in the non-relativistic limit, which means when the particle moves slowly compared with the speed of light: $\dot{x} \ll c$. We can then use the Taylor formula⁴

⁴ See appendix B.3

$$-mc^2 \sqrt{1 - \frac{\dot{x}^2}{c^2}} = -mc^2 \left(1 - \frac{1}{2} \frac{\dot{x}^2}{c^2} + \dots \right). \quad (10.12)$$

In the limit $\dot{x} \ll c$ we can neglect higher order terms and the Lagrangian reads

$$\mathcal{L} = -mc^2 + \frac{1}{2} m \dot{x}^2 - V(x) \quad (10.13)$$

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

$$\mathcal{L} = \frac{1}{2}m\dot{x}^2 - V(x). \quad (10.14)$$

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

$$\begin{aligned} & \frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{x}} \right) = 0 \\ \rightarrow & \frac{\partial}{\partial x} \left(\frac{1}{2}m\dot{x}^2 \right) - \frac{d}{dt} \left(\frac{\partial}{\partial \dot{x}} \left(\frac{1}{2}m\dot{x}^2 - V(x) \right) \right) = 0 \\ \rightarrow & -\frac{\partial}{\partial x} V(x) - \frac{d}{dt} m\dot{x} = 0 \rightarrow \frac{d}{dt} m\dot{x} = -\frac{\partial}{\partial x} V(x) \end{aligned} \quad (10.15)$$

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 Chap. 7 one of the most important equations of classical electrodynamics: The inhomogeneous Maxwell equations (Eq. 7.22)

$$\partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) = J^\rho \quad (11.1)$$

or in a more compact form using the electromagnetic tensor¹

$$\partial_\sigma F^{\sigma\rho} = J^\rho. \quad (11.2)$$

$$^1 F^{\sigma\rho} \equiv (\partial^\sigma A^\rho - \partial^\rho A^\sigma)$$

We discovered in Sec. 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 field-strength 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

$$F^{i0} = \partial^i A^0 - \partial^0 A^i \quad (11.3)$$

and three are

$$F^{ij} = \partial^i A^j - \partial^j A^i = (\delta_l^i \delta_m^j - \delta_m^i \delta_l^j) \partial^l A^m \underbrace{=}_{\epsilon^{ijk} \epsilon^{klm}} \epsilon^{ijk} \epsilon^{klm} \partial^l A^m \quad (11.4)$$

This is a standard relation for the multiplication of two ϵ with one coinciding index

The standard way to label these components is

$$\partial^i A^0 - \partial^0 A^i \equiv E^i \quad (11.5)$$

$$\epsilon^{ijk} \partial^j A^k \equiv -B^i \quad (11.6)$$

and therefore

$$F^{i0} = E^i \quad (11.7)$$

$$F^{ij} = \epsilon^{ijk} \epsilon^{klm} \partial^l A^m = -\epsilon^{ijk} B^k \quad (11.8)$$

If we now rewrite the inhomogeneous Maxwell equations² as

² Plural because we have one equation for each component ρ .

$$\partial_\sigma F^{\rho\sigma} = \partial_0 F^{\rho 0} - \partial_k F^{\rho k} = J^\rho \quad (11.9)$$

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 components³ ($\rho \rightarrow i$)

$$\partial_0 F^{i0} - \partial_k F^{ik} = \partial_0 E^i + \epsilon^{ikl} \partial_k B^l = J^i$$

$$\rightarrow \partial_t \vec{E} + \nabla \times \vec{B} = \vec{J} \quad (11.10)$$

and for the time-component ($\rho \rightarrow 0$)

$$\underbrace{\partial_0 F^{00}}_{=0 \text{ see the definition of } F^{\rho\sigma}} - \underbrace{\partial_k F^{0k}}_{\substack{\text{Because } F^{\mu\nu} = -F^{\nu\mu} \\ \text{Eq. 11.7}}} = \partial_k E^k = J^0$$

$$\rightarrow \nabla \cdot \vec{E} = J^0 \quad (11.11)$$

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 antisymmetric⁴

$$\tilde{F}^{\mu\nu} = \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma} \quad (11.12)$$

the derivative ∂_μ of this new object $\tilde{F}^{\mu\nu}$, which is called the dual electromagnetic tensor, vanishes:

$$\partial_\mu \tilde{F}^{\mu\nu} = \partial_\mu \epsilon^{\mu\nu\rho\sigma} (\partial_\sigma A_\rho - \partial_\rho A_\sigma) = 0. \quad (11.13)$$

This follows from the fact that if we contract two symmetric indices with two antisymmetric indices, the result is always zero⁵. This can be seen, focussing for brevity on the first term

$$\begin{aligned} \epsilon^{\mu\nu\rho\sigma} \partial_\mu \partial_\sigma A_\rho &= \frac{1}{2} (\epsilon^{\mu\nu\rho\sigma} \partial_\mu \partial_\sigma A_\rho + \epsilon^{\mu\nu\rho\sigma} \partial_\mu \partial_\sigma A_\rho) \\ &\stackrel{\text{Renaming dummy indices}}{=} \frac{1}{2} (\epsilon^{\mu\nu\rho\sigma} \partial_\mu \partial_\sigma A_\rho + \epsilon^{\sigma\nu\rho\mu} \partial_\sigma \partial_\mu A_\rho) \\ &\stackrel{\text{Because } \epsilon^{\mu\nu\rho\sigma} = -\epsilon^{\sigma\nu\rho\mu} \text{ and } \partial_\mu \partial_\sigma = \partial_\sigma \partial_\mu}{=} \frac{1}{2} (\epsilon^{\mu\nu\rho\sigma} \partial_\mu \partial_\sigma A_\rho - \epsilon^{\mu\nu\rho\sigma} \partial_\mu \partial_\sigma A_\rho) = 0 \quad \checkmark \end{aligned} \quad (11.14)$$

⁴ $\epsilon^{\mu\nu\rho\sigma}$ is the four-dimensional Levi-Civita symbol, which is defined in appendix B.5.5.

⁵ This is explained in appendix B.5.4.

Equally the second term is zero. The equations⁶

$$\partial_\mu \tilde{F}^{\mu\nu} = 0 \quad (11.15)$$

⁶ Plural because we have one equation for each component $\nu = 0, 1, 2, 3$.

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 time-component $\nu \rightarrow 0$:

$$\begin{aligned}
 0 &= \partial_\mu \tilde{F}^{\mu 0} \\
 &= \partial_\mu \epsilon^{\mu 0 \rho \sigma} F^{\rho \sigma} \\
 &= \partial_0 \underbrace{\epsilon^{00 \rho \sigma}}_{=0} F^{\rho \sigma} - \partial_i \epsilon^{i 0 \rho \sigma} F^{\rho \sigma} \\
 &= \underbrace{-\partial_i \epsilon^{i 0 j k} F^{j k}}_{\text{Because } \epsilon^{i 0 \rho \sigma} = 0 \text{ for } \rho=0 \text{ or } \sigma=0} \\
 &= \partial_i \epsilon^{0 i j k} F^{j k} \\
 &= \underbrace{-\partial_i \epsilon^{0 i j k} \epsilon^{l j k}}_{\text{See Eq. 11.8} \quad = 2\delta_{il}} B^l \\
 &= -2\partial_i \delta_{il} B^l = -2\partial_i B^i
 \end{aligned}$$

$$\Rightarrow \partial_i B^i = 0 \quad \text{or in vector notation} \quad \nabla \cdot \vec{B} = 0 \quad (11.16)$$

Analogously, we can take a look at the spatial-components $\nu \rightarrow i$ and derive

$$\nabla \times \vec{E} + \partial_t \vec{B} = 0 \quad (11.17)$$

This is the conventional form of the homogeneous Maxwell equation that is used in applications.

11.2 The Lorentz Force

We can use the connection between quantum and classical mechanics (the Ehrenfest theorem) we discovered in the last chapter 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 field⁷, see Eq. 8.23:

$$i\partial_t \Psi = \underbrace{\left(\frac{1}{2m} (\vec{p} - q\vec{A})^2 + q\Phi \right)}_{\equiv H} \Psi. \quad (11.18)$$

If we define the momentum⁸ of this system $\vec{\Pi} = \vec{p} - q\vec{A}$ we can write the Hamiltonian as

$$H = \frac{1}{2m} \vec{\Pi}^2 + q\Phi.$$

⁷ Derived from the Klein Gordan 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.

⁸ This is actually the momentum of the system following from invariance under translations using the Noether theorem $\frac{\partial L}{\partial \vec{x}} = \vec{\Pi}$.

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 this time carrying the partial time-derivative with us, because the operator we are going to look at is time dependent:

$$\begin{aligned}\frac{d}{dt} \langle \hat{O} \rangle &= \frac{1}{i} \langle [O, H] \rangle + \langle \frac{\partial O}{\partial t} \rangle \\ \rightarrow \frac{d}{dt} \langle \vec{\Pi} \rangle &= \frac{1}{i} \langle [\vec{\Pi}, H] \rangle + \langle \frac{\partial \vec{\Pi}}{\partial t} \rangle,\end{aligned}$$

We can see that $\frac{\partial \vec{\Pi}}{\partial t} \neq 0$ because A can depend on time. If we now put the explicit form of H into the equation we get

$$\begin{aligned}\rightarrow \frac{d}{dt} \langle \vec{\Pi} \rangle &= \frac{1}{i} \langle [\vec{\Pi}, \frac{1}{2m} \vec{\Pi}^2 + q\Phi] \rangle + \langle \frac{\partial \vec{\Pi}}{\partial t} \rangle \\ \rightarrow \frac{d}{dt} \langle \vec{\Pi} \rangle &= \frac{1}{i} \langle [\vec{\Pi}, \frac{1}{2m} \vec{\Pi}^2] \rangle + \underbrace{\frac{1}{i} \langle [\vec{\Pi}, q\Phi] \rangle}_{= \langle q \nabla \Phi \rangle} + \langle \frac{\partial \vec{\Pi}}{\partial t} \rangle.\end{aligned}$$

In the last step we use that $[\vec{\Pi}, q\Phi]$ can be computed analogously 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

$$[\Pi_i, \Pi_j] = -\frac{q}{i} \left(\frac{\partial A_j}{\partial x_i} - \frac{\partial A_i}{\partial x_j} \right) = -\frac{q}{i} \epsilon_{ijk} \underbrace{B_k}_{= \epsilon_{klm} \frac{\partial}{\partial x_l} A_m} \quad (11.19)$$

with the usual definition of the magnetic field $B = \nabla \times A$ written in index form. If we now define the speed⁹ 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

$$\begin{aligned}\frac{d}{dt} \langle \vec{\Pi} \rangle &= \langle q \nabla \Phi \rangle - \langle q(v \times B) \rangle + \underbrace{\langle \frac{\partial \vec{\Pi}}{\partial t} \rangle}_{= -q \frac{\partial \vec{A}}{\partial t}} \\ \frac{d}{dt} \langle \vec{\Pi} \rangle &= -q \langle (v \times B) \rangle + q \underbrace{\langle \nabla \Phi - \frac{\partial \vec{A}}{\partial t} \rangle}_{= \langle E \rangle \text{ see Eq. 11.5}},\end{aligned}$$

We finally arrive at

$$\frac{d}{dt} \langle \vec{\Pi} \rangle \equiv F_{\text{Lorentz}} = -q(\langle (v \times B) \rangle + \langle E \rangle). \quad (11.20)$$

This is the equation of motion that describes, when solved, the classical trajectory of a particle in an external electromagnetic field.

⁹ The momentum divided by the mass of the particle: $p = mv$

11.3 Coulomb Potential

We learned in Chap. 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 that¹⁰ $\partial_\mu A^\mu = 0$. This is called the Lorentz gauge. Using this gauge simplifies the inhomogeneous Maxwell equations to

$$\partial_\sigma(\partial^\sigma A^\rho - \partial^\rho A^\sigma) = \partial_\sigma \partial^\sigma A^\rho = J^\rho \quad (11.21)$$

We now have a look at the physical situation where a fixed, static charge is located in a 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

$$\partial_\sigma \partial^\sigma A^\rho = \partial_0 \partial_0 A^\rho - \partial_i \partial_i A^\rho = 0. \quad (11.22)$$

We now use that we are considering a static ($\partial_0 A^\rho = 0$), symmetric system, by rewriting the equation using spherical coordinates¹¹. Then we can neglect all terms¹² but the term involving the ∂_r derivative. This yields

$$\rightarrow \frac{\partial^2}{\partial r^2}(rA^\mu) = 0. \quad (11.23)$$

The general solution of this equation is

$$A^\mu = \epsilon^\mu \frac{C}{r} + \epsilon^\mu D \quad (11.24)$$

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 component¹³ of A^μ this is the famous Coulomb potential

$$A^0 = \Phi = \frac{C}{r} = \frac{Ze}{r}, \quad (11.25)$$

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 angles 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.

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

Further Reading Tips

- **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

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}. \quad (12.1)$$

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 tensor¹ $T_{\mu\nu}$. 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 straightforward². How? 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. 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)

$$\partial^\mu T_{\mu\nu} = 0. \quad (12.2)$$

Next we need something to describe curvature mathematically.

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 of course much faster.

³ Up to this point we were only considering the Minkowski metric

$$\eta^{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \text{ and the}$$

$$\text{Euclidean metric } \delta^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

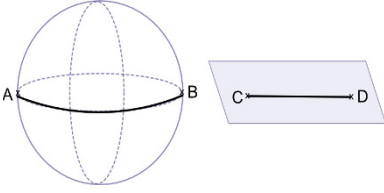


Fig. 12.1: Distance between two points in a curved and a flat space

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.

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 points³. In a curved space the distance between two points is different than in a flat space as illustrated in Fig. 12.1 and 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-free⁴ function of the metric $g_{\mu\nu}$ and at most their first and second partial derivatives. 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

$$T_{\mu\nu} = CG_{\mu\nu} \quad \text{that} \quad \partial^\mu T_{\mu\nu} = 0 \rightarrow \partial^\mu G_{\mu\nu} = 0 \quad (12.3)$$

must hold, too. The Einstein tensor is a second rank tensor⁵ 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 $R = R^\nu_\nu$

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \quad (12.4)$$

where the Ricci Tensor $R_{\mu\nu}$ is defined in terms of the Christoffel symbols $\Gamma^\mu_{\nu\rho}$

$$R_{\alpha\beta} = \partial_\rho \Gamma^\rho_{\beta\alpha} - \partial_\beta \Gamma^\rho_{\rho\alpha} + \Gamma^\rho_{\rho\lambda} \Gamma^\lambda_{\beta\alpha} - \Gamma^\rho_{\beta\lambda} \Gamma^\lambda_{\rho\alpha} \quad (12.5)$$

and the Christoffel Symbols are defined in terms of the metric

$$\Gamma_{cab} = \frac{1}{2} \left(\frac{\partial g_{ca}}{\partial x^b} + \frac{\partial g_{cb}}{\partial x^a} - \frac{\partial g_{ab}}{\partial x^c} \right) = \frac{1}{2} (\partial_b g_{ca} + \partial_a g_{cb} - \partial_c g_{ab}). \quad (12.6)$$

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 space-time. 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**

$$\frac{d^2 x^\lambda}{dt^2} + \Gamma^\lambda_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} = 0. \quad (12.7)$$

The geodesic is the locally shortest⁶ curve between two points on a manifold.

It is interesting to note that Einstein thought about the Christoffel Symbols as the gravitational field⁷

If the $\Gamma_{\nu\rho}^{\mu}$ 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 Einstein⁸

We can understand what Einstein means by looking at Eq. 12.7. For $\Gamma_{\nu\rho}^{\mu} = 0$ the geodesic equation reduces to

$$\frac{d^2 x^{\lambda}}{dt^2} = 0. \quad (12.8)$$

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

$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}. \quad (12.9)$$

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 able⁹ to compare them.

The derivative becomes in a curved space the **covariant derivative**¹⁰

$$D_b v^a \equiv \partial_b v^a + \Gamma_{bc}^a v^c. \quad (12.10)$$

Therefore, if we want any equation we derived so far to be valid in curved spacetime, we need to change

$$\partial_b \rightarrow D_b = \partial_b + \Gamma_{bc}^a. \quad (12.11)$$

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 in the prescription

$$\partial_{\mu} \rightarrow D_{\mu} \equiv \partial_{\mu} + ieA_{\mu}. \quad (12.12)$$

⁶ This is a bit oversimplified, but the correct definition needs some terms from differential geometry we haven't introduced here.

⁷ Others prefer to think of the metric as the gravitational field.

⁸ Albert Einstein. The foundation of the general theory of relativity. 1916

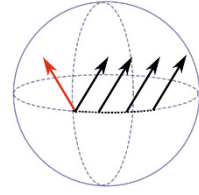


Fig. 12.2: In order to be able to compare the red arrow with the black arrow, we transport the black arrow to the location of the red arrow.

⁹ In fact in differential geometry one has only local coordinate systems. The defining feature of a manifold is that it's locally flat=Euclidean, as explained in Sec. 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.

¹⁰ In this context the Christoffel symbols Γ_{bc}^a are often called connection coefficients, because of their property to connect the points we want to compare.

Remember that this wasn't just a mathematical gimmick. This prescription gives us the correct theory of electromagnetism. The same is true for the weak

$$\partial_\mu \rightarrow D_\mu \equiv \partial_\mu + ig \underbrace{W_\mu}_{=W_\mu^i \sigma^i} \quad (12.13)$$

and the strong interaction (see Eq. 7.152)

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig' \underbrace{G_\mu}_{=T^C G_\mu^C} . \quad (12.14)$$

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 Chap. 10 that the equation of motion for the average value is just the classical one and no quantum effects are measured.

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.

¹¹ Albert Einstein and Francis A. Davis. *The Principle of Relativity*. Dover Publications, reprint edition, 6 1952. ISBN 9780486600819

We are able to 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

$$G_{\alpha\beta} = (\delta_{\alpha}^{\gamma}\delta_{\beta}^{\zeta} - \frac{1}{2}g_{\alpha\beta}g^{\gamma\zeta})(\partial_{\epsilon}\Gamma_{\gamma\zeta}^{\epsilon} - \partial_{\zeta}\Gamma_{\gamma\epsilon}^{\epsilon} + \Gamma_{\epsilon\sigma}^{\epsilon}\Gamma_{\gamma\zeta}^{\sigma} - \Gamma_{\zeta\sigma}^{\epsilon}\Gamma_{\epsilon\gamma}^{\sigma}). \quad (12.15)$$

Thus maybe we can think of the Einstein equation as the field equation¹² for $\Gamma_{\zeta\sigma}^{\epsilon}$, and the terms generated by the prescription $\partial_b \rightarrow D_b = \partial_b + \Gamma_{bc}^a$ as the corresponding coupling between the gravitational field $\Gamma_{\zeta\sigma}^{\epsilon}$ and the other fields?

¹² Analogous to the Maxwell equation for the electromagnetic field.

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, ... representation of the Poincare 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 working¹³ theory of quantum gravity and for further information have a look at the books mentioned in the next section.

¹³ Many attempts result in an infinite number of infinity terms, which is quite bad for probability predictions.

Further Reading Tips

For more information about the standard theory of gravity, Einstein's general relativity, see

- **Ta-Pei Cheng - Relativity, Gravitation and Cosmology**¹⁴ is a great, rather low-level introduction to general relativity with many very enlightening explanations. Perfect for getting a quick overview.
- **A. Zee - Einstein Gravity in a Nutshell**¹⁵, 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**¹⁶ is a really, really big book, but often offers in depth explanations for points that remain unclear in most other books.

¹⁴ Ta-Pei Cheng. *Relativity, Gravitation and Cosmology: A Basic Introduction*. Oxford University Press, 2nd edition, 1 2010. ISBN 9780199573646

¹⁵ Anthony Zee. *Einstein Gravity in a Nutshell*. Princeton University Press, 1st edition, 5 2013. ISBN 9780691145587

¹⁶ 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

¹⁷ John C. Baez and Javier P. Muniain. *Gauge Fields, Knots, and Gravity*. World Scientific Pub Co Inc, 1st edition, 9 1994. ISBN 9789810220341

- **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 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

$$\vec{v} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \quad \vec{w} = \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} \quad \rightarrow \quad \vec{v} + \vec{w} = \begin{pmatrix} v_x + w_x \\ v_y + w_y \\ v_z + w_z \end{pmatrix} \quad (\text{A.1})$$

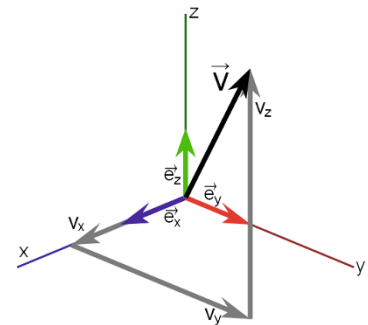
or multiplied

$$\vec{v} \cdot \vec{w} = \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \cdot \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} = v_x w_x + v_y w_y + v_z w_z. \quad (\text{A.2})$$

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:

$$\text{length}(\vec{v}) = \sqrt{\vec{v} \cdot \vec{v}}. \quad (\text{A.3})$$

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:

$$\begin{pmatrix} T \\ P \\ H \end{pmatrix}. \quad (\text{A.4})$$

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 perspective¹. 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.

¹ This will be made explicit in a moment.

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 independent² 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:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (\text{A.5})$$

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 vector³ \vec{w} we introduced above 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

$${}^3\vec{w} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}$$

$$\tilde{\vec{e}}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \tilde{\vec{e}}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \quad \tilde{\vec{e}}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \quad (\text{A.7})$$

In this basis the vector \vec{w} looks quite different:

$$\vec{w} = 2\sqrt{2}\tilde{\vec{e}}_1 - 2\sqrt{2}\tilde{\vec{e}}_2 + 0\tilde{\vec{e}}_3 = 2\sqrt{2}\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} - 2\sqrt{2}\frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix}. \quad (\text{A.8})$$

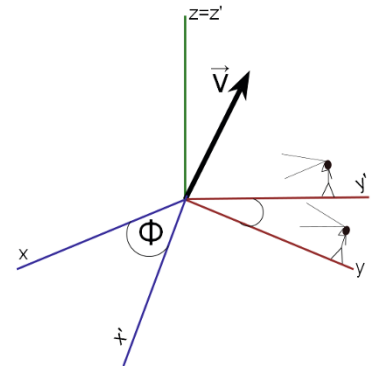
Therefore we can write \vec{w} in terms of components with respect to this new basis as

$$\tilde{\vec{w}} = \begin{pmatrix} 2\sqrt{2} \\ -2\sqrt{2} \\ 0 \end{pmatrix}.$$

This is not a different vector just a different description! To be precise, $\tilde{\vec{w}}$ 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

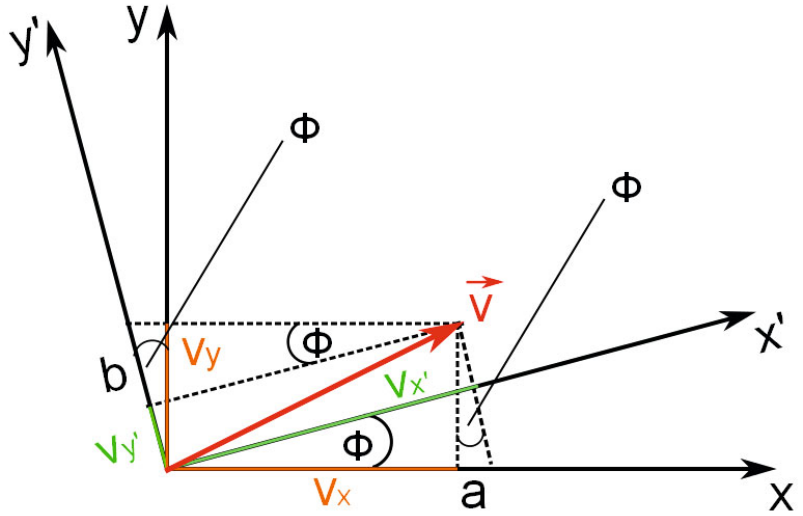
The connection between different coordinate systems can be made precise through the use of matrices. Two different coordinate systems can mean that we have two different observers that look at our experiment from different perspectives or this can simply mean that **one** observer decides to use a different 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.

Fig. A.1: Illustration of the components of a vector in two different coordinate systems. Details can be found in the text.



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

$$\begin{aligned} v_{x'} &= (v_x + v_y \tan(\phi)) \cos(\phi) = \left(v_x + v_y \frac{\sin(\phi)}{\cos(\phi)} \right) \cos(\phi) \\ &= v_x \cos(\phi) + v_y \sin(\phi). \end{aligned}$$

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 $\sin^2(\phi) + \cos^2(\phi) = 1$

$$\begin{aligned} v_{y'} &= v_y \frac{1}{\cos(\phi)} - v_{x'} \tan(\phi) = v_y \frac{1}{\cos(\phi)} - (v_x \cos(\phi) + v_y \sin(\phi)) \frac{\sin(\phi)}{\cos(\phi)} \\ &= v_y \frac{\sin^2(\phi) + \cos^2(\phi)}{\cos(\phi)} - v_x \sin(\phi) - v_y \frac{\sin^2(\phi)}{\cos(\phi)} = v_y \cos(\phi) - v_x \sin(\phi) \end{aligned}$$

Therefore $v_{y'} = -v_x \sin(\phi) + v_y \cos(\phi)$.

We can write this using a rotation matrix:

$$\begin{aligned} \begin{pmatrix} v_{x'} \\ v_{y'} \\ v_{z'} \end{pmatrix} &= R_z(\phi) \vec{v} = \begin{pmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} \\ &= \begin{pmatrix} \cos(\phi)v_x + \sin(\phi)v_y \\ -\sin(\phi)v_x + \cos(\phi)v_y \\ v_z \end{pmatrix}. \end{aligned} \quad (\text{A.9})$$

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

$$\vec{v} \cdot \vec{w} = \vec{v}^T \vec{w} = \begin{pmatrix} v_x & v_y & v_z \end{pmatrix} \begin{pmatrix} w_x \\ w_y \\ w_z \end{pmatrix} = v_x w_x + v_y w_y + v_z w_z, \quad (\text{A.10})$$

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

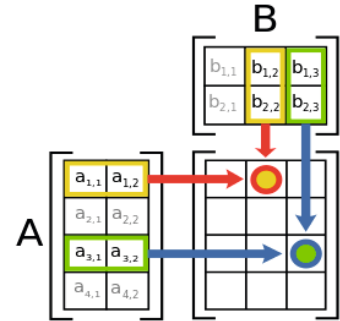


Fig. A.2: Schematic matrix multiplication. The important thing to keep in mind is **row times column**. The first index denotes the row number, the second the column number. In the example, the red element of the product matrix is $c_{1,2} = a_{1,1}b_{1,2} + a_{1,2}b_{2,2}$ and the blue element is $c_{3,3} = a_{3,1}b_{1,3} + a_{3,2}b_{2,3}$. In general $c_{i,j} = a_{i,k}b_{k,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \dots$. 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

$$M = \begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix} \quad N = \begin{pmatrix} 0 & 1 \\ 4 & 8 \end{pmatrix}$$

$$MN = \begin{pmatrix} 2 & 3 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 4 & 8 \end{pmatrix} = \begin{pmatrix} 2 \cdot 0 + 3 \cdot 4 & 2 \cdot 1 + 3 \cdot 8 \\ 1 \cdot 0 + 0 \cdot 4 & 1 \cdot 1 + 0 \cdot 8 \end{pmatrix} = \begin{pmatrix} 12 & 26 \\ 0 & 1 \end{pmatrix} \quad (\text{A.11})$$

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

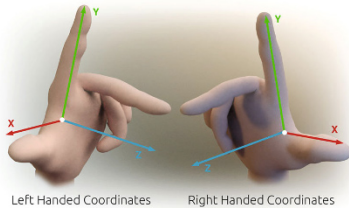


Fig. A.3: Right-handed and left-handed coordinate system. Figure by Primalshell (Wikimedia Commons) released under a CC-BY-SA-3.0 licence: <http://creativecommons.org/licenses/by-sa/3.0/deed.en>. URL: http://commons.wikimedia.org/wiki/File:3D_Cartesian_Coordinate_Handedness.jpg, Accessed: 1.12.2014

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

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

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \rightarrow \begin{pmatrix} -v_1 \\ -v_2 \\ -v_3 \end{pmatrix}, \quad (\text{A.12})$$

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

$$\vec{v} \rightarrow \vec{v}' = P\vec{v} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} -v_1 \\ -v_2 \\ -v_3 \end{pmatrix}. \quad (\text{A.13})$$

B

Calculus

B.1 Product Rule

The product rule

$$\frac{d(f(x)g(x))}{dx} = \left(\frac{df(x)}{dx}\right)g(x) + f(x)\left(\frac{dg(x)}{dx}\right) \equiv f'g + fg' \quad (\text{B.1})$$

follows directly from the definition of derivatives

$$\begin{aligned} \frac{d}{dx}[f(x)g(x)] &= \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{[f(x+h)g(x+h) - f(x+h)g(x)] + [f(x+h)g(x) - f(x)g(x)]}{h} \\ &= \lim_{h \rightarrow 0} f(x+h) \frac{g(x+h) - g(x)}{h} + g(x) \frac{f(x+h) - f(x)}{h} \\ &= f(x)g'(x) + g(x)f'(x) \end{aligned}$$

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 Wong¹

¹ Told on www.math.stackexchange.com

An important rule for integrals follows directly from the product rule. Integrating the product rule²

² See Eq. B.1 and we then use for the first term the fundamental theorem of calculus $\int_a^b dx h'(x) = h(b) - h(a)$.

$$\underbrace{\int_a^b dx \frac{d(f(x)g(x))}{dx}}_{=f(x)g(x)\big|_a^b} = \int_a^b dx \left(\frac{df(x)}{dx} \right) g(x) + \int_a^b dx f(x) \left(\frac{dg(x)}{dx} \right) \quad (\text{B.2})$$

and rearranging the terms yields

$$\int_a^b dx \left(\frac{df(x)}{dx} \right) g(x) = f(x)g(x)\big|_a^b - \int_a^b dx f(x) \left(\frac{dg(x)}{dx} \right). \quad (\text{B.3})$$

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 infinity³.

³ We discover in Sec. 2.4 that nothing can move faster than the speed of light. Therefore the field configuration infinitely far away mustn't have any influence on physics at finite x .

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

$$f(x) = f(a) + (x-a)f'(a) + \frac{1}{2}(x-a)^2 f''(a) + \dots \quad (\text{B.4})$$

- 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 write⁴

$$f(y + \Delta y) = f(y) + (y + \Delta y - y)f'(y) + \dots = f(y) + \Delta y f'(y) + \dots \quad (\text{B.5})$$

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 $\Delta y \rightarrow \epsilon$, the change of the function can be written by one (the linear) term of the Taylor series.

$$\Delta f = f(y + \epsilon) - f(y) = f(y) + \epsilon f'(y) + \dots - f(y) = \epsilon f'(y) + \underbrace{\dots}_{\approx 0 \text{ for } \epsilon^2 \approx 0} \quad (\text{B.6})$$

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 $a = y$.

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

$$\int_a^x dt f'(t) = f(x) - f(a) \rightarrow f(x) = f(a) + \int_a^x dt f'(t). \quad (\text{B.7})$$

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 $\int_a^b v'u = vu|_a^b + \int v u'$ 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

$$f(x) = f(a) + \int_a^x dt f'(t) = f(a) + g(t)f(t)\Big|_a^x - \int_a^x dt g(t)f''(t). \quad (\text{B.8})$$

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 useful⁵ 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

$$g(t)f(t)\Big|_a^x = (t - x)f(t)\Big|_a^x = \underbrace{(x - x)}_{=0}f(x) - (a - x)f(a) = (x - a)f(a) \quad (\text{B.9})$$

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 with⁶ $v' = (x - t)$ and $u = f''(t)$:

$$\rightarrow \int_a^x dt \underbrace{(x - t)}_{=v'} \underbrace{f''(t)}_{=u} = \underbrace{-\frac{1}{2}(x - t)^2}_{=v} \underbrace{f''(t)}_{=u} \Big|_a^x - \int_a^x dt \underbrace{\left(-\frac{1}{2}(x - t)^2\right)}_{=v} \underbrace{f'''(t)}_{=u'}$$

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

$$f(x) = f(a) + (x - a)f'(a) + \frac{1}{2}(x - a)^2 f''(a) + \int_a^x dt \frac{1}{2}(x - t)^2 f'''(t). \quad (\text{B.11})$$

⁵ The equation holds for arbitrary c and of course you're free to choose something different, but you won't get our formula. We choose the constant c such that we get a useful formula for $f(x)$. Otherwise $f(x)$ would appear on the left- and right-hand side.

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.

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 :

$$\begin{aligned} f(x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)(x-a)^n}{n!} \\ &= \frac{f^{(0)}(a)(x-a)^0}{0!} + \frac{f^{(1)}(a)(x-a)^1}{1!} + \frac{f^{(2)}(a)(x-a)^2}{2!} \\ &\quad + \frac{f^{(3)}(a)(x-a)^3}{3!} + \dots, \end{aligned} \quad (\text{B.12})$$

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

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}, \quad (\text{B.13})$$

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:

$$e^x = \sum_{n=0}^{\infty} \frac{e^0(x-0)^n}{n!} = \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad (\text{B.14})$$

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 uneven n vanishes, which we can use if we split the sum. Observe that

$$\begin{aligned} \sum_{n=0}^{\infty} n &= \sum_{n=0}^{\infty} (2n+1) + \sum_{n=0}^{\infty} (2n) \\ 1+2+3+4+5+6\ldots &= 1+3+5+\ldots + 2+4+6+\ldots \quad (\text{B.15}) \end{aligned}$$

Splitting the sum for $\sin(x)$ yields

$$\begin{aligned} \sin(x) &= \sum_{n=0}^{\infty} \frac{\sin^{(2n+1)}(0)(x-0)^{2n+1}}{(2n+1)!} + \underbrace{\sum_{n=0}^{\infty} \frac{\sin^{(2n)}(0)(x-0)^{2n}}{(2n)!}}_{=0} \\ &= \sum_{n=0}^{\infty} \frac{\sin^{(2n+1)}(0)(x-0)^{2n+1}}{(2n+1)!}. \end{aligned} \quad (\text{B.16})$$

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

$$\begin{aligned} \sin(x)^{(1)} &= \cos(x) \\ \sin(x)^{(2)} &= \cos'(x) = -\sin(x) \\ \sin(x)^{(3)} &= -\sin'(x) = -\cos(x) \\ \sin(x)^{(4)} &= -\cos'(x) = \sin(x) \\ \sin(x)^{(5)} &= \sin'(x) = \cos(x) \end{aligned} \quad (\text{B.17})$$

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 formula⁷. We can therefore rewrite Eq. B.16 as

$$\begin{aligned} {}^7 \sin^{(1)}(x) &= \sin^{(2 \cdot 0 + 1)}(x) = \\ (-1)^0 \cos(x) &= \cos(x), \sin^{(3)}(x) = \\ \sin^{(2 \cdot 1 + 1)}(x) &= (-1)^1 \cos(x) = -\cos(x) \end{aligned}$$

$$\begin{aligned} \sin(x) &= \sum_{n=0}^{\infty} \frac{\sin^{(2n+1)}(0)(x-0)^{2n+1}}{(2n+1)!} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n \cos(0)(x-0)^{2n+1}}{(2n+1)!} \\ &\stackrel{\cos(0)=1}{=} \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n+1}}{(2n+1)!} \end{aligned} \quad (\text{B.18})$$

This is the Taylor series for $\sin(x)$, which again can be seen as a definition of $\sin(x)$. Analogous we can derive

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n}}{(2n)!}, \quad (\text{B.19})$$

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

$$\begin{aligned} \sum_{n=0}^{\infty} n &= \sum_{n=0}^{\infty} (2n+1) + \sum_{n=0}^{\infty} (2n) \\ 1 + 2 + 3 + 4 + 5 + 6 \dots &= 1 + 3 + 5 + \dots + 2 + 4 + 6 + \dots, \end{aligned} \quad (\text{B.20})$$

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 ?

$$\begin{aligned} e^{ix} &= \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(ix)^{2n}}{(2n)!} + \sum_{n=0}^{\infty} \frac{(ix)^{2n+1}}{(2n+1)!} \end{aligned} \quad (\text{B.21})$$

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

$$\begin{aligned} e^{ix} &= \sum_{n=0}^{\infty} \frac{(ix)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n}}{(2n)!} + i \sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n+1}}{(2n+1)!} \\ &= \underbrace{\sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n}}{(2n)!}}_{=\cos(x) \text{ see Eq. B.19}} + i \underbrace{\sum_{n=0}^{\infty} \frac{(-1)^n (x)^{2n+1}}{(2n+1)!}}_{=\sin(x) \text{ see Eq. B.18}} \\ &= \cos(x) + i \sin(x) \end{aligned} \quad (\text{B.22})$$

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

$$a_n b_n \equiv \sum_n a_n b_n. \quad (\text{B.23})$$

Other examples are

$$a_n b_n c_m \equiv \sum_n a_n b_n c_m. \quad (\text{B.24})$$

$$a_m b_n c_m \equiv \sum_m a_m b_n c_m. \quad (\text{B.25})$$

but

$$a_n b_m \neq \sum_n a_n b_m, \quad (\text{B.26})$$

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 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 nothing⁸, as long as n is contracted

$$a_n b_n c_m = a_k b_k c_m \equiv \sum_n a_n b_n c_m \equiv \sum_k a_k b_k c_m. \quad (\text{B.27})$$

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

$$F_i = \epsilon_{ijk} a_j b_k. \quad (\text{B.28})$$

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

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.

at $\epsilon_{ijk}a_jb_k$ we can change the names of j and k as we like, because these indices are contracted. For example $j \rightarrow u, k \rightarrow z$, which yields $\epsilon_{iuz}a_ub_z$ is really the same. On the other hand i is not a dummy index and we can't rename it $i \rightarrow m$: $\epsilon_{muz}a_ub_z$, because then our equation would read

$$F_i = \epsilon_{muz}a_ub_z. \quad (\text{B.29})$$

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

$$M_{ij} \equiv \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}. \quad (\text{B.30})$$

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

$$MN \equiv (MN)_{ij} = M_{ik}N_{kj}. \quad (\text{B.31})$$

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

$$\begin{pmatrix} 9 & 3 \\ 3 & 17 \end{pmatrix} \quad (\text{B.32})$$

A matrix is called **totally antisymmetric** if $M_{ij} = -M_{ji}$ for **all** i, j holds. An example would be.

$$\begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix} \quad (\text{B.33})$$

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 antisymmetric in the same indices, the result is zero:

$$\sum_{ij} a_{ij} b_{ij} = 0 \quad (\text{B.34})$$

if $a_{ij} = -a_{ji}$ and $b_{ij} = b_{ji}$ holds for all i, j . We can see this by writing

$$\sum_{ij} a_{ij} b_{ij} = \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} + \sum_{ij} a_{ij} b_{ij} \right) \quad (\text{B.35})$$

As explained earlier we are free to rename our dummy indices $i \rightarrow j$ and $j \rightarrow i$, which we use in the second term

$$\rightarrow \sum_{ij} a_{ij} b_{ij} = \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} + \sum_{ij} a_{ji} b_{ji} \right) \quad (\text{B.36})$$

Then we use the symmetry of b_{ij} and antisymmetry of a_{ij} , to switch the indices in the second term, which yields⁹

$$\begin{aligned} \rightarrow \sum_{ij} a_{ij} b_{ij} &= \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} + \sum_{ij} \underbrace{a_{ji}}_{=-a_{ij}} \underbrace{b_{ji}}_{=b_{ij}} \right) \\ &= \frac{1}{2} \left(\sum_{ij} a_{ij} b_{ij} - \sum_{ij} a_{ij} b_{ij} \right) = 0 \end{aligned} \quad (\text{B.37})$$

⁹ If this looks like a cheap trick to you, 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

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (\text{B.38})$$

In index notation the unit matrix is called the Kronecker symbol, denoted δ_{ij} , which is then defined for arbitrary dimensions by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (\text{B.39})$$

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 by

$$\epsilon_{ij} = \begin{cases} 1 & \text{if } (i, j) = \{(1, 2)\} \\ 0 & \text{if } i = j \\ -1 & \text{if } (i, j) = \{(2, 1)\} \end{cases} \quad (\text{B.40})$$

In matrix form

$$\epsilon_{ij} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (\text{B.41})$$

In three dimensions the Levi-Civita symbol is

$$\epsilon_{ijk} = \begin{cases} 1 & \text{if } (i, j, k) = \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\} \\ 0 & \text{if } i = j \text{ or } j = k \text{ or } k = i \\ -1 & \text{if } (i, j, k) = \{(1, 3, 2), (3, 2, 1), (2, 1, 3)\} \end{cases} \quad (\text{B.42})$$

and in four dimensions

$$\epsilon_{ijkl} = \begin{cases} 1 & \text{if } (i, j, k, l) \text{ is an even permutation of } \{(1, 2, 3, 4)\} \\ -1 & \text{if } (i, j, k, l) \text{ is an uneven permutation of } \{(1, 2, 3, 4)\} \\ 0 & \text{otherwise} \end{cases} \quad (\text{B.43})$$

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:

$$\epsilon_{ijk} = -\epsilon_{jik}, \epsilon_{ijk} = -\epsilon_{ikj} \text{ etc. or in two dimensions } \epsilon_{ij} = -\epsilon_{ji}.$$

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 by

$$M_{ij}^* = \begin{pmatrix} M_{11}^* & M_{12}^* \\ M_{21}^* & M_{22}^* \end{pmatrix}, \quad (\text{C.1})$$

which means we simply take the complex conjugate of each element¹.

The **transpose of a matrix** is defined by $M_{ij}^T = M_{ji}$, in matrix form

$$M_{ij} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \rightarrow M_{ij}^T = \begin{pmatrix} M_{11} & M_{21} \\ M_{12} & M_{22} \end{pmatrix}, \quad (\text{C.2})$$

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

$$MN \equiv (MN)_{ij} = M_{ik} N_{kj}$$

$$(MN)^T \equiv ((MN)_{ij})^T = (MN)_{ji} = (M_{ik} N_{kj})^T$$

$$(M_{ik} N_{kj})^T = M_{ik}^T N_{kj}^T = M_{ki} N_{jk} = N_{jk} M_{ki} \equiv N^T M^T, \quad (\text{C.3})$$

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 $z^* = a - ib$.

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 this series

$$e^M = \sum_{n=0}^{\infty} \frac{M^n}{n!}. \quad (\text{C.4})$$

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 invertible². 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

$$\det(A) = \sum_{i=1}^3 \sum_{j=1}^3 \sum_{k=1}^3 \epsilon_{ijk} A_{1i} A_{2j} A_{3k} \quad (\text{C.5})$$

and analogously for n-dimensions

$$\det(A) = \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_n=1}^n \epsilon_{i_1 i_2 \dots i_n} A_{1i_1} A_{2i_2} \dots A_{ni_n}. \quad (\text{C.6})$$

It is instructive to look at an explicit example in two dimensions:

$$\det(A) = \det \begin{pmatrix} 3 & 1 \\ 5 & 2 \end{pmatrix} = (3 \cdot 2) - (5 \cdot 1) = 1$$

² A matrix M is invertible, if we can find an inverse matrix, denoted by M^{-1} , with $M^{-1}M = 1$.

Or for a general three dimensional matrix

$$\det \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} = a_1(b_2c_3 - b_3c_2) - a_2(b_1c_3 - b_3c_1) + a_3(b_1c_2 - b_2c_1) \quad (\text{C.7})$$

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

$$M\vec{v} = \lambda\vec{v}. \quad (\text{C.8})$$

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 form³

$$M = N^{-1}DN, \quad (\text{C.9})$$

where the matrix N consists of the eigenvectors as its column and D is diagonal with the eigenvalues of M on its diagonal:

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} = N^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} N = (\vec{v}_1, \vec{v}_2)^{-1} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} (\vec{v}_1, \vec{v}_2) \quad (\text{C.10})$$

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 vectors¹ ($\vec{e}_1, \vec{e}_2, \vec{e}_3$). In ordinary Euclidean space the most common choice is

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (\text{D.1})$$

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 functions². 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

$$f(x) = \sum_{k=0}^{\infty} (a_k \cos(kx) + b_k \sin(kx)) \quad (\text{D.3})$$

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 sum³

¹ This is explained in more detail in appendix A.1.

² In a more abstract sense, functions are abstract vectors. This is meant in the sense that functions are elements of some vector space. For different kinds of functions a different vector space. Such abstract vector spaces are defined similar to the usual Euclidean vector space, where our ordinary position vectors live (those with the little arrow $\vec{}$). For example, take note that we can add two functions, just as we can add two vectors, and get another function. In addition, it's possible to define a scalar product.

³ Recall that an integral is just the limit of a sum, where the discrete k in \sum_k becomes a continuous variable in $\int dk$.

$$f(x) = \int_0^\infty dk (a_k e^{ikx} + b_k e^{-ikx}), \quad (\text{D.4})$$

which we can also write as

$$f(x) = \int_{-\infty}^\infty dk f_k e^{-ikx}. \quad (\text{D.5})$$

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 delta⁴ is to sums. We can use the Kronecker delta δ_{ij} to pick one specific term of any sum. For example, consider

$$\sum_{i=1}^3 a_i b_j = a_1 b_j + a_2 b_j + a_3 b_j \quad (\text{D.6})$$

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

$$\sum_{i=1}^3 \delta_{2i} a_i b_j = \underbrace{\delta_{21}}_{=0} a_1 b_j + \underbrace{\delta_{22}}_{=1} a_2 b_j + \underbrace{\delta_{23}}_{=0} a_3 b_j = a_2 b_j. \quad (\text{D.7})$$

Or more general

$$\sum_{i=1}^3 \delta_{ik} a_i b_j = a_k b_j. \quad (\text{D.8})$$

The delta distribution $\delta(x - y)$ is defined by

$$\int dx f(x) \delta(x - y) = f(y). \quad (\text{D.9})$$

Completely analogous to the Kronecker delta, the delta distribution picks one term⁵ from the integral. In addition, we can use this analogy to motivate from the equality

$$\frac{\partial x_i}{\partial x_j} = \delta_{ij} \quad (\text{D.10})$$

the equality

$$\frac{\partial f(x_i)}{\partial f(x_j)} = \delta(x_i - x_j). \quad (\text{D.11})$$

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.

⁴ The Kronecker delta is defined in appendix B.5.5.

⁵ The term where $x = y$. For example, $\int dx f(x) \delta(x - 2) = f(2)$.

Bibliography

Ian J.R. Aitchison and Anthony J.G. Hey. *Gauge Theories in Particle Physics*. CRC Press, 4th edition, 12 2012. ISBN 9781466513174.

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.

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.

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.

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.

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.

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 2008a. ISBN 978-0387782140.

John Stillwell. *Naive Lie Theory*. Springer, 1st edition, 8 2008b. ISBN 9780387782140.

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, 92
- adjoint representation, 164
- angular momentum, 100
- annihilation operator, 210
- anticommutator, 214

- B field, 233
- Baker-Campbell-Hausdorff formula, 40
- basis generators, 42
- basis spinors, 213
- Bohmian mechanics, 193
- boost, 62
- boost x-axis matrix, 64
- boost y-axis matrix, 64
- boost z-axis matrix, 64
- boson, 169
- bra, 185
- broken symmetry, 147

- Cartan generators, 53
- Casimir elements, 52
- Casimir operator $SU(2)$, 56
- charge conjugation, 81
- classical electrodynamics, 233
- Closure, 40
- commutator, 40
- commutator quantum field theory, 116
- commutator quantum mechanics, 114
- complex Klein-Gordon field, 119
- conjugate momentum, 108, 115, 207
- continuous symmetry, 26
- Coulomb potential, 237
- coupling constants, 119
- covariance, 21, 118
- covariant derivative, 134
- creation operator, 210
- dagger, 34, 87
- Dirac equation, 123
- Dirac equation solution, 213
- dotted index, 70
- doublet, 127, 139, 145
- Dyson series, 220

- E field, 233
- Ehrenfest theorem, 228
- Einstein, 11
- Einstein summation convention, 17
- electric charge, 137
- electric charge density, 137
- electric field, 233
- electric four-current, 137, 233
- electrodynamics, 233
- elementary particles, 85
- energy field, 104
- Energy scalar field, 208
- energy-momentum relation, 174
- Energy-Momentum tensor, 103
- Euclidean space, 20
- Euler-Lagrange equation, 95
 - field theory, 97
 - particle theory, 96
- expectation value, 177, 227
- extrema of functionals, 93

- Fermat's principle, 92
- fermion, 169
- fermion mass, 156
- Feynman path integral formalism, 193
- field energy, 104
- field momentum, 104
- field-strength tensor, 233
- four-vector, 18
- frame of reference, 5
- free Hamiltonian, 216

- functional, 92
- gamma matrices, 122
- gauge field, 145
- gauge symmetry, 129
- Gaussian wave-packet, 181
- Gell-Mann matrices, 165
- general relativity, 239
- generator, 38, 39
- generator boost x-axis, 63
- generator boost y-axis, 63
- generator boost z-axis, 63
- generator $SO(3)$, 42
- generators, 40
- generators $SU(2)$, 46
- global symmetry, 130
- gluons, 168
- Goldstone bosons, 149
- gravity, 239
- Greek indices, 17
- group axioms, 29
- groups, 26
- Hamiltonian, 100
- Hamiltonian scalar field, 209
- Heisenberg picture, 217
- Higgs potential, 147
- Higgs-field, 149
- homogeneity, 11
- homogeneous Maxwell equations, 137
- inertial frame, 11
- infinitesimal transformation, 38
- inhomogeneous Maxwell equations, 125, 134, 233
- interaction Hamiltonian, 216
- interaction picture, 218
- internal symmetry, 106
- invariance, 20, 118
- invariant of special-Relativity, 14
- invariant of special-relativity, 12
- invariant subspace, 52
- irreducible representation, 52
- isotropy, 11
- ket, 185
- Klein-Gordon equation, 207
- Klein-Gordon equation solution, 207
- Klein-Gordon equation, 119
- ladder operators, 54
- Lagrangian formalism, 93
- left-chiral spinor, 69
- Lie algebra
 - definition, 40
 - Lorentz group, 66
 - modern definition, 44
 - Poincare group, 84
 - $SO(3)$, 43
 - $SU(2)$, 46
- Lie bracket, 40
- local symmetry, 130
- Lorentz force, 235
- Lorentz group, 59
- Lorentz group components, 60
- Lorentz transformations, 19
- magnetic field, 233
- Majorana mass terms, 120
- mass-shell condition, 225
- matrix multiplication, 253
- metric, 18
- minimal coupling, 135
- Minkowski metric, 17
- momentum, 99
- momentum field, 104
- natural units, 5
- Newton's second law, 228
- Noether current, 105
- Noether's Theorem, 97
- non-continuous symmetry, 26
- operators of quantum mechanics, 174
- parity, 60, 78
- particle in a box, 181, 187
- Pauli matrices, 46
- Pauli-exclusion principle, 214
- Pauli-Lubanski four-vector, 85
- Poincare group, 29, 84
- principle of locality, 17
- principle of relativity, 11
- probabilistic interpretation, 176
- probability amplitude, 176
- Proca equation, 124
- Proca equation solution, 215
- quantum gravity, 244
- quantum mechanics, 173
- quantum operators, 174

- quaternions, 33
- relativity, 11
- representation
 - definition, 50
 - Lorentz group, 61
 - (0,0), 68
 - (0,0)(1/2,0), 68
 - (0,1/2), 69
 - (1/2,1/2), 75
 - spin 0, 86
 - spin 1, 86
 - spin 1/2, 86
 - SU(2), 53
- right-chiral spinor, 70
- rotation matrices, 33
- rotational symmetry, 26
- rotations in Euclidean space, 20
- rotations with quaternions, 36
- scalar product, 249
- scalar representation, 86
- scatter Amplitude, 216
- Schrödinger picture, 218
- Schur's Lemma, 52
- similarity transformation, 51
- SO(2), 29
- SO(3), 41
- special relativity, 11, 12
- spin, 85
- spin representation, 85
- spinor, 69
- spinor metric, 71
- spinor representation, 86
- Standard Model, 7
- strong interaction, 169
- SU(2), 35, 45, 139
- SU(3), 164
- SU(n), 87
- subrepresentation, 52
- sum over histories, 193
- summation convention, 17
- superposition, 176
- symmetry, 20
- symmetry group, 29
- time evolution of states, 216
- total derivative, 102
- trace, 41
- triplet, 165
- U(1), 31, 130
- unit complex number, 31
- unit complex numbers, 130
- unit quaternions, 33
- unitary gauge, 149
- vacuum state, 149
- vacuum value, 149
- Van der Waerden notation, 70
- variational calculus, 92
- vector representation, 86
- wave function, 176, 185
- Weyl spinor, 70, 79
- Yukawa coupling, 157