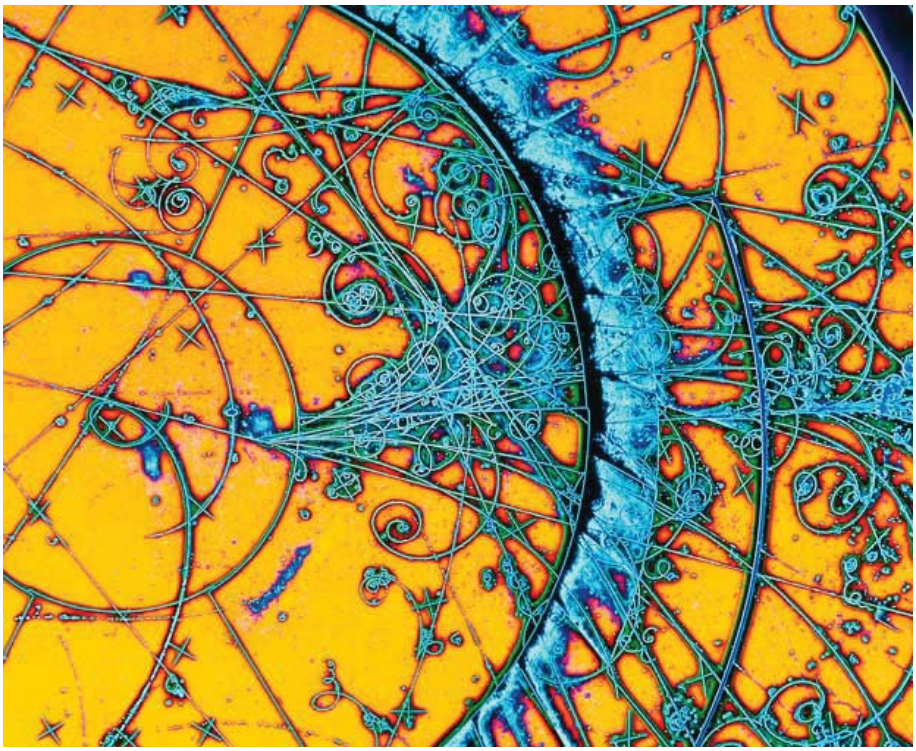
CHALMERS



Group Theory and Symmetries in Particle Physics

Bachelor thesis in Engineering Physics

Saladin Grebović, Axel Radnäs, Arian Ranjbar, Malin Renneby, Carl Toft and Erik Widén

Department of Fundamental Physics
Chalmers University of Technology
Gothenburg, 2012-06-08
Project FUFX02-12-04

Group Theory and Symmetries in Particle Physics

Authors: Saladin Grebović, Axel Radnäs, Arian Ranjbar, Malin Renneby, Carl Toft and Erik Widén

Contact:

saladin@student.chalmers.se
radnas@student.chalmers.se
ranjbar@student.chalmers.se
renneby@student.chalmers.se
tcarl@student.chalmers.se
erikwi@student.chalmers.se

FUFX02 - Bachelor Thesis at Fundamental Physics
Bachelor Thesis N° FUFX02-12-04

Supervisor: Daniel Persson, Division of Mathematical Physics
Examiner: Christian Forssén, Division of Subatomic Physics

Department of Fundamental Physics
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
+ 46 (0) 31 772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2012

Cover: *The figure on the front page displays an artistically enhanced collection of tracks left by subatomic particles passing through a bubble chamber at CERN.* [1]

Group Theory and Symmetries in Particle Physics

Saladin Grebović Axel Radnäs Arian Ranjbar
Malin Renneby Carl Toft Erik Widén

Supervisor:
Daniel Persson

Examiner:
Christian Forssén

June 8, 2012

Department of Fundamental Physics
Chalmers University of Technology

Abstract

This report presents a survey of symmetries and the various applications of their mathematical framework, group theory, in modern particle physics. The text is a result of literature studies and is of introductory character, chiefly aimed at undergraduate students and graduate students with some prior knowledge of quantum mechanics, special relativity and analytical mechanics who wish to learn elementary group theory. Hence, both the pedagogical nature of the text and the physical and mathematical reasoning are of importance. The survey is illustrated in the main part of this report in form of a book, constructed in such a manner that it potentially may serve as future course literature. In the first chapters of the book, mathematical formalism is presented which is required to understand the physical theories. The latter chapters focus on physical applications, especially the classification of elementary particles according to external and internal symmetries. Finally, an overview is provided of several occurrences of these symmetries in the realm of particle physics. Moreover, a discussion regarding the challenges and opportunities of this field with a teaching aspect is conducted as an evalutation of the project.

Sammanfattning

Den här rapporten är en övergripande undersökning av symmetrier och deras matematiska ramverk gruppteori, med tillämpningar inom modern partikelfysik. Texten är ett resultat av litteraturstudier och har ett introducerande upplägg. Den riktar sig främst till tredjeårsstudenter vid Teknisk Fysik, Chalmers Tekniska Högskola, samt studenter på motsvarande nivå som sedan tidigare är bekanta med kvantmekanik, speciell relativitetsteori och analytisk mekanik och som önskar lära sig om gruppteori. Därmed fästs stor vikt vid det pedagogiska inslaget utöver de fysikaliska och matematiska resonemangen. Resultatet redovisas i huvuddelen av rapporten skriven i form av en bok så att den potentiellt kan nyttjas som framtida kurslitteratur. De första kapitlen av boken behandlar den matematiska formalism som fordras för att kunna förstå den fysikaliska teorin. Därefter berör de senare kapitlen fysikaliska tillämpningar, speciellt klassificering av elementarpartiklar enligt externa och interna symmetrier. Slutligen ges en överblick över tillfällen då dessa symmetrier dyker upp i partikelfysik. Huvuddelen föregås av en diskussion gällande fältets svårigheter och möjligheter ur lärandesynpunkt som en utvärdering av arbetsprocessen.

Contents

Part I

Rapportdel

Bakgrund

Människan har sedan tidernas begynnelse varit fascinerad av symmetrier. Ibland kan man skönja fjälltopparnas silhuett speglad i sjöarnas blanka yta eller konstnärens strävan efter att skapa den perfekta mosaikbilden. Även inom matematiken har symmetrier länge studerats med intresse och beundran, som de antika grekernas speciella relation till sfären. I modern tid har begreppet symmetri fått en allt större betydelse inom fysikforskningen. Det är i sökandet efter dessa som vi försöker förstå vad som pågår på de kortaste avstånden, och på så sätt hur hela universum är uppbyggt. I och med kvantmekanikens intåg under 1900-talet och konstruktionen av kraftfulla partikelacceleratorer såsom LHC i CERN, Schweiz, har vi fått nya verktyg att utröna hur fenomen uppkommer och observera huruvida teoretiska förutsägelser stämmer. Dessutom har Einsteins relativitetsteori visat oss vägar att beskriva galaxers rörelser och möten mellan det stora och lilla ger oss spännande möjligheter att sätta vår verklighetsuppfattning på prov. När man talar om symmetrier i dessa sammanhang används ordet i en vidare mening än vardagliga speglings- och rotationssymmetrier hos olika objekt. Symmetri avser här någon viss typ av transformation, vilken egentligen kan utgöras av vilken matematisk operation som helst, som lämnar en generell egenskap oförändrad. Om en ny fysikalisk teori har uppvisat symmetriegenskaper har den ofta visat sig vara den empiriskt korrekta, vilket gör undersökningen av sådana särdrag till en vetenskaplig prioritet. När det gäller studier av symmetrier inom fysiken var några av de viktigaste förgrundsgestalterna den ungersk-amerikanske matematikern/fysikern Eugene Wigner (1902-1995) och den tyska matematikern Emmy Noether (1882-1935), varav den sistnämnda formulerade det som numera kallas Noethers teorem - vilket utgör en milstolpe i modern fysik.

Det matematiska ramverket för studier av symmetrier kallas för gruppteori. Fältet grundades under 1800-talet med viktiga insatser av de norska matematikerna Niels Henrik Abel (1802-1829) och Sophus Lie (1842-1899). Deras namn återkommer i begrepp som abelsk, liegrupper och liealgebra. Genom att tillgodogöra sig den utförliga formalismen är det möjligt att beskriva vitt skilda saker som olika virusmolekylers beteende i sökande efter effektiva läkemedel inom bioteknik eller kristallstrukturers karaktär inom det fasta tillståndets fysik. Vad detta arbete fokuserar på är dess enorma betydelse inom just partikelfysik. Gruppteori erbjuder nya pusselbitar till hur vi uppfattar kvantmekaniskt spinn, kvarkar och mesoner i subatomära experiment och fundamentet i dagens partikelfysik, standardmodellen, samt ger ledtrådar om vad som väntas bortom den. Vi avser därmed att undersöka detta spännande matematiska fält och se hur spinn och elementarpartiklar kan få en plats i ett större sammanhang.

Syfte

Syftet med detta kandidatarbete är att själva som grupp bekanta oss med gruppteori och symmetrier såsom de förekommer i fysiken, att lära oss vetenskapliga arbetsmetoder samt att producera en rapport i bokform som på bästa sätt kan sprida kunskapen till andra studenter och intresserade. Dessa förutsätts ha grundläggande kännedom kring kvantmekanik, speciell relativitetsteori och analytisk mekanik samt matematisk analys. Ett delmål under processen har också varit att uppmärksamma vår egen inlärningsprocess för att både utveckla vårt förhållningssätt till vetenskaplig litteratur och tidskriftsartiklar samt en pedagogisk vetenskaplig kommunikation.

Uppgiftsformulering

Uppgiften har bestått av att bilda oss en överblick över matematisk gruppteori och symmetriers roll i den nuvarande fysikaliska verklighetsbeskrivningen och på ett pedagogiskt sätt presentera kunskapen i en lärobok riktad till medstuderande. Målet för skriften har varit att den ska innehålla tillräckligt med underlag för att en medstuderande ska kunna nå samma nivå av kunskap inom ämnet som författarna har fått under våren. Ämnesstudierna har bestått av tre delar.

Förstå och beskriva:

(i) Grunderna i gruppteori utifrån det kvantmekaniska begreppet spinn

(ii) Uppdelningen av *externa* och *interna* symmetrier

(iii) Klassificeringen av partiklar utifrån ovanstående begrepp.

Dessutom introduceras geometriska och topologiska samband i nära anknytning till dessa symmetrier för att bidra till en djupare förståelse av gruppteorins samspel med verkligheten.

Metod

Arbetet har huvudsakligen bestått av litteraturstudier samt beräkning för hand och samtal kring olika definitioner, teorem och implikationer. Till följd av det omfattande materialet har huvudansvaret för de olika kapitlena i boken delats upp. För att fortfarande upprätthålla en bred kunskapsnivå har gruppen utöver vanliga diskussioner hållit längre föreläsningar för varandra. Därutöver har också redigeringen av varandras utkast gjorts korsvis.

Vidare har kontinuerliga möten och diskussioner med vår handledare spelat en stor roll. I genomsnitt har mötena ägt rum varannan till var tredje vecka under vårterminen. Inledningsvis hade dessa en mer föreläsningsbetonad karaktär med presentationer av de flesta begreppen. Under de senare mötena diskuterades främst upplägget av rapporten och specifika frågor på det material som bearbetades för tillfället. Kommunikation med handledare och inbördes inom gruppen har skett via mail, vår handledares kandidatprojektshemsida `http://www.danper.se/Daniels_homepage/GroupTheory.html` och en gemensam `Dropbox`-mapp med allt material och en kommentarsfil. I gruppen har vi därutöver haft veckovisa möten och extrainsatta tillfällen när det har varit gynnsamt under intensiva arbetsveckor.

En ytterligare värdefull del av metoden har varit möjligheten att närvara vid föreläsningar i en doktorandkurs i gruppteori och symmetrier inom fysik, given av Institutionen för Fundamental fysik under vårterminen.

Skrivarprocessen bistods av två möten med centrum för fackspråk och kommunikation, varav det första utgjordes av responsarbete tillsammans med grupp KBTX01-12-14.

Litteraturmässigt har vi utgått från *Group theory: A Physicist's Survey* av Pierre Ramond [2]. Därtill har vi framförallt hämtat information och inspiration från *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction* av Brian C. Hall, *Semi-Simple Lie Algebras And Their Representations* av Robert N. Cahn [3] och *Lie algebras in Particle Physics* av Howard Georgi [4]. Vidare har vi för beskrivning av spinn och rörelsemängdsmoment studerat kapitel 3 i *Modern Quantum Mechanics* av J.J. Sakurai och för diskussioner kring Lorentz-

och Poincarégruppen kapitel 2 i *The Quantum Theory of Fields vol. I - Foundations* av Steven Weinberg [5]. Kunskap kring konstruktion av gaugeteorier har främst hämtats från avsnitt 15.1-15.2 ur *An Introduction to Quantum Field Theory* av Michael E. Peskin och Daniel V. Schroeder [6], samt ur kapitel 7 och 8 i Weinberg [5]. Topologiska begrepp återges huvudsakligen från *Geometry, Topology and Physics* av Mieko Nakahara [7]. Därutöver redovisas andra viktiga, kompletterande källor löpande genom rapporten och finns sammanställda i referenslistan.

Figurer som illustrerar relationer, begrepp och förlopp har tillverkats med hjälp av `Adobe Illustrator`, `Mathematica`, `Paint` och `Xfig` enligt programmens förutsättningar för olika behov.

Avgränsning

Då vårt ämnesområde har sin ena fot i matematiken och den andra i den teoretiska fysiken uppstår en del svårigheter vad gäller detaljnivå och gestaltning. Allmänt kan sägas att matematiker är mer noggranna när det gäller definitioner än fysiker, som å andra sidan har fördelen av att de lättare kan ta till sig större mängder material och se samband i helheten. Vi har försökt att ta till oss de nya begreppen ur både matematiker- och fysikerperspektivet, men lutar åt det sistnämnda då vi annars omöjligen skulle ha kunnat genomföra mer än en bråkdel av projektet. Genomgående i arbetet används fysikerkonvention och fysikers förklaring av olika fenomen med matematiska inslag. Detaljnivån i inlärning och framställning har fördelats någorlunda jämnt och har valts så att ämnesbredden kunnat sträckas ända till en inblick av standardmodellen som en intressant anhalt för introduktionen i ämnet. Rapportens huvuddel är skriven på engelska då majoriteten av all litteratur inom fältet enbart finns på det språket samt för att kunna vända sig till en bredare, icke-svensktalande målgrupp.

Resultat

Litteraturstudierna har resulterat i en introducerande bok som ämnar att kunna leda en tredjeårs fysik- eller matematikstudent från introduktionen av begreppen grupp och symmetri till en översiktlig insikt rörande deras roll inom modern fysik. Boken inleds med en kort bakgrund varefter den matematiska formalismen gällande allmänna definitioner, ändliga och kontinuerliga grupper introduceras. Därefter presenteras grundläggande idéer kring liealgebra och representationsteori. Av pedagogiska skäl följs detta av en fysikalisk motivering till studierna av gruppteori via en genomgång av det kvantmekaniska begreppet spinn. Exemplet visar hur gruppteori och liealgebra uppstår naturligt ur fysikers beskrivning av verkligheten och fungerar som en konkret illustration av många koncept ur det efterföljande kapitlet. Detta senare kapitel fördjupar den matematiska behandlingen av liealgebror och representationsteori. Praktiskt sätt sker det genom införande av rot- och viktrum, cartanmatriser och dynkindiagram. Förfarandet når resultat som används för den senare hälften av boken.

Den andra halvan fokuserar mer på gruppteorins användning inom fysiken och betonar indelningen av relevanta symmetrier i externa och interna. Undersökningen av de förstnämnda inleds med en presentation av symmetrigrupperna i speciell relativitetsteori, Lorentz- och Poincaré-gruppen. Vidare fortsätter kapitlet med en explicit härledning av liealgebran för den definerande representationen av Poincarégruppen, en diskussion kring algebrans implikationer

och framtagning av andra grundläggande representationer. Resultatet blir en första klassificering av elementarpartiklar med kvanttalen massa och spinn. Behandlingen av externa symmetrier avslutas med en närmare studie av hur representationen av Poincarétransformationer på kvantmekaniska hilbertrum inte uppfyller gruppaxiomen fullt ut i ett parti om projektiva representationer. Avsnittet i fråga ger en något djupare inblick i de topologiska aspekterna av gruppteori vilka har antydits tidigare i boken. Nästa kapitel berör interna symmetrier i ljuset av s.k. gaugeteorier, hur dessa teorier kan konstrueras och listar de gaugeteorier som är relevanta för det avslutande kapitlet. Avslutningen föregås dock av ett kompletterande avsnitt om matematiken bakom gaugeteorier, vilket erbjuder en ytterligare utblick mot gruppteorins samspel med geometri och topologi. Det sista kapitlet går översiktligt igenom klassificeringen av elementarpartiklar, vilken vilar på en gruppteoretisk grund, med nya kvanttal som isospin och färg. Kvarkar och leptoner beskrivs i en framställning som grovt följer den historiska utvecklingen av området och återkoppling görs till tidigare kapitel. Som avslutning nämns något om gaugesymmetrigruppen för standardmodellen och möjliga underlag till framtida arbeten inom ämnet vilket förhoppningsvis inspirerar till fortsatt engagemang och nyfikenhet hos läsaren.

Till boken följer också ett relativt omfattande appendix som täcker några av de förkunskaper som är önskvärda för att kunna förmedla innehållet till en bredare målgrupp samt diverse räkningar.

Det inlärningsmässiga resultatet återspeglas huvudsakligen i den skriftliga produkten. Gruppmedlemmarnas förkunskaper inom området var begränsade. Därför reflekterar i princip allt innehåll i boken, bortsett från grundläggande kvantmekanik och speciell relativitetsteori, kunskaper som har förvärvats under arbetets gång. En ofrånkomlig följd av de enskilda avsnittens specialisering är att varje gruppmedlem inte har uppnått samma kunskapsdjup inom varje enskilt område. Emellertid har vi alla med detta arbete tillägnat oss en bred allmänbildning inom fältet med individuella fördjupningar vilket ger nya, utvecklade förutsättningar att ta till sig innehållet i framtida studier inom teoretisk fysik.

Diskussion

Arbetsprocessen i det här kandidatarbetet har inneburit stora skillnader mot tidigare studier då raka läsanvisningar, givna räkneuppgifter och tydliga avgränsningar för vad som förväntas kunnas har bytts mot ett öppet fält och mängder av källor, vilka ibland inte riktigt säger samma sak. Detta har inneburit utmaningar och stundom frustration gällande vilka kunskapsmässiga krav som har varit rimliga och över svårigheterna att finna lämpliga källor. Ett genomgående problem har varit att röra sig mellan matematik- och fysiklitteraturen där den förstnämnda är mycket detaljerad medan fysikpresentationerna ofta lämnar mycket underförstått. Tillsammans med svårigheterna att finna källor med passande förkunskapskrav har detta dock bidragit till viktiga lärdomar, både inför en framtid utan skräddarsydda mallar och för den pedagogiska kvaliteten hos vår slutprodukt. Vi har därmed försökt att vara tydliga med förkunskapskraven för vår text och motiverats till att inkludera ett utförligt appendix. Glappet mellan de två disciplinerna har vi försökt överbygga genom att varva exempel och teori, speciellt genom att föregå svåra definitioner med illustrerande exempel, samt fysik och matematik. En viktig erfarenhet av arbetet är hur inlärning sker på olika nivåer av abstraktionsgrad men att den inledningsvis alltid fordrar en strikt noggrannhet i framställningen. Kravet kan lättas först efter att man har uppnått en viss nivå, men då med fördel i och med

att det ofta är lättare att urskilja kärnan i resonemang när begreppen är välkända. Som en följd bör många mindre stringenta fysikkonventioner introduceras med särskild pedagogisk omtanke.

Hur stor omfattning den slutgiltiga texten skulle ha har varit en genomgående fråga under arbetets gång. Resultatet har blivit att vissa delar har en mer övergripande karaktär än andra för att läsare snabbt skall kunna bilda sig en uppfattning om innehållet. När det har varit möjligt har vi å andra sidan försökt gå in mer på djupet och illustrerat viktiga koncept genom pedagogiska räkneexempel och figurer. Uträkningar har på så vis ofta en mer instruktiv utförlighet än vad som är norm i artiklar och flertalet böcker.

Arbetsmetodiken har generellt fungerat bra. Det stora fokuset på textens utformning kan dock ha varit något hämmande för den egna inlärningsprocessen. Kunskapen hade möjligen befästs bättre om en större del av tiden hade nyttjats för egna studier, framförallt fler egna räkningar, istället för det omfattande skrivandet. Samtidigt har det funnits stora inlärningskvaliteter i att tvingas formulera kunskapen så lättbegripligt som möjligt. Likaså gäller detta i betydande grad även våra egna föreläsningar inom gruppen. Därtill har faktumet att andra möjligen kan få ut något av vårt arbete agerat som ytterligare motivation.

Slutsats

Gruppteori är ett mycket spännande ämne som har en avgörande roll i den moderna beskrivningen av den fysikaliska verkligheten. Litteraturstudierna har givit oss en allmänbildande orientering i ämnet med individuella fördjupningar vilket ger en god grund inför fortsatta fysikstudier. Det sammanlagda arbetet har resulterat i en enligt vår uppfattning välanpassad bok för medstudenter och som potentiellt sett kan användas i undervisning eller för självstudier.

Part II

Huvuddel

Chapter 1

Introduction

Beauty is bound up with symmetry - Hermann Weyl, German mathematician. [8]

If you happen upon a butterfly you immediately notice the patterns which appear on both wings as well as the wings' almost identical forms. Our minds seek patterns in order to establish recognition, seeking out each visible hint of symmetry. Since early history mankind has shown a keen interest in the symmetries of nature. This passion has been manifested in elaborately crafted arts and architecture, such as the magnificent Taj Mahal. Observing a kayak on a tranquil mountain lake, in our eyes it is subjected to a parity transformation and hence mirrored in the water upside-down. However, what would happen if we mirrored an equation of motion, a differential equation, instead? To improve our understanding of the laws of nature, physicists seek out the symmetry of theories and see if the theory itself can be expressed in a new, different way and perhaps combined with something else. When constructing a new theory, symmetry properties generally indicate that it is on the right track. [9] [10]

The mathematical framework which depicts symmetries is known as "group theory" and was developed by the Norwegian mathematicians Niels Henrik Abel and Sophus Lie among others during the nineteenth century. After the advent of quantum mechanics this theory soon found a way in the description and unification of phenomena in theoretical physics. Together with the progress in construction of massive colliders and other experimental setup the field of particle physics was created during the twentieth century. Today with the Large Hadron Collider (LHC) at CERN we may be able find new phenomena, requiring new theories based on symmetries.

From quantum mechanics, we recall the conserved quantities associated with temporal, spatial and rotational invariance of a given problem through Emmy Noether's theorem. Yet, subatomic experiments indicate that there may be other symmetries corresponding to other conserved quantities when particles suddenly transform into one another. The first symmetries mentioned in this paragraph are known as *external symmetries* (the symmetries of spacetime) and the latter *internal symmetries* . Our work attempts to describe both of these symmetries.

The main aim of this book is to convey a brief survey of group theory and explore and discuss its usages in describing the behavior of elementary particles. It will be presented as

our bachelor project in Engineering Physics at Chalmers University of Technology. The first chapters focus on the mathematical aspects, pertinent definitions and introduce the framework in lucid, illustrative examples whereas the latter ones put emphasis on applications in different areas of particle physics. Complementary material on topology, tensor notation, Hilbert and dual spaces et cetera helpful in order to understand the arguments throughout the book as well as supplementary calculations are found in the assorted appendices. A reader's guide is provided below which explains the structure of the text and the connections between different chapters and sections.

1.1 Reader's Guide

Mathematical Foundations

Chapter 1 presents a brief introduction of the subject whereas chapter 2 introduces necessary mathematical foundations such as the group axioms, basic notions, finite groups as well as Lie groups and provides a brief treatment of Lie algebra, which we return to in chapter 5 in more detail. The isomorphism between the Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$ is mentioned and their group level relations will be developed in chapter 4. Its topological implications will be explained in the last section of chapter 6.

Next we proceed with representation theory in chapter 3 with mathematical preliminaries and the chapter is concluded with a section about the tensor product and the direct sum which is required in the next chapter to understand the addition of angular momenta. The definitions and results are used frequently in the following chapters.

On one hand, the next chapter may seem a bit far off, titled "A physical approach", but the interested reader soon discovers that by expanding the initial rotational matrices one arrives in the same theory as the previous chapter regarding the Lie groups $SU(2)$ and $SO(3)$. Their topological relation is determined in the last section of chapter 6. Moreover, the chapter provides discussions about spin 1/2-systems as well as spinors, which we will encounter later in chapter 6, and finishes with the familiar ladder operators of quantum mechanics and investigates how these can be used in order to create irreducible representations.

Afterwards we dive into chapter 5, where a more formal scrutiny of semi-simple Lie algebras is performed. Notions such as roots and weights, the Chevalley-Serre presentation, Cartan matrices, Dynkin diagrams and Casimir operators are introduced and most of the theory is summarized in an example with the Lie algebra $\mathfrak{sl}(3,\mathbb{C})$. This Lie algebra will return in the description of the Eightfold Way and $SU(3)$-gauge group of quarks in particle physics, which are developed in the end of chapter 8.

Physical Applications

The latter chapters 6, 7 and 8 illustrate the physical significance of the prior mathematical treatment. Chapter 6 focus on external symmetries encoded by the Lorentz and Poincaré groups. Firstly, basic properties of the both groups are summarized. Thereafter, we derive the defining representation's Lie algebra and ponder its implications. Moreover, a brief survey of the various manifestations of the Lorentz and Poincaré groups are conducted where we once more come across spinors. With these new tools as well as the Casimir operators of chapter

5 we are able to classify 1-particle states. Lastly, the subject of projective representations is elaborated both in a Lie algebra, Lie group and a topological picture.

Chapter 7 and chapter 8 chiefly examine the nature of internal symmetries. The first one, conveniently named "Internal symmetries", deduces the nature of these symmetries with gauge theory. It begins in Maxwell's classical electromagnetism and we analyze how one may construct a gauge theory from it. Afterwards we move on to a more general case, constructing an arbitrary gauge theory and discern how the procedure can be performed with Lagrangian formalism. To conclude we encounter non-Abelian gauge theories and come upon the gauge groups $SU(2)$ and $SU(3)$, which reappear in chapter 8. An excursion to the underlying topological constructs known as fiber bundles is provided for the interested reader (hence the asterisk in the table of contents).

The fascinating conclusion to this book is presented in chapter 8, where we encounter group theory in modern particle physics with additional information on historical contributions and experiments. At first we present the Coleman-Mandula theorem, consider its consequences and connecting it with previous statements in chapter 6 and 7. The chapter continues with a brief overview of the particles of the Standard Model of particle physics. Feynman diagrams are introduced and the last sections concern how we may view isospin, quarks and mesons on group theoretical premises. Here we once more use the addition of angular momenta of chapter 4, the gauge symmetries of chapter 7 and the Lie algebra of $\mathfrak{sl}(3, \mathbb{C})$ in the Eightfold Way of mesons as well as the quarks. The entire book is concluded with a few remarks on the gauge groups of the Standard Model and some interesting topics suitable for future studies. Attached are various appendices provided with some valuable prerequisites and more information to enhance the reader's understanding of the work presented in the book's chapters.

We wish you an agreeable reading experience and hope you will enjoy our work.

Chapter 2

Basic Group Theory

2.1 Introduction

In this section we will give the very basics of group theory on which this text will be built upon. We will start with alot of definitions but also present some examples. A few propositions will also be given, intended to show that this is a consistent theory.

Definition 2.1.1. *A* **group** *is a set $\mathcal{G}$, together with an operation $\star$, denoted $(\mathcal{G},\star)$, satisfying the group axioms:*

(i) **Closure** – *For every ordered pair $(a_i,\ a_j)$ of elements in $\mathcal{G}$, there exists an* unique *element a_k such that,*

$$a_i \star a_j = a_k \in \mathcal{G} \tag{2.1.1}$$

(ii) **Associativity** – *For arbitrarily chosen $a_i,\ a_j,\ a_k \in \mathcal{G}$*

$$(a_i \star a_j) \star a_k = a_i \star (a_j \star a_k) \tag{2.1.2}$$

(iii) **Unit element** – *There is a* unique *element $e \in \mathcal{G}$ such that*

$$e \star a_i = a_i \star e = a_i,\ \forall a_i \in \mathcal{G} \tag{2.1.3}$$

(iv) **Inverse element** – *For every element $a_i \in \mathcal{G}$, there is an* unique *inverse $(a_i)^{-1} \in \mathcal{G}$ such that*

$$a_i \star (a_i)^{-1} = (a_i)^{-1} a_i = e \tag{2.1.4}$$

The number of elements in a group can be finite, countably infinite or uncountably infinite. We say that $(\mathcal{G},\star)$ is a **finite group** if the set $\mathcal{G}$ is finite, a **countably infinite group** if $\mathcal{G}$ is countably infinite and an **uncountably infinite group** if the set $\mathcal{G}$ is uncountably infinite. Further on the set $\mathcal{G}$ is called the **underlying set** of the group $(\mathcal{G},\star)$. However $\mathcal{G}$ is commonly used to refer to the full group $(\mathcal{G},\star)$, and we will from now on do so in this text.

The operation $\star$ is called the **composition** or **the product operation** but is often referred to as multiplication and the star is often omitted, $a_1 \star a_2 = a_1 a_2$. We can also see that the group axioms do not require the multiplication to be commutative, however, it is such an important special case that it has been given its own name.

Definition 2.1.2. *Suppose we have a group $\mathcal{G}$. If, $a_i \star a_j = a_j \star a_i; \forall a_i, a_j \in \mathcal{G}$, then $\mathcal{G}$ is said to be* **Abelian**.

To get a better picture of what a group can be we will start by studying an example of a finite group. Finite groups in difference to infinite groups have an *order*.

Definition 2.1.3. *Consider a finite group $\mathcal{G} = \{a_1, a_2, \ldots, a_n\}$, the number of elements, n, is called the* **order** *of the group and is denoted $|\mathcal{G}|$.*

We skip the trivial group of order one and begin to study the group of order two. Worth mentioning is that there exists only one group of order two, often called $\mathbb{Z}_2$, which can be described using a **multiplication table**.

Example 2.1.1. The first element in $\mathbb{Z}_2$ must be the identity operation, e, and the second element, a, must be its own inverse, to satisfy the group axioms. The resulting multiplication table will then be as follows. It is now easy to see that all the group axioms are satisfied. We

Table 2.1: Multiplication table of $\mathbb{Z}_2$.

$\star$	e	a
e	e	a
a	a	e

have a closure since every ordered pair of $\{e, a\}$ together with the multiplication yields an element in the group. We also have associativity, the unit element, e and both a and e are their own inverses. We can also add, from the multiplication table, that the multiplication is commutative and thus $\mathbb{Z}_2$ is Abelian.

To give an illustration of $\mathbb{Z}_2$, a so called *representation*, one can think that a operates as an 180° rotation of a picture in a plane and e, the identity operation, leaves the picture as it is. Thus applying a two times, $a \star a$, results in the same picture as when we started.

Before we continue with more examples and definitions we will stop and prove some important properties of groups. The first one deal with the third property of our first definition.

Proposition 2.1.1: The unit element e, sometimes called the *identity element* or simply the *identity*, is unique for a given group $\mathcal{G}$.

Proof. Let $e_2 \neq e$ be another element of $\mathcal{G}$ with the property (iii) defined above for e:

$$e_2 \star a_i = a_i \star e_2 = a_i, \; \forall a_i \in \mathcal{G} \tag{2.1.5}$$

The left-hand side of equation (2.1.3) gives that $e \star e_2 = e_2$, and the right-hand side of equation (2.1.5) gives that $e \star e_2 = e$. Consequently $e_2 = e$, which contradicts the premise $e_2 \neq e$. Thus, e_2 *must* be equal to e and the unit element is therefore uniquely determined. □

Note that this holds for *all* groups since no extra requirements was used on the group $\mathcal{G}$. The next three propositions deals with the fourth property of our first definition.

PROPOSITION 2.1.2: For every element a_i in a given group $\mathcal{G}$, the inverse $(a_i)^{-1}$ is uniquely determined.

Proof. The proof is similar to the one above. Assume that $(a_i)_2^{-1} \neq (a_i)^{-1}$ is another element of $\mathcal{G}$ with the property (iv) defined above for $(a_i)^{-1}$:

$$a_i \star (a_i)_2^{-1} = (a_i)_2^{-1} \star a_i = e \tag{2.1.6}$$

Together, equation (2.1.4) and (2.1.6) give that $(e =)\ a_i \star (a_i)^{-1} = a_i \star (a_i)_2^{-1}$. Multiplying from the left on both sides of this equation with $(a_i)^{-1}$ gives

$$(a_i)^{-1} \star (a_i \star (a_i)^{-1}) = (a_i)^{-1} \star (a_i \star (a_i)_2^{-1}) \overset{(2.1.2)}{\Longleftrightarrow} \tag{2.1.7}$$

$$((a_i)^{-1} \star a_i) \star (a_i)^{-1} = ((a_i)^{-1} \star a_i) \star (a_i)_2^{-1} \overset{(2.1.4)}{\Longleftrightarrow} \tag{2.1.8}$$

$$(a_i)_2^{-1} = (a_i)^{-1} \tag{2.1.9}$$

But equation (2.1.9) contradicts the assumption that $(a_i)_2^{-1} \neq (a_i)^{-1}$, which proves that the two elements are in fact one and the same. In other words, the inverse of any given element $a_i \in \mathcal{G}$ is uniquely determined. □

PROPOSITION 2.1.3: Let a and b be elements of a group $\mathcal{G}$. If *either* $a \star b = e$ *or* $b \star a = e$, then b is the uniquely determined inverse of a in $\mathcal{G}$ (and vice versa).

Proof. Suppose that $a \star b = e$. Then

$$\begin{aligned} b \star (a \star b) &= b \star e \Leftrightarrow \\ (b \star a) \star b &= b \end{aligned} \tag{2.1.10}$$

As an element of $\mathcal{G}$, b must have an inverse $b^{-1} \in \mathcal{G}$. Righ-multiplying both sides of equation (2.1.10) by b^{-1} gives

$$\begin{aligned} ((b \star a) \star b) \star b^{-1} &= b \star b^{-1} \Leftrightarrow \\ (b \star a) \star (b \star b^{-1}) &= e \Leftrightarrow \\ (b \star a) \star e &= e \Leftrightarrow \\ b \star a &= e \end{aligned} \tag{2.1.11}$$

Since $a \star b = e$ and $b \star a = e$, proposition 2.1.2 gives that b is the unique inverse of a in $\mathcal{G}$ (and vice versa). Now suppose that $b \star a = e$. By left-multiplying both sides of this equation by first a and then a^{-1}, we find that $a \star b = e$. This proves that if **either** $a \star b = e$ **or** $b \star a = e$, then b is the unique inverse of a in $\mathcal{G}$ (and vice versa). □

PROPOSITION 2.1.4: If a and b are elements of a group $\mathcal{G}$, then the inverse $(a \star b)^{-1}$ of $a \star b$ is equal to $b^{-1} \star a^{-1}$ (where a^{-1} and b^{-1} denote the inverse of a respectively b).

Proof. By a straightforward calculation we find:

$$\begin{aligned}(a \star b) \star (b^{-1} \star a^{-1}) &= ((a \star b) \star b^{-1}) \star a^{-1} \\ &= (a \star (b \star b^{-1})) \star a^{-1} \\ &= (a \star e) \star a^{-1} \\ &= a \star a^{-1} \\ &= e \end{aligned} \tag{2.1.12}$$

By proposition 2.1.3, $b^{-1} \star a^{-1}$ is the (unique) inverse of $a \star b$. □

We will now introduce some new notation necessary for upcoming subjects.

Definition 2.1.4. *A* **subgroup** $\mathcal{H}$ *of a group* $\mathcal{G}$ *is a subset* $\mathcal{H} \subseteq \mathcal{G}$ *with the properties:*

(i) $h_1, h_2 \in \mathcal{H} \Rightarrow h_1 \star h_2 \in \mathcal{H}$

(ii) The unit element is an element of $\mathcal{H}$*.*

(iii) $h \in \mathcal{H} \Rightarrow h^{-1} \in \mathcal{H}$

Note that these properties are identical with the group axioms (i), (iii) respectively (iv) listed above. Since every element of $\mathcal{H}$ is also an element of $\mathcal{G}$, the associativity property (ii) follows automatically from the fact that $\mathcal{G}$ is a group. This means that every subgroup of a group $\mathcal{G}$ is in itself a group.

The number of subgroups varies, but every group $\mathcal{G}$ has at least two subgroups, namely $\{e\}$ and $\mathcal{G}$ itself. These are called **trivial subgroups**. A subgroup where $\mathcal{H}$ is a proper subset of $\mathcal{G}$, i.e. $\mathcal{H} \subset \mathcal{G}$, is called a **proper subgroup**.

Definition 2.1.5. *Consider two groups* $\mathcal{G}$ *and* $\mathcal{S}$*, the* **direct product** *of the groups* $\mathcal{G} \times \mathcal{S}$ *is then defined as follows:*

(i) The elements of $\mathcal{G} \times \mathcal{S}$ *are ordered pairs* (g, s)*,* $g \in \mathcal{G}$ *and* $s \in \mathcal{S}$*.*

(ii) The composition $\star$ *on* $\mathcal{G} \times \mathcal{S}$ *is defined component wise:*

$$(g_1, s_1) \star (g_2, s_2) = (g_1 \star g_2, s_1 \star s_2)$$

We will now present the two groups of order four that can exist. Since the group of order three is much like the one of order two in the previous example we will skip it for now.

Example 2.1.2. There are two ways to approach a group of order four. If we use the same illustration of rotations as in the previous example, we can think of a cyclic series of four rotations. The first element, a_1, corresponds to a 90° rotation, the second element corresponds to a 180° rotation, the third corresponds to a 270° rotation and finally the unit element corresponds to no rotation at all.

We will now get, for example, by applying a_1 two times, $a_1 a_1$, two 90° rotations which obviously is the same as a 180°, thus $a_1 a_1 = a_2$ and so on. A more correct way to construct this group however is by using a multiplication table, since we would not need the use of an illustration. This group is commonly known as $\mathbb{Z}_4$ however, as mentioned before, we can construct a second group of order four different from $\mathbb{Z}_4$. This second group is called the

Table 2.2: Multiplication table of $\mathbb{Z}_4$.

$\star$	e	a_1	a_2	a_3
e	e	a_1	a_2	a_3
a_1	a_1	a_2	a_3	e
a_2	a_2	a_3	e	a_1
a_3	a_3	e	a_1	a_2

Table 2.3: Multiplication table of $\mathcal{D}_2$.

$\star$	e	a_1	a_2	a_3
e	e	a_1	a_2	a_3
a_1	a_1	e	a_3	a_2
a_2	a_2	a_3	e	a_1
a_3	a_3	a_2	a_1	e

dihedral group of order 2 and has the multiplication table according to table 2.3. From the multiplication table we can see that it is Abelian due to the fact it is symmetrical about its diagonal.

Another feature both of these groups share is that we can construct non-trivial subgroups of their elements. This is obviously true in the first case, since $\{e,a_2\}$ forms $\mathbb{Z}_2$ (the "180° rotation" together with the identity), in other words $\mathbb{Z}_2 \subset \mathbb{Z}_4$. However, this is the case of $\mathcal{D}_2$ as well. In fact, one can form three different $\mathbb{Z}_2$ subsets from $\mathcal{D}_2$, both (e, a_1), (e, a_2) and (e, a_3). Since $\mathcal{D}_2$ was Abelian the elements of the first two $\mathbb{Z}_2$ commute with one another, and thus we can express $\mathcal{D}_2$ as the direct product between the first two $\mathbb{Z}_2$, $\mathcal{D}_2 = \mathbb{Z}_2 \times \mathbb{Z}_2$.

Definition 2.1.6. *Consider a group $\mathcal{G}$, a subgroup $\mathcal{H} \subset \mathcal{G}$ and an element $g \in \mathcal{G}$ then,*

$$g\mathcal{H} = \{gh : h \in \mathcal{H}\} \text{ and} \tag{2.1.13}$$

$$\mathcal{H}g = \{hg : h \in \mathcal{H}\} \tag{2.1.14}$$

*is said to be a **left coset of** $\mathcal{H}$ in $\mathcal{G}$ respectively a **right coset of** $\mathcal{H}$ in $\mathcal{G}$.*

It is important to note that a coset is, in general, not a subgroup. Cosets play an important role, mainly in the field of finite groups. To show what they can be used for we will prove *Lagrange's theorem*.

Theorem 2.1.1. (*Lagrange's Theorem*) If a group $\mathcal{G}$ of order N has a subgroup $\mathcal{H}$ of order n then the **index** of $\mathcal{H}$ in $\mathcal{G}$, defined as N/n, must be an integer.

Proof. Take an element $g_1 \in \mathcal{G}$, but not in $\mathcal{H}$ and form the left coset $g_1\mathcal{H}$.

$$g_1\mathcal{H} = g_1 h_i,\ h_i \in \mathcal{H},\ i = 1,2,...,n.$$

Since $g_1 \notin \mathcal{H}$, $g_1\mathcal{H}$ and $\mathcal{H}$ must be disjoint. If $g_1\mathcal{H}$ and $\mathcal{H}$ were not disjoint then there exists an element $h \in \mathcal{H}$ such that $g_1 h_i = h$ for some i. But this imply that $g_1 \in \mathcal{H}$ and we have a contradiction.

Now, take another $g_2 \in \mathcal{G}$ that is not in $\mathcal{H}$ or $g_1\mathcal{H}$ and form the left coset $g_2\mathcal{H}$.

$$g_2\mathcal{H} = g_2 h_i,\ h_i \in \mathcal{H},\ i = 1, 2, ..., n.$$

Since g_2 is not in $g_1\mathcal{H}$ or $\mathcal{H}$, $g_2\mathcal{H}$ are disjoint to both $g_1\mathcal{H}$ and $\mathcal{H}$. If they were not disjoint then $g_2 h_j = g_1 h_i$ for some i, j, but by taking the invers, h_j^{-1}, on both sides we get $g_2 = g_1 h_i h_j^{-1}$ and g_2 would have to be in $g_1\mathcal{H}$ since $h_i h_j^{-1} \in \mathcal{H}$. That $g_2\mathcal{H}$ and $\mathcal{H}$ are disjoint follows from above. If we continue in this way the coset space formed by the cosets $g_j\mathcal{H}$, $j = 1, 2, ..., k$, and $\mathcal{H}$ will cover $\mathcal{G}$. Since all of these cosets are disjoint N must be a multiple integer $k + 1$ of n. $\mathcal{G}$ can be expressed as a sum of cosets

$$\mathcal{G} = \mathcal{H} + \sum_{i=1}^{k} g_i\mathcal{H}.$$

□

An intresting consequence of this is that every group with a prime number order has no non-trivial subgroups.

Definition 2.1.7. *Consider a group $\mathcal{G}$, the* **conjugate***, $\tilde{g}_i$, of any element $g_i \in \mathcal{G}$ with respect to another element $g_j \in \mathcal{G}$ is defined as,*

$$\tilde{g}_i = g_j g_i g_j^{-1}. \tag{2.1.15}$$

This operation plays an important role since it can be used to organize the elements of any finite group into a set of *conjugacy classes*.

Definition 2.1.8. *Let g_b be any element in $\mathcal{G}$. The set C_b defined as following,*

$$C_a :\ \tilde{g}_b = g_a g_b g_a^{-1}, \forall g_a \in \mathcal{G} \tag{2.1.16}$$

is said to be a ***conjugacy class****.*

We will finally introduce three important concepts.

Definition 2.1.9. *Let $\mathcal{G}$ and $\mathcal{H}$ be two groups. A map $T : G \to H$ is said to be a* **homomorphism** *if it satisfies $g_i g_j = g_k \Rightarrow T(g_i)T(g_j) = T(g_k)$, $\forall g_i, g_j \in \mathcal{G}$. If T is also bijective it is said to be an* **isomorphism** *between $\mathcal{G}$ and $\mathcal{H}$, denoted $\mathcal{G} \cong \mathcal{H}$. An isomorphism of a group with itself is called an* **automorphism**.

Note that the definition of isomorphism can be extended to a more general case, i.e. if a map and its inverse between two sets which have some algebraic structures, such as addition or multiplication, preserve these structures, then the map is isomorphic.

To resume this section we will now give one of the most common examples of a group, namely the one of an equilateral triangle.

Example 2.1.3. Consider the equilateral triangle ABC in figure 2.1. There are six symmetry transformations that we may perform that leaves the triangle unchanged, including the identity transformation (leaving the triangle as it was). We can perform a flip or reflection

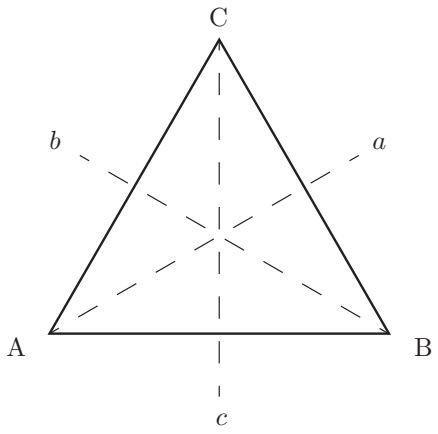


Figure 2.1: An equilateral triangle with the three axes a,b and c indicated for the flip operations.

on one of the axis a, b or c. We can also rotatate it 120° and 240°. The first rotation will be called r and the second s.

If we perform the reflection a, b or c two times we will be back with the original triangle, that is to say $aa = e$, $bb = e$, $cc = e$. Two of the r rotations yields the s rotation and r together with s obviously gives $rs = sr = e$. Further on will an r rotation together with an a reflection give $ra = b$. Since the r rotation will give the triangle CAB and the a reflection will flip it into CBA, which the b reflection in the same way will flip ABC into CBA. However it is worth to notice that if we change the order in which we perform the rotation and the flip we will end up with another result. In our case if we perform a first we will end up with the triangle ACB and the rotation r will transform it into BAC. In other words, the group is non-Abelian.

By following this logic we will eventually reach the multiplication table in table 2.4.

Table 2.4: Multiplication table of $\mathcal{D}_3$.

$\star$	e	r	s	a	b	c
e	e	r	s	a	b	c
r	r	s	e	b	c	a
s	s	e	r	c	a	b
a	a	c	b	e	s	r
b	b	a	c	r	e	s
c	c	b	a	s	r	e

As seen in the example above multiplication tables becomes quite a diffuse way to present information in, for larger groups. However there exists a more effective way to give information in, through *presentations*. To be able to define a presentation we need something called *generators*.

A *generating set* of a group is a subset of the group such that every element of the group can be expressed as a combination, under the group operation, of finitely many elements and their inverses of the subset. More generally we define the generating set as follows.

Definition 2.1.10. *Consider a group $\mathcal{G}$ with a subgroup $S \subset \mathcal{G}$. The subgroup $< S >$, the* **subgroup generated by** *S, is then the subgroup of $\mathcal{G}$ containing all elements in $\mathcal{G}$ which can be expressed by a finite product of elements in S.*

If $\mathcal{G} =< S >$, then S **generates** *$\mathcal{G}$ and the elements of S are called the* **generators** *of $\mathcal{G}$.*

We can now see that, for example, the group $\mathbb{Z}_2$ from example 2.1.1 and the group $\mathbb{Z}_4$ from example 2.1.2 are both generated from a single generator. Since in the first case $aa = e$ and in our second case $a_1a_1 = a_2$ and $a_1a_1a_1 = a_2a_1 = a_3$ and therefore $< a >= \mathbb{Z}_2$, $< a_1 >= \mathbb{Z}_4$. This can of course be generalized to build the family of **cyclic groups** $Z_n =< x >= \{x^0, x^1, x^2, ..., x^n\}$.

Definition 2.1.11. *A* **presentation** *of a group is a set of generators S and a set R of relations among those generators. We then say $\mathcal{G}$ has the presentation $\langle S|R\rangle$.*

The information in table 2.4 can now be given through the shorter way of a presentation, $\mathcal{D}_3 = \langle a, b \,|\, a^3 = b^2 = e;\, bab^{-1} = a^{-1}\rangle$. The same holds for $\mathcal{D}_2$ from example 2.1.2, $\mathcal{D}_2 = \left\langle a_1, a_2 \,|\, a_1^2 = a_2^2 = e;\, a_2a_1a_2^{-1} = a_1^{-1}\right\rangle$. This can as well be generalized into the family of **dihedral groups**, $\mathcal{D}_n = \langle r, s \,|\, r^n = s^2 = e, srs^{-1} = r^{-1}\rangle$.

2.1.1 Normal Subgroups

This small section will be used to introduce some concepts which are of importance for subjects presented later on in this report.

Definition 2.1.12. *Consider a group $\mathcal{G}$, a subgroup $N \subseteq \mathcal{G}$ is then called a* **normal subgroup**, *denoted $N \triangleleft G$, if it is invariant under conjugation i.e.,*

$$\forall n \in N, \forall g \in \mathcal{G} : \; gng^{-1} \in N. \tag{2.1.17}$$

The two trivial subgroups $\{e\}$ and $\mathcal{G}$ itself are obviously normal subgroups. The same applies for *all* subgroups of Abelian groups since their multiplications are commutative and thus $gN = Ng$.

One important result from the normal subgroup is the *semidirect product* which is a generalization of the direct product.

Definition 2.1.13. *Consider a group $\mathcal{G}$ with a unity element e, a normal subgroup $N \triangleleft G$ and a subgroup $\mathcal{H} \subseteq \mathcal{G}$. If*

$$\mathcal{G} = N\mathcal{H} = \mathcal{H}N \text{ and } N \cap \mathcal{H} = e, \tag{2.1.18}$$

then $\mathcal{G}$ is said to be the **semidirect product** *of N and $\mathcal{H}$, denoted $G = N \rtimes \mathcal{H}$.*

This will show to be a useful tool for constructing groups in upcoming chapters. We will see that it is not necessary for N and H to be subgroups of any given group in order to form a semidirect product.

Consider a group $\mathcal{G} = N \rtimes H$, where N and H not necessarily are subgroups of $\mathcal{G}$, then define a map $\varphi : H \to \text{Aut}(N)$. Here $\text{Aut}(N)$ denotes the group of all automorphisms of N. The purpose of this map is to tell how the group operation in $\mathcal{G}$ will work. By defining φ as,

$$\varphi(h) = \varphi_h(n) = hnh^{-1}, \forall h \in H, \forall n \in N, \tag{2.1.19}$$

it forms a group homomorphism and we will see that together with N and H, it will determine $\mathcal{G}$ up to isomorphism. We say that $\mathcal{G} = N \rtimes_\varphi H$ is the **semidirect product of N and H with respect to** φ. As a set $N \rtimes_\varphi H$ will just be the cartesian product $N \times H$. The group operation will, as mentioned, be determined through φ. The operation is defined as,

$$\star : (N \rtimes_\varphi H) \times (N \rtimes_\varphi H) \to (N \rtimes_\varphi H) \tag{2.1.20}$$

$$(n_1,h_1) \star (n_2,h_2) = (n_1\varphi_{h_1}(n_2),h_1h_2) \tag{2.1.21}$$

for all n in N and all h in H. To get a better picture of this we will recall example 2.1.3.

Example 2.1.4. In example 2.1.3 we saw how we could construct the symmetry group of the equilateral triangle. Except from the identity transformation we had five other tranformations which transformed the triangle to its original appearance, two rotations and three reflections.

With our new tool, the semidirect product, we will now be able to construct the symmetry group from two cyclic groups since it will show that $\mathcal{D}_3 \cong \mathbb{Z}_3 \rtimes_\varphi \mathbb{Z}_2$ where φ is the identity transformation for $\varphi(0)$ (the first element in $\mathbb{Z}_2$) and inversion for $\varphi(1)$. As a set we will get the product,

$$\mathbb{Z}_3 \times \mathbb{Z}_2 = \{(e', e), (y_1, e), (y_2, e), (e', x), (a_1, x), (a_2, x)\} \tag{2.1.22}$$

where $\mathbb{Z}_3 = \{e', y_1, y_2\}$ and $\mathbb{Z}_2 = \{e, x\}$. The rotation elements is now (y_1,e) and (y_2,e), corresponding to r and s in table 2.4, and the reflection elements are (e',x), (y_1,x) and (y_2,x) corresponding to a, b and c in table 2.4. As an example we have $rs = (y_1, e) \star (y_2, e) = (y_1\varphi_e(y_2), ee) = (y_1y_2, e) = (e', e) = e$ as we can recall from the multiplication table 2.4. To get an example where φ acts as an inversion we can multiply an rotation with a reflection.

This method can be generalized for all dihedral groups and sometimes this is gives as an alternative definition $\mathcal{D} \cong \mathbb{Z}_n \rtimes_\varphi \mathbb{Z}_2$.

Finally we will define something called the *quotient group*. It is somewhat analouge to division of integers but instead of numbers we divide groups. The normal subgroups will serve as the divisors here much in the same sense as for example 2 is a divisor of 6. Since groups contain much more information than a single number we will preserve or get more structure in the final quotient.

Definition 2.1.14. *Consider a group $\mathcal{G}$ with a normal subgroup $N \triangleleft G$. The set of all left cosets of N in $\mathcal{G}$, $\{aN : a \in \mathcal{G}\}$, is then said to be the* **quotient group** *denoted $\mathcal{G}/N$. The group operation is defined as the product between the cosets.*

The fact that the group operation works builds on the properties of the normal subgroup. Since N is a normal subgroup of $\mathcal{G}$, aN is equal to Na for all a in G. In other words, the left and right cosets of N in $\mathcal{G}$ are equal. The product of two cosets is thereby,

$$(aN)(bN) = a(Nb)N = a(bN)N = (ab)N, \forall a, b \in G. \tag{2.1.23}$$

Since $\mathcal{G}$ is a group, ab is also in $\mathcal{G}$ and $(ab)N \in \mathcal{G}/N$. To illustrate this we will give an example.

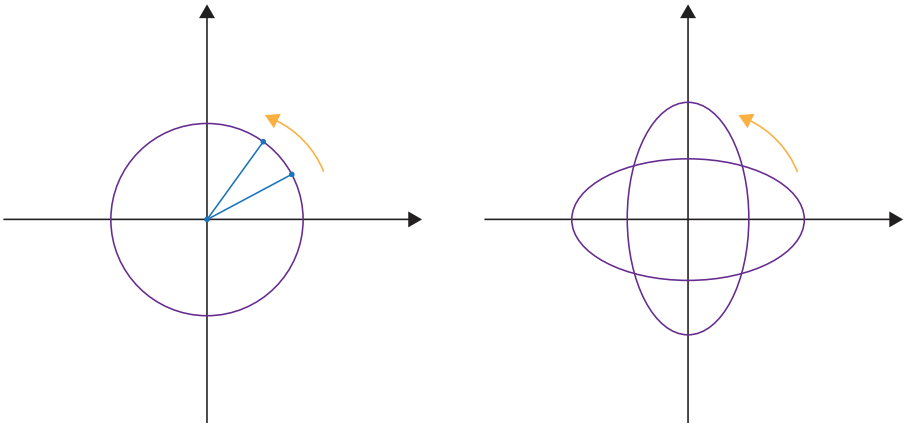


Figure 2.2: This figure illustrates the difference between continuous and discrete symmetry groups. We can rotate a circle with an arbitrary angle and still receive the same picture, i.e. a continuous rotational symmetry. If we rotate a vector on the circle with an arbitrary angle, we end up on another point on the circle since the transformation must conserve the vector's length. However, for the ellipse we have to rotate it 180° in order to obtain the same picture and hence we have a discrete rotational symmetry.

Example 2.1.5. Consider the set $\{0, 1, 2, 3, 4, 5\}$ under addition modulo 6. This forms a group $\mathcal{G}$ with a subgroup $N = 0, 3$ which in fact shows to be normal. The quotient group is now,

$$\begin{aligned} \mathcal{G}/N &= \{aN : a \in \mathcal{G}\} = \{aN : a \in \{0, 1, 2, 3, 4, 5\}\} \\ &= \{\{0, 3\}, \{1, 4\}, \{2, 5\}\} \end{aligned}$$

We are now finished with the first introductory section. Although all of our examples have been on finite groups it is worth to remember that all methods and expressions here can be used on infinite groups as well. In the next section we will leave the finite groups to only discuss infinite, in a special case named *Lie groups*.

2.2 Lie Groups

In physical applications we mostly deal with so called *Lie groups* which are continuous in contrast with the finite groups described throughout the previous section. The difference is shown in figure 2.2. These new continuous groups may also be considered as manifolds (a definition of this concept can be found in appendix A). To get some clues what this statement implies, let us study a few examples. Consider a vector $\boldsymbol{v}$ in $\mathbb{R}^2$ which we rotate with an arbitrary angle θ around the origin in the positive mathematical direction. This can be described by a matrix R according to

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{2.2.1}$$

Let $\boldsymbol{v'}$ be the rotated vector. Observing that it ought to have the same length as the original vector $\boldsymbol{v}$, the relation $\boldsymbol{v'}^2 = \boldsymbol{v}^2$ has to be satisfied. Moreover, since the norm can be written as $\boldsymbol{v}^T\boldsymbol{v}$, then $\boldsymbol{v'}^2 = (R\boldsymbol{v})^T(R\boldsymbol{v}) = \boldsymbol{v}^T R^T R\boldsymbol{v} = \boldsymbol{v}^T\boldsymbol{v}$, since $R^T R = \mathbb{1}$, which means that R is orthogonal in order to preserve the norm of the vector. This matrix is a representation of $SO(2)$, a continuous *Lie group*, which describes the rotations in $\mathbb{R}^2$. S stands for *special* which means det $R = 1$, O is orthogonal and 2 implies 2×2 - matrices. By continuous we mean that one can take θ arbitrary for an arbitrary rotation.

The same thing happens when one multiplies a complex number $z \in \mathbb{C}$ with a phase $e^{i\phi}$. As we recall from elementary complex analysis, $z = |z|e^{i\theta}$ and hence when multiplied with a phase $z \to z' = |z|e^{i(\theta+\phi)}$. This also corresponds to a rotation about the origin with another group $U(1)$, which is the one-dimensional unitary group (we will soon see what this means). We have an *isomorphism* between $U(1)$ and $SO(2)$, $U(1) \cong SO(2)$, consistent with definition 2.1.9.

$SO(2)$ is an example of a so-called *classical Lie group* or *matrix Lie group*. These are formed from the **general linear group** with complex entries $GL(n;\mathbb{C})$ which consists of all $n \times n$ invertible matrices with complex entries. Of course, the general linear group with real entries $GL(n;\mathbb{R})$ is a subgroup of $GL(n;\mathbb{C})$. After the definition we will examine a few examples and eventually reach the formal definition of a general Lie group.

Definition 2.2.1. *Let A_n be a sequence of complex matrices. A_n* **converges** *to a matrix A if each entry of A_n converges to the corresponding entry of A.*

Definition 2.2.2. *A* **matrix Lie group (classical Lie group)** *is any subgroup $\mathcal{H}$ of $GL(n;\mathbb{C})$ with the following property. If A_n is an arbitrary sequence of matrices in $\mathcal{H}$ and A_n converges to some matrix A then either $A \in \mathcal{H}$ or A is not invertible.*

A more informal description is that a matrix Lie group is a group whose *defining representation* are matrices. This concept will be thoroughly explained in example 3.1.2 in chapter 3. The definition above of a matrix Lie group implies that $\mathcal{H}$ is a closed subset of $GL(n;\mathbb{C})$ and that all matrix Lie groups are closed subgroups of $GL(n;\mathbb{C})$. It is possible to locate counter-examples, i.e. some subgroups of $GL(n;\mathbb{C})$ which are not closed and thus not matrix Lie groups, but we will not digress and take the issue further, see for instance ([11], p. 4). Instead, let us get acquainted with a few examples.

Example 2.2.1. The special linear groups $SL(n;\mathbb{R})$ and $SL(n;\mathbb{C})$ are the $n \times n$ invertible matrices with real or complex entries respectively having determinant one. They are subgroups of $GL(n;\mathbb{C})$ and by taking a sequence of matrices A_n with determinant one they will converge to a matrix A with determinant one due to the fact that the determinant is a continuous function. $SL(n;\mathbb{C})$ will play an important role later on.

Example 2.2.2. The special orthogonal groups $SO(n)$ and orthogonal groups $O(n)$ are the groups of all $n \times n$ matrices with real entries with orthonormal columns, i.e.

$$\sum_i A_{ij}A_{ki} = \delta_{jk},\ \forall j,\ k = 1,\ \ldots,\ n. \tag{2.2.2}$$

We have already seen an example of this in $SO(2)$. An equivalent statement is that A is orthogonal if it preserves the inner product, that is if $\boldsymbol{x'} = A\boldsymbol{x}$ and $\boldsymbol{y'} = \mathrm{A}\ \boldsymbol{y}then$

$$\boldsymbol{x'} \cdot \boldsymbol{y'} = \langle A\boldsymbol{x},\, A\boldsymbol{y}\rangle = \boldsymbol{x} \cdot \boldsymbol{y} \tag{2.2.3}$$

for all vectors $x,\, y \in \mathbb{R}^n$ where $\langle \cdot, \cdot \rangle$ denotes the ordinary inner product in $\mathbb{R}^n$. Likewise, it can also be written as $A^T A = \mathbb{1}$ where $A^T = A^{-1}$ where $(A^T)_{ij} = A_{ji}$. Also, we observe that $\det(A^T A) = \det A^T \cdot \det A = (\pm 1)^2$, because $\det A = \det A^T$. The set of all $n \times n$ orthogonal matrices with real entries is the orthogonal group $O(n)$, a subgroup of $GL(n; \mathbb{C})$. Due to that the relation $A^T A = \mathbb{1}$ is preserved under limits a sequence of such matrices $O_n \in O(n)$ converge to $O \in O(n)$ and $O(n)$ is thus a matrix Lie group.

Similarly, the special orthogonal group $SO(n)$ is the set of all $n \times n$ orthogonal matrices with real entries with determinant one. In contrast, the group elements of $O(n)$ have a determinant of ± 1. Thus $SO(n)$ can be thought of as "half" of $O(n)$. Clearly, the condition on the determinant as well as the orthogonality are preserved under limits which means that $SO(n)$ is a matrix Lie group.

It is possible to comprehend these groups in terms of their geometric realization. The elements of $O(n)$ are rotations or rotations combined with parity (because of the determinant being ± 1) while the elements of $SO(n)$ are simply rotations. This means that the elements of $SO(n)$ are **proper matrices** (a proper matrix is an orthogonal matrix with determinant one).

Example 2.2.3. The Lorentz group $O(3, 1)$ is a famous example in physics and is a subgroup of **the generalized orthogonal group** with real entries. The generalized orthogonal group $O(n, k)$ is the group of $(n + k) \times (n + k)$ real matrices, where n and k are integers, and from which one obtains a scalar product with n negative entries and k positive or vice versa. $O(1,\, 3)$ is a subgroup of $O(n,\, k)$ where $n, k \in \mathbb{N}$ which in turn is a subgroup of $GL(n+k;\, \mathbb{R})$. However, we mostly deal with the restricted Lorentz group $SO(1,\, 3)$, i.e. all group elements of $O(1, 3)$ which have determinant one, since $O(1,\, 3)$ does not preserve orientations of space and time. If compared to an ordinary special orthogonal group $SO(1, 3)$, we observe that $SO(1,\, 3)$ preserve the Minkowski inner product instead of the ordinary scalar product in Euclidean space $\mathbb{R}^n$. This subject will be developed in chapter 6.

Example 2.2.4. The special unitary group $SU(n)$ **and unitary group** $U(n)$ are the $n \times n$ matrices with complex entries with orthonormal columns. Let $*$ denote the conjugate of a complex number. If $A \in U(n)$ then

$$\sum_i A_{ij}(A_{ki})^* = \delta_{jk}. \tag{2.2.4}$$

This is equivalent to stating that A is unitary if $\langle A\boldsymbol{x}|A\boldsymbol{y}\rangle = x_i^* y_i\,,\ \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ where $\langle\ |\ \rangle$ denotes the standard scalar product in the Hilbert space $\mathbb{C}^n$, see appendix C. It is also the same as that if A is unitary, then $A^\dagger A = \mathbb{1}$ where $A^\dagger = (A^t)^*$ is the adjoint or Hermitian conjugate of A such that $A^\dagger = A^{-1}$.

If we compare with the real case of the $O(n)$ group, we find similary that for a complex unitary matrix $\det A^\dagger = (\det A)^*$ which yields $\det(A^\dagger A) = |\det A|^2 = 1$. This means that $|\det A| = 1$ and implies that A is invertible. Since all unitary matrices are invertible, the group axioms are fulfilled. The same argument as before regarding the determinant and also that the limit of unitary matrices is unitary lead us to recognize that $U(n)$ and $SU(n)$ are matrix Lie groups where the latter has determinant one. They are both subgroups of $GL(n; \mathbb{C})$.

Note that the group elements of $U(n)$ may have any determinant $e^{i\theta}$, since $|e^{i\theta}| = 1$. Hereby, the group elements are determined to the degree of a phase θ. Both $U(n)$ and $SU(n)$ will be essential in our later work, since they encode the behavior of particles in the symmetry group of the Standard Model of particle physics, $U(1) \times SU(2) \times SU(3)$. We will discover that there are in fact two different kinds of $SU(2)$ which representations describe different properties of particles in the following chapters.

2.2.1 Definition and Some Topological Notions

The examples above are of utmost importance, but they still leave us with some unresolved issues. For instance, we have simply examined matrix Lie groups and not provided a full definition of a Lie group. To answer this, let us return to the rotational example in $\mathbb{R}^2$ with the groups $U(1)$ and $SO(2)$ where we begun this section and study some interesting properties.

One can think of $SO(2)$ and $U(1)$ as a geometric object - a circle S^1 since they correspond to rotations of a certain, arbitary angle around the origin. By drawing a continuous curve between infinitesimal angle alternations the circle is formed, see figure 2.2. Formally, the branch of mathematics which studies mathematical structures manifested as preserved properties of a continuously deformed object is known as *topology* and the constructs are called *manifolds* (more on this issue soon).

Next, have a look at the group $SU(2)$ which is the group of special, unitary 2×2 matrices. Let a matrix $U \in SU(2)$ and write U as

$$U = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{2.2.5}$$

with $a, b, c, d \in \mathbb{C}$. The determinant of U is $\det U = ad - bc = 1$, and U has the inverse

$$U^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{2.2.6}$$

But since U is unitary, $U^\dagger = U^{-1}$ and thus

$$\begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}. \tag{2.2.7}$$

By equating elements we see that $c = -b^*$ and $d = a^*$. An arbitrary element of $SU(2)$ can therefore be written

$$U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \tag{2.2.8}$$

with the additional constraint that

$$\det U = ad - bc = |a|^2 + |b|^2 = 1. \tag{2.2.9}$$

Now, write $a = x_1 + ix_2$ and $b = x_3 + ix_4$ where i is the imaginary unit and $x_k \in \mathbb{R}$, $k = 1, 2, 3, 4$. Hence, the condition on the determinant becomes $x_1^2 + x_2^2 + x_3^2 + x_4^2 = 1$. This equation describes the isomorphic embedding of the unit 3-sphere in $\mathbb{R}^4$, S^3 (recall that for

$SO(2)$ we had $x^2 + y^2 = 1$ for the circle, i.e. a one-dimensional surface for two variables, therefore we get a construct with one dimension less than the number of variables). Hereby, we can conclude that the group $SU(2)$ viewed as a manifold is isomorphic to S^3 which means that every group element can be represented as a point on the surface of the 3-sphere. If one takes a multiple of transformations, they will lie on a continuous curve on the surface.

Hereby, with this geometric reasoning, we arrive at the full definition of a Lie group.

Definition 2.2.3. *A* **Lie group** *is a differentiable manifold $\mathcal{G}$ which is also a group, where the group product*

$$\mathcal{G} \times \mathcal{G} \to \mathcal{G} \ : \ (g_1, \, g_2) \mapsto g_1 \star g_2$$

and the inverse map $\mathcal{G} \to \mathcal{G} \ : \ g \mapsto g^{-1}$ are also differentiable.

We have already encountered some examples of compact manifolds such as the circle S^1 and the sphere S^3. Well then, how are these abstract concepts related to matrix Lie groups? Fortunately, there is a theorem which states the following:

Theorem 2.2.1. Every matrix Lie group (or classical Lie group) is a Lie group.

For a rigorous proof, please see appendix E. This means that we can study these Lie groups in terms of matrices, or since it is the case, their defining matrix representations. Most of the Lie groups which we will investigate in this study are matrix Lie groups and that facilitates the analysis considerably.

If we return to the rotational matrix from the beginning of this section, we observe that it was characterized by the angle θ - a smooth parameter. We see that the group elements g in a Lie group $\mathcal{G}$ may depend on a smooth parameter θ. By "smooth" we mean that there is a notion of closeness between different group elements, when the difference in θ is small (p. 43, [4]). If we explore this dependence, we will find that every Lie group can be formed by exponentiating a so-called *Lie algebra*, which includes the continuous parameter. This exciting theory will be explored in the next section, starting with studying the identity element of the Lie group on the manifold.

2.3 Lie Algebras

To every Lie group $\mathcal{G}$ there is an associated vector space $\mathfrak{g}$ known as the Lie algebra of the group. This space can be thought of as the tangent space to $\mathcal{G}$ near the identity and it consists of all infinitesimal displacements away from unity. Like a Lie group, a Lie algebra is an abstract construction, and, just like in the Lie group case, its properties are examined by use of a defining matrix representation. We illustrate the concept with a simple example before presenting the formal definition.

Consider the group $SO(2)$ that rotates a point in the plane counterclockwise an arbitrary angle ϕ about the origin, represented by the matrix

$$R(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}. \tag{2.3.1}$$

If we consider only small rotations, we can expand the sine and cosine terms to first order and obtain

$$R(\phi) \approx \begin{pmatrix} 1 & -\phi \\ \phi & 1 \end{pmatrix} = \mathbb{1} - i\phi T \tag{2.3.2}$$

where T is given by

$$T = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}. \tag{2.3.3}$$

Keep in mind that equation (2.3.2) is only valid for small values of ϕ, i.e. for small displacements away from $\mathbb{1}$. With T, we can create a finite rotation by successively compounding small rotations. If we want achieve a rotation by an angle ϕ, we can simply rotate N times through an angle $\frac{\phi}{N}$, and let N tend to infinity:

$$R(\phi) = \lim_{N\to\infty} \left(\mathbb{1} - i\frac{\phi T}{N}\right)^N = \mathrm{e}^{-i\phi T}. \tag{2.3.4}$$

By expanding the exponential in the right hand side of (2.3.4) in its Taylor series, we find (after some calculations) that

$$R(\phi) = \mathrm{e}^{-i\phi T} = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}, \tag{2.3.5}$$

i.e., we get the usual rotation matrix (2.3.1) by compounding infinitesimal rotations, as expected. Now, recall what was said at the beginning: the Lie algebra consists of all infinitesimal displacements away from unity. But this is exactly ϕT in equation (2.3.2). So we have reached the following conclusion:

$$\mathfrak{so}(2) = \{i\phi T, \phi \in \mathbb{R} \text{ and } T \text{ as in (2.3.3)}\} \tag{2.3.6}$$

is called the **Lie algebra** of the group of rotations defined by (2.3.1). Note that $\mathfrak{so}(2)$ is a linear space over the real numbers, spanned by the **generator** iT (familiar as the basis element in linear algebra).

By exponentiating an element of the Lie algebra, we obtain an element of the Lie group, and this is the way a Lie algebra is related to its Lie group. Remember that both the Lie group and the Lie algebra are abstract spaces, and exponentiating an element of the Lie algebra simply means exponentiating a representation of an element in the algebra (which is a well defined operation), whereupon we obtain a representation of an element in the Lie group.

This example hopefully gives something of a intuitive feel of what a Lie algebra is, but we will now proceed with a little more rigour. As we go, it will be helpful to look back and see how the formalism applies to the $SO(2)$ example. We begin with a definition:

Definition 2.3.1. *A function* $A: \mathbb{R} \to GL(n, \mathbb{R})$ *is called a* **one-parameter group** *if*

(i) A *is continuous*

(ii) $A(0) = \mathbb{1}$

(iii) $A(s+t) = A(s)A(t)$ *for all real* t *and* s.

It is not a coincidence that A is called a group: it is easily checked that A satisfies the group axioms. Note that $SO(2)$ is a one-parameter group. Recall from equation (2.3.5) that the matrix representation of $SO(2)$ could be written as $R(\phi) = \mathrm{e}^{\phi X}$, where X is a matrix. The following theorem, which we present without proof, states that this is not something that is unique to $SO(2)$, but is rather a general property of one-parameter groups.

Theorem 2.3.1. To every one-parameter group A there is a unique matrix X such that

$$A(t) = \mathrm{e}^{tX}. \tag{2.3.7}$$

One crucial observation is that if we have an expression for $A(t)$, then X can be obtained by simply computing the derivative of A at $t = 0$, i.e.

$$\frac{dA}{dt}|_{t=0} = X\mathrm{e}^{tX}\Big|_{t=0} = X. \tag{2.3.8}$$

We have finally reached the point where we are ready to present the definition of the Lie algebra of a matrix Lie group.

Definition 2.3.2. *If $\mathcal{G}$ is a matrix Lie group, then its Lie algebra $\mathfrak{g}$ is the set of all matrices X such that $\mathrm{e}^{tX} \in \mathcal{G} \ \ \forall t \in \mathbb{R}$.* [1]

Before we examine the meaning of this definition in more detail, we first note some very important immediate consequences of this definition. Suppose X and Y are two elements of the Lie algebra $\mathfrak{g}$ of some Lie group $\mathcal{G}$. That $X \in \mathfrak{g}$ means that $\mathrm{e}^{tX} \in \mathcal{G}$ for all real t, but then $sX \in \mathfrak{g}$ for all real s, i.e. a (real) multiple of an element in a Lie algebra is also an element of the same Lie algebra. In fact, the sum $X + Y$ also belongs to $\mathfrak{g}$. To see this, exponentiate $t(X + Y)$:

$$\mathrm{e}^{t(X+Y)} = \lim_{m\to\infty} \left(\mathrm{e}^{\frac{tX}{m}} \mathrm{e}^{\frac{tY}{m}}\right)^m. \tag{2.3.9}$$

Both of the exponentials under the limit belong to $\mathcal{G}$, and thus the entire expression under the limit belongs to $\mathcal{G}$. But the limit of a sequence of matrices in a matrix Lie group $\mathcal{G}$ belongs to the group (if the limit is invertible, see definition 2.2.1). Thus $\mathrm{e}^{t(X+Y)} \in \mathcal{G}$ for all real t, and hence $X + Y$ is an element of the Lie algebra of $\mathcal{G}$. To summarize: *A linear combination of elements in a Lie algebra belongs to the same Lie algebra.* In other words, the Lie algebra of a matrix Lie group is closed under addition and multiplication by real scalars. One should take care to note that there is a difference between real and complex Lie algebras; real Lie algebras allow linear combinations over the real numbers, and complex Lie algebras allow linear combinations over the complex numbers.

The simplest example is the Lie algebra of a one-parameter group. Recall that a one-parameter group A is defined **for all** real arguments t and that $A(t) = \mathrm{e}^{tX}$, thus according to definition 2.3.2, X is an element of the Lie algebra of A, and X can be computed with equation (2.3.8).

Let's return for a moment to the $SO(2)$ example. Equation (2.3.1) gives us an explicit description of $SO(2)$ in terms of a one-parameter group, and we can thus immediately find its Lie algebra as

[1] In physics literature it is customary to let the Lie algebra of a matrix Lie group $\mathcal{G}$ be the set of matrices X such that e^{itX} belongs to $\mathcal{G}$ for all real t.

$$X = \frac{dR}{d\phi}\Big|_{\phi=0} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{2.3.10}$$

But this is a generator of the Lie algebra we found for $SO(2)$ in equation (2.3.6). Let us now look at the Lie algebras of a few other familiar groups.

Example 2.3.1. Another example of a one-parameter group is the unitary group of order one, U(1). Recall from the previous section that this group consists of all complex numbers of the form $\mathrm{e}^{i\theta}$, where θ is a real number, and that this group is isomorphic to $SO(2)$. The group is obviously parameterized by

$$R(\theta) = \mathrm{e}^{i\theta}, \tag{2.3.11}$$

and from this we find by using equation (2.3.8) that the generator of the Lie algebra $\mathfrak{u}(1)$ of U(1) is i, and thus $\mathfrak{u}(1)$ (it is conventional to denote a Lie algebra in lowercase gothic letters) consists of all pure imaginary numbers.

Since $U(1)$ is isomorphic to $SO(2)$ we might expect so see some relationship between their Lie algebras, and we can make one interesting observation. Note that the generator of $\mathfrak{u}(1)$ is i, which when squared becomes $i^2 = -1$, i.e. the negative of the identity. Similarly, we see that the generator (2.3.10) of $\mathfrak{so}(2)$ satisfies

$$X^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -\mathbb{1}, \tag{2.3.12}$$

i.e. just like for $\mathfrak{u}(1)$, the square of the generator of $\mathfrak{su}(2)$ equals the negative of the identity.

As a parenthesis, we can mention that this kind of structure that squares to $-\mathbb{1}$ in a vector space is called a *complex structure*.

Example 2.3.2. Recall that the general linear group $GL(n,\mathbb{C})$ consists of all invertible $n \times n$ matrices with complex entries. Since the exponential of *any* $n \times n$ matrix is invertible, it follows that e^{tX} is invertible for all $n \times n$ matrices X and all real numbers t. Thus the Lie algebra of $GL(n, \mathbb{C})$, denoted by $\mathfrak{gl}(n,\mathbb{C})$ consists of all $n \times n$ matrices with complex entries.

Before finding the (complex) Lie algebra of $SU(2)$, we note that if we have found a parametrization of a matrix Lie group $\mathcal{G}$ in a number of parameters, i.e. $R = R(t_1, t_2, \ldots, t_n)$, with $R(0,0,\ldots,0) = \mathbb{1}$, then $R(0, \ldots, 0, t_k, 0, \ldots, 0)$ is a one parameter group $\mathcal{G}_k \subset \mathcal{G}$ and hence there is a matrix X_k such that $\mathrm{e}^{tX_k} \in \mathcal{G}_k$ for all real t, or, more specifically since $\mathcal{G}_k$ is a subset of $\mathcal{G}$, $\mathrm{e}^{tX_k} \in \mathcal{G} \ \ \forall t \in \mathbb{R}$. In other words, X_k is an element of the Lie algebra of $\mathcal{G}$.

But X_k can be computed as

$$X_k = \frac{dR(0, \ldots, 0, t_k, 0, \ldots, 0)}{dt_k}\Big|_{t_k=0}. \tag{2.3.13}$$

To sum it up: if we have a parametrization of a matrix Lie group in terms of n different parameters, we can obtain the generators of the Lie algebra of the group by simply differentiation the parameterization with respect to the n different parameters at the identity element. We illustrate the method by computing the generators of the Lie algebra of $SU(2)$.

Example 2.3.3. Recall that $SU(2)$ is the Lie group whose defining representation is that of 2×2 unitary matrices with determinant equal to one. If we can find the general form of an

element of $SU(2)$, in terms of a number of n parameters, we can then use equation (2.3.13) to find the generators of its Lie algebra, denoted by $\mathfrak{su}(2)$.

As a convenient matter of fact, some of the work was already done in the section 2.2.1. Since $SU(2)$ is isomorphic to the sphere S^3 we know a priori that the number of parameters has to be three. By proceeding from equation (2.2.9) we note that it is just the equation of the unit circle, and so there exists a real η such that $|a| = \cos\eta$ and $|b| = \sin\eta$, and hence

$$a = \mathrm{e}^{i\xi}\cos\eta \tag{2.3.14a}$$

$$b = \mathrm{e}^{i\zeta}\sin\eta \tag{2.3.14b}$$

for real but otherwise arbitrary ξ, η and ζ.

With this and equation 2.2.8 we have now found the general form of the elements of $SU(2)$, parameterized by the real numbers η, ζ and ξ:

$$U = \begin{pmatrix} \mathrm{e}^{i\xi}\cos\eta & \mathrm{e}^{i\zeta}\sin\eta \\ -\mathrm{e}^{-i\zeta}\sin\eta & \mathrm{e}^{-i\xi}\cos\eta \end{pmatrix}. \tag{2.3.15}$$

Finally, we obtain the generators of $SU(2)$ by differentiating (2.3.15) at the identity ($\eta = 0$, $\zeta = 0$, $\xi = 0$).

So the generators of $\mathfrak{su}(2)$ are

$$\frac{\partial}{\partial\eta}U: \qquad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = i\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = i\sigma_2 \tag{2.3.16a}$$

$$\frac{\partial}{\partial\xi}U: \qquad \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = i\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = i\sigma_3. \tag{2.3.16b}$$

However, note that if we differentiate (2.3.15) at the unity we obtain a matrix with all elements equal to zero. To avoid this and obtain the last generator of $\mathfrak{su}(2)$ we must parametrize $SU(2)$ in another way. Recall that a general element of $SU(2)$ can be written in the form

$$U = \begin{pmatrix} a & b \\ -b^* & a^* \end{pmatrix} \tag{2.3.17}$$

with the constraint that $|a|^2 + |b|^2 = 1$. Now, to find the last generator, we let $b = i\beta$ and $a = \sqrt{1-\beta^2}$. The generator thus becomes

$$\frac{\partial}{\partial\beta}\begin{pmatrix} \sqrt{1-\beta^2} & i\beta \\ i\beta & \sqrt{1-\beta^2} \end{pmatrix}\Big|_{\beta=0} = i\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = i\sigma_1. \tag{2.3.18}$$

The matrices denoted σ_1, σ_2 and σ_3 are called the **Pauli matrices**, and with a factor $\frac{i}{2}$ in front they occur as, for example, the generators of angular momentum in quantum mechanics. This is to be investigated later on in section 4.3.1. $\mathfrak{su}(2)$ has three generators and is thus three dimensional.

The Pauli matrices obey the following relations

$$\begin{aligned} \sigma_1\sigma_1 = \sigma_2\sigma_2 = \sigma_3\sigma_3 = \mathbb{1} \\ [\sigma_a, \sigma_b] = 2i\epsilon_{abc}\sigma_c \end{aligned} \tag{2.3.19}$$

where ϵ_{abc} is the anti-symmetric Levi-Civita tensor.

The generators T of $\mathfrak{su}(2)$ obey the commutation relations

$$[T_a, T_b] = -2\epsilon_{abc}T_c. \tag{2.3.20}$$

The Lie algebra is a linear space because a *linear combination of two elements of the Lie algebra corresponds to a composition of elements in the Lie group.* One might ask whether the property $\mathrm{e}^{X+Y} = \mathrm{e}^X\mathrm{e}^Y$ is still true for x and y in the Lie algebra. It turns out that this is true when X and Y commute, but *not* in the general case.

The composition of $\mathrm{e}^X\mathrm{e}^Y$ is given by the **Baker-Campbell-Hausdorff formula**

$$\mathrm{e}^X\mathrm{e}^Y = \mathrm{e}^{X+Y+\frac{1}{2}[X,Y]+\ldots} \tag{2.3.21}$$

where $X, Y \in \mathfrak{g}$ and ... denotes higher order terms.

From definition 2.3.2 we know that by exponentiating an element of the matrix Lie algebra, we obtain an element of the corresponding matrix Lie group. A natural question is whether the mapping

$$\exp:\ \mathfrak{g} \to \mathcal{G} \tag{2.3.22}$$

is one-to-one and onto. In other words, by exponentiating an element of the Lie algebra $\mathfrak{g}$ we obtain an element of the Lie group $\mathcal{G}$, but can all elements of $\mathcal{G}$ be reached in this way, ie. does it exist an element $X \in \mathfrak{g}$ such that $\mathrm{e}^X = G$ for all G in $\mathcal{G}$? And if so, is this X unique?

The answer to both of these questions is *no*, the exponential mapping (2.3.22) is neither one-to-one nor onto. While this is not true in general, it is true that the exponential mapping is always locally one-to-one and onto (in the special case when the exponential map is onto, i.e. when it covers the whole group, the group is said to be an *exponential group*). Stated formally:

Theorem 2.3.2. If $\mathcal{G}$ is a matrix Lie group with a corresponding Lie algebra $\mathfrak{g}$, there exists a neighbourhood U of zero in $\mathfrak{g}$ and a neighbourhood V of the identity in $\mathcal{G}$ such that the mapping

$$\exp:\ U \to V \tag{2.3.23}$$

is homomorphic.

While we can not obtain all elements of $\mathcal{G}$ by exponentiating an element of $\mathfrak{g}$, it turns out that all elements of $\mathcal{G}$ *can* be created by compounding elements of $\mathcal{G}$ that can be written as e^X, with X in the Lie algebra. In other worlds:

Theorem 2.3.3. If $\mathcal{G}$ is a connected matrix Lie group, then every element A of $\mathcal{G}$ can be written as

$$A = \mathrm{e}^{X_1}e^{X_2}\ldots\mathrm{e}^{X_n} \tag{2.3.24}$$

where $X_1, X_2, \ldots, X_n \in \mathfrak{g}$.

So while not all elements of $\mathcal{G}$ can be obtained directly from $\mathfrak{g}$, every element of $\mathcal{G}$ can be created by compounding elements of $\mathcal{G}$ that can.

We will now present the formal definition of a Lie algebra:

Definition 2.3.3. *A **Lie algebra** $\mathfrak{g}$ (real or complex) is a vector space over $\mathbb{R}$ or $\mathbb{C}$, together with a binary operation, [.,.], called the **Lie bracket**, satisfying*

(i) Closure:

$$[.,.] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g} \tag{2.3.25}$$

(ii) Bilinearity:

$$[ax + bx', y] = a[x, y] + b[x', y] \tag{2.3.26a}$$

$$[x, ay + by'] = a[x, y] + b[x, y'] \tag{2.3.26b}$$

where a,$b \in \mathbb{R}$ or $\mathbb{C}$, x,x',y,$y' \in \mathfrak{g}$

(iii) Alternating:

$$[x, x] = 0 \quad \forall x \in \mathfrak{g} \tag{2.3.27}$$

(iv) Jacobi identity:

$$[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0 \quad \forall x,y,z \in \mathfrak{g} \tag{2.3.28}$$

But in the $SO(2)$ example, what is the Lie bracket? Since we are working with matrices, we can simply take the Lie bracket to be the ordinary commutator of matrices:

$$[x, y] = xy - yx \tag{2.3.29}$$

and it can easily be shown that the commutator satisfies the four axioms in definition 2.3.3.

But once again we should remind ourselves that the elements of the Lie algebra are *not* matrices; the matrices are just *representations* of the group elements. In fact, representations of Lie algebras are defined in such a way that the representation of the element $[x,y]$ of the algebra is $[\rho(x), \rho(y)]$ where $\rho(x)$ and $\rho(y)$ are the representations (matrices) of x and y respectively, and the bracket is the usual commutator. But we will deal with representation in more depth in the following chapter.

Another important concept for a Lie algebra is its **structure constants** with respect to a given basis of the algebra. Let $T_{i i=1}^{n}$ be a basis for a Lie algebra $\mathfrak{g}$, then the commutator between any two Ts can be expressed as a linear combination

$$[T_i, T_j] = \sum_{k=1}^{n} C_{i,j}^{k} T_k \tag{2.3.30}$$

where the constants $C_{i,j}^{k}$ are called the structure constants. Note that the antisymmetry of the commutator implies that $C_{i,j}^{k} = -C_{j,i}^{k}$. The structure constants determine the Lie algebra.

Example 2.3.4. It turns out that the Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$ have the same structure constants and are thus isomorphic. Since rotations in three dimensions have three degrees of freedom, $\mathfrak{so}(3)$ will obviously be three dimensional, and to obtain the generators we need three different one-parameter groups in $SO(3)$, and we can simply take these to be rotations about the $x-$, $y-$ and $z-$axes, parameterized with the usual rotation matrices, and by computing the derivative of these at the identity the Lie algebra is easily obtained.

By computing the commutation relations one finds that they are exactly the same as those for $\mathfrak{su}(2)$, and hence the two Lie algebras are isomorphic.

One may ask if this in turn implies that the two Lie groups $SO(3)$ and $SU(2)$ are isomorphic. The answer to this is no. It does imply that they are *locally* isomorphic, but not globally. $SU(2)$ is the *double cover* of $SO(3)$, which roughly means that to every element in $SO(3)$ there exists two corresponding elements in $SU(2)$, i.e. we have the isomorhpism $SU(2) \cong SO(2)/\mathbb{Z}_2$. This isomorphism will be explored further in section 6.7.2.

A given real Lie algebra can also be *complexified*. By the complexification $\mathfrak{g}_{\mathbb{C}}$ of a real Lie algebra $\mathfrak{g}$ we simply mean the set of elements of the form $X + iY$ where X and Y belong to $\mathfrak{g}$. The Lie bracket has a unique extension from $\mathfrak{g}$ to $\mathfrak{g}_{\mathbb{C}}$, given by

$$[X_1 + iX_2,\, Y_1 + iY_2] = ([X_1,\, Y_1] - [X_2,\, Y_2]) + i([X_1,\, Y_2] + [X_2,\, Y_1]). \tag{2.3.31}$$

Note that (2.3.31) is the only possible way to define the Lie bracket on $\mathfrak{g}_{\mathbb{C}}$ if it is to remain bilinear. That the other two properties of the Lie bracket still hold is a matter of a simple calculation and is left as an exercise to the reader.

The complexification of a real Lie algebra can simply be thought of as allowing linear combinations over the complex numbers.

Example 2.3.5. The Lie algebra $\mathfrak{sl}(2, \mathbb{C})$ is obtained as the complexification of $\mathfrak{su}(2)$, i.e.

$$\mathfrak{sl}(2, \mathbb{C}) = \mathfrak{su}(2) \oplus i\mathfrak{su}(2). \tag{2.3.32}$$

This Lie algebra is three dimensional and its generators, often denoted e, f, h satisfy the following commutation relations:

$$\begin{aligned} [e,\, f] &= h \\ [h,\, e] &= 2e \\ [h,\, f] &= -2f. \end{aligned} \tag{2.3.33}$$

Since the commutation relations (2.3.33) are real, we can define the real Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ as the set of all real linear combinations of e, f and h, and $\mathfrak{sl}(2, \mathbb{C})$ as the set of all complex linear combinations of e, f and h.

We will encounter this algebra more in later chapters.

In this chapter we have laid the foundations of the theory of Lie groups and Lie algebras. We looked at Lie groups, which are groups that depend on one or more continuous parameters, and their corresponding Lie algebras, the elements corresponding to infinitesimal displacements away from the unity of the group and saw how these are related via the exponential mapping. In the next chapter we look more into the theory of representations that we alluded to in this chapter.

Chapter 3

Representation Theory

In group theory we chiefly deal with the group element's *representations* rather than the group elements themselves. Our endeavor in physics is to describe rotations, translations etc and these actions are interpreted through representations of group elements. Hence, learning about representations is a crucial ingredient to understand the patterns of the universe. One of the chief applications of representation theory is to identify the symmetries of a given problem. If we locate a set of symmetries we know that they constitute a group and having an idea about which representations that most likely occur in the specific space one may simplify the problem substantially. In this chapter we will develop some fundamental formalism regarding this theory.

A *representation* is a mapping of the group elements $g \in \mathcal{G}$ onto a set of linear operators which preserves the group properties and the multiplication table. Moreover, it is necessary that e is mapped to 1 where e is the unit element of the group $\mathcal{G}$ and 1 is the unit element of the representation. In physics, the linear operator is often a matrix and operations within the group are easily manifested as matrix multiplications. In fact the representation consists of two parts, the *realization* which is/are the linear operator(s) and the *module* - some object which the realization *acts* on. The module is often a vector space when it occurs in physics. However, there are several different objects which may serve as that. For instance, the 180° rotation of a picture in a plane, introduced in section 2.1, is an object on which the $\mathbb{Z}_2$ group acts upon and there could also be tensors, spinors[1] and even a Lie algebra itself subjected to actions of some linear operator. Yet, since the realization and the module always appear together, physicists ubiquitously tend to refer to both objects with "representation".

In this chapter we will exclusively work with *matrix representations*, i.e. matrices acting on vectors in vector spaces, and present several examples of representations. The aim of this chapter is to give a mathematical foundation for representation theory which can be applied in the following chapters. We begin with some formal definitions.

[1] See section 6.5.

3.1 Basics

Definition 3.1.1. *A* **finite dimensional complex matrix representation** Π *of a Lie group* $\mathcal{G}$ *acting on a finite dimensional complex vector space* V*, dim* $V \geqslant 1$*, is a Lie group homomorphism*

$$\Pi : \mathcal{G} \longrightarrow GL(n; \mathbb{C}) \tag{3.1.1}$$

or more generally $\Pi : \mathcal{G} \longrightarrow GL(V)$*. Because of the Lie group homomorphism we have that*

$$\Pi(g_1)\Pi(g_2) = \Pi(g_1 g_2), \ \forall g_1, g_2 \in \mathcal{G}. \tag{3.1.2}$$

A **finite dimensional complex matrix representation** π *of a Lie algebra* $\mathfrak{g}$ *acting on a finite dimensional complex vector space* V*, dim* $V \geqslant 1$*, is a Lie algebra homomorphism*

$$\pi : \mathfrak{g} \longrightarrow \mathfrak{gl}(n; \mathbb{C}) \tag{3.1.3}$$

or more generally $\pi : \mathfrak{g} \longrightarrow \mathfrak{gl}(V)$*. In the same way as for the Lie group, homomorphism implies*

$$[\pi(t_1), \pi(t_2)] = \pi([t_1, t_2]), \ \forall t_1, t_2 \in \mathfrak{g}. \tag{3.1.4}$$

If there is a bijection between the group $\mathcal{G}$ *and the representation* Π *then the representation is called* **faithful***. Analogously, if there is a bijection between the algebra* $\mathfrak{g}$ *and the representation* π *then the representation is called* **faithful***. A representation is thus a mapping from the abstract group elements or abstract Lie algebra elements*[2] *to a linear operator which acts on a vector space* V*.*

Note 1. Observe that $g_1 g_2 = g_1 \star g_2$ is the group product operation, and that $\Pi(g_1)\Pi(g_2)$ is ordinary matrix multiplication or the multiplication defined for the specific linear operators. Even more important: observe that $[t_1, t_2]$ is *not* the commutator. It is the abstract Lie bracket, but of course $[\pi(t_1), \pi(t_2)]$ is the usual commutator of two operators.

We saw in section 2.3 how to relate a Lie group to its Lie algebra by exponentiation. A natural question then is whether the representations of Lie group and its Lie algebra have a similar relation. We claimed in section 2.3 that by exponentiating a representation of the Lie algebra we obtain a representation of the Lie group. But is this really true? To answer this question, we are in need of an important result.

Theorem 3.1.1. If $\mathcal{G}$ and $\mathcal{H}$ are two Lie groups with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$, respectively, and we have the Lie group homomorphism $\Phi : \mathcal{G} \longrightarrow \mathcal{H}$, then there exists a *unique* real linear map $\Phi' : \mathfrak{g} \longrightarrow \mathfrak{h}$ such that

$$\Phi(e^t) = e^{\Phi'(x)}. \tag{3.1.5}$$

This theorem tells us that when we have a Lie group homomorphism between two Lie groups then there exists a Lie algebra homomorphism between the corresponding Lie algebras. With this result we can answer the question above.

Proposition 1: A Lie group $\mathcal{G}$ with a representation Π and a Lie algebra $\mathfrak{g}$ with a representation π has the following relation between their representations:

$$e^{\pi(t)} = \Pi(e^t), \tag{3.1.6}$$

$t \in \mathfrak{g}$. Moreover, the representation π is *unique*.

[2] Depends on whether we consider a representation for the group or for the Lie algebra.

Proof. We use the results of the last theorem. Let $\Pi = \Phi$ and $\mathrm{GL}(V) = \mathcal{H}$,
i.e. $\Pi : \mathcal{G} \longrightarrow \mathrm{GL}(V)$. The associated Lie algebra homomorphism is then $\pi : \mathfrak{g} \longrightarrow \mathrm{gl}(V)$, π is a representation. From this we get

$$e^{\pi(t)} = \Pi(e^t). \tag{3.1.7}$$

□

The fact that π is unique follows from theorem 3.1.1. Before we continue, we need a concept that will be used extensively later.

Definition 3.1.2. *Consider a group $\mathcal{G}$ and a representation Π of $\mathcal{G}$ acting on a vector space V with a subspace W. The subspace W is* **invariant** *if $\Pi(g_i)w \in W$ for all $g_i \in \mathcal{G}$ and for all $w \in W$. If $W \neq \{0\}$ and $W \neq V$, then it is called a* **non-trivial subspace***. The representation Π is then a* **reducible** *representation. A representation with no non-trivial subspaces is called* **irreducible***.*

One can say that we have an irreducible representation when *all* the elements in our vector space can be reached when we act on *any* of the elements in our vector space with all our group elements. In other words, we can transform any element in our vector space to every other element in the vector space through all the transformations from the group. It is important to realize that a representation involves both operators and a **module**, e.g. matrices. Moreover, a group can different representations with different dimension of the module.

As soon as we have an expression of how the group elements act on an object we have a representation. In physics we like to think of representations in the following way: to every group element we associate a linear operator, matrices, acting on a vector space V. In fact, we have already encountered several representations such as the matrices for $SU(2)$. Recall that $SU(2)$ is the group of unitary matrices with determinant equal to unity which acts on the vector space $\mathbb{C}^2$. But this is exactly what a representation is and from now on we shall think of the unitary matrices, $SU(2)$, as a representation. In fact, this is the representation that defines the group, see example 3.1.2. Let us look at some examples of representations.

Example 3.1.1. The trivial representation: If $\mathcal{G}$ is a Lie group and $\mathfrak{g}$ is the Lie algebra, then

$$\Pi : \mathcal{G} \longrightarrow GL(1, \mathbb{C}) \tag{3.1.8}$$

is the **trivial representation** of the Lie group $\mathcal{G}$ by $\Pi(g_i) = \mathbb{1}$, for all $g_i \in \mathcal{G}$ acting on the vector space $\mathbb{C}$ and

$$\pi : \mathfrak{g} \longrightarrow \mathfrak{gl}(1, \mathbb{C}) \tag{3.1.9}$$

is the **trivial representation** of the Lie algebra $\mathfrak{g}$ by $\pi(t_i) = 0$, for all $t_i \in \mathfrak{g}$ acting on the vector space $\mathbb{C}$.

Example 3.1.2. The defining representation: Let $\mathcal{G}$ be a Lie group and $\mathfrak{g}$ be its Lie algebra, then

$$\Pi : \mathcal{G} \longrightarrow \Pi(\mathcal{G}) = \mathcal{G} \subset GL(n, \mathbb{C}) \tag{3.1.10}$$

is **the defining representation** of $\mathcal{G}$. For example: the defining representation of $SO(3)$ is the representation in which $SO(3)$ acts on $\mathbb{R}^3$ by the orthogonal 3×3 matrices. The defining

representation for $SU(2)$ is the representation in which $SU(2)$ acts by the unitary 2×2 matrices on $\mathbb{C}^2$. In the same way we have for the Lie algebra that the defining representation for $\mathfrak{g}$ is $\mathfrak{gl}(n, \mathbb{C})$, i.e

$$\pi : \mathfrak{g} \longrightarrow \pi(\mathfrak{g}) = \mathfrak{g} \subset \mathfrak{gl}(n, \mathbb{C}). \tag{3.1.11}$$

Example 3.1.3. The adjoint representation: Consider a Lie group $\mathcal{G}$ with a Lie algebra $\mathfrak{g}$. The mapping:

$$\mathrm{Ad} : \mathcal{G} \longrightarrow GL(\mathfrak{g}), \tag{3.1.12}$$

is a Lie group homomorphism defined by

$$\mathrm{Ad}(g)t = gtg^{-1}, \tag{3.1.13}$$

$g \in \mathcal{G}$ and $t \in \mathfrak{g}$, which constitutes **the adjoint representation** of $\mathcal{G}$ acting on the vector space formed by the Lie algebra, i.e $\mathrm{Ad}(g) : \mathfrak{g} \longrightarrow \mathfrak{g}$. Moreover, the mapping

$$\mathrm{ad} : \mathfrak{g} \longrightarrow \mathfrak{gl}(\mathfrak{g}) \tag{3.1.14}$$

is a Lie algebra homomorphism defined by

$$\mathrm{ad}(t)h = [t, h], \tag{3.1.15}$$

for all t,$h \in \mathfrak{g}$, i.e $\mathrm{ad}(t) : \mathfrak{g} \longrightarrow \mathfrak{g}$. This is **the adjoint representation** of the Lie algebra $\mathfrak{g}$. The module of the representation is the algebra itself which makes the dimension of the representation equal to the dimension of the Lie algebra. A representation must satisfy the same commutation relations as the group, so to show that this is a representation we have to show that

$$[t, h] = z \quad \Rightarrow \quad [\mathrm{ad}(t), \mathrm{ad}(h)] = \mathrm{ad}(z), \tag{3.1.16}$$

for all t,h,$z \in \mathfrak{g}$, e.i that the commutation relations are preserved. $[\cdot,\cdot]$ in the first expression stands for the lie brackets and $[\cdot,\cdot]$ in the second expression stands for the usual commutator.

$$\begin{aligned} &[\mathrm{ad}(t), \mathrm{ad}(h)]w = (\mathrm{ad}(t)\mathrm{ad}(h) - \mathrm{ad}(h)\mathrm{ad}(t))w = \\ &= \mathrm{ad}(t)[h, w] - \mathrm{ad}(h)[t, w] = [t, [h, w]] - [h, [t, w]] = [t, [h, w]] + [[t, w], h] \\ &= /\text{Jacobi identity}/ = -[w, [t, h]] = [[t, h], w] = [z, w] = \mathrm{ad}(z)w, \end{aligned} \tag{3.1.17}$$

where we have used $[t, h] = z$ and that $w \in \mathfrak{g}$.

The adjoint representation is one of the most important representation because it acts on the module by commuting Lie algebra elements. Since the algebra is defined from its commuting relations, we can use this to derive a Lie groups Lie algebra. In chapter 6 we will do this explicitly.

A more precise example is the representation of the Lie algebra $\mathfrak{su}(2)$. We have seen in section 2.3 that the generators of $\mathfrak{su}(2)$, i.e. the Lie algebra, can be represented by the Pauli matrices, multiplied with an i, acting on the 2 dimensional complex vector space. The Pauli matrices multiplied with $\hbar/2$ are in fact the matrix expression for the spin operators in quantum mechanics and this is what we mean when we say that the Pauli matrices multiplies with an i are the generators of spin. We will continue to make use of linear operators and we will denote them by $\hat{A}$, i.e. by a "hat". The spin operators $\hat{S}_x$,$\hat{S}_y$,$\hat{S}_z$ can thus be expressed as the following example suggests.

Example 3.1.4. The Pauli matrices: The map

$$\pi : \mathfrak{su}(2,\mathbb{C}) \longrightarrow \mathfrak{gl}(2,\mathbb{C}) \tag{3.1.18}$$

gives the three 2×2 self-adjoint matrices

$$\pi(i\hat{S}_x) = i\frac{\hbar}{2}\sigma_x = i\frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \pi(i\hat{S}_y) = i\frac{\hbar}{2}\sigma_y = i\frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad , \pi(i\hat{S}_z) = i\frac{\hbar}{2}\sigma_z = i\frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{3.1.19}$$

that forms the 2-dimensional defining representation for $\mathfrak{su}(2)$. $\hbar$ is Planck´s constant h divided by 2π.

Example 3.1.5. The defining representation of SO(3) is as mentioned above the 3×3 orthogonal matrices with determinant equal to unity. For example, a rotation about the z-axis, y-axis, x-axis by a finite angle ϕ is represented, respectively, as

$$\begin{pmatrix} \cos(\phi) & -\sin(\phi) & 0 \\ \sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} \cos(\phi) & 0 & -\sin(\phi) \\ 0 & 1 & 0 \\ \sin(\phi) & 0 & \cos(\phi) \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{pmatrix}. \tag{3.1.20}$$

Example 3.1.6. A general method of finding a matrix representation for an operator $\hat{A}$ acting on a finite vector space V is to study how $\hat{A}$ acts on the elements of V. Because an arbitrary element of V can be expressed as a linear combination of the elements of the basis, it is sufficient that we examine $\hat{A}$:s action on the base elements. Let $|\Psi_i\rangle$, $i = 1, 2, ..., n$ be a basis for V and $\dim(V) = n$. The first column in the representation matrix A is the coefficients for $\hat{A}|\Psi_1\rangle$ in our particular basis, i.e. $A_{i1} = \langle\Psi_i|\hat{A}\Psi_1\rangle$. The i,j:th component of A is thus $\langle\Psi_i|\hat{A}\Psi_j\rangle$ and we get

$$\begin{pmatrix} \langle\Psi_1|\hat{A}\Psi_1\rangle & \langle\Psi_1|\hat{A}\Psi_2\rangle & \dots & \langle\Psi_1|\hat{A}\Psi_n\rangle \\ \langle\Psi_2|\hat{A}\Psi_1\rangle & \langle\Psi_2|\hat{A}\Psi_2\rangle & \dots & \langle\Psi_2|\hat{A}\Psi_n\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle\Psi_n|\hat{A}\Psi_1\rangle & \langle\Psi_n|\hat{A}\Psi_2\rangle & \dots & \langle\Psi_n|\hat{A}\Psi_n\rangle \end{pmatrix}. \tag{3.1.21}$$

Of course, the matrix A depends on our choice of basis for V. If we choose another basis for V we get a different representation but with the same dimension and we can relate these two representations through a **similarity transform**, $A' = SAS^{-1}$. A' is the transformed matrix and S is the change of coordinate matrix. That is, whose columns are the coordinate vectors of the old base expressed in the new base. A' and A are said to be **equivalent** representations.

In order to proceed with representation theory, we need to define even more concepts.

Definition 3.1.3. *Consider a Lie group $\mathcal{G}$, or a Lie algebra $\mathfrak{g}$, and a representation Π, or π, acting on a vector space V. Let W be an invariant subspace to V and U be an invariant subspace to W. If there exists an invariant subgroup $\bar{U}$, $\bar{U} \cap U = \{0\}$, such that $W = U \oplus \bar{U}$, then Π is called* **completely reducible**.

From a physical perspective it is interesting to see how this works on a Hilbert space $\mathcal{H}$ of finite dimension N.

Proposition 2: A *unitary* representation Π of a Lie group $\mathcal{G}$ acting on a finite dimensional complex Hilbert space $\mathcal{H}$ is completely reducible.

Proof. Since $\mathcal{H}$ is a Hilbert space we have an inner product $\langle\psi_i|\psi_j\rangle$, for all i, j. We can divide the Hilbert space as $\mathcal{H} = \mathcal{H}_\perp^{(1)} \oplus \mathcal{H}^{(1)}$, where $\mathcal{H}_\perp^{(1)}$ is the orthogonal complement to $\mathcal{H}^{(1)}$ and $\mathcal{H}_\perp^{(1)}$ is an assumed invariant subgroup. Once again, divide $\mathcal{H}^{(1)} = (\mathcal{H}_\perp^{(2)} \cap \mathcal{H}^{(1)}) \oplus \mathcal{H}^{(2)}$ where $\mathcal{H}^{(2)}$ is another assumed invariant subspace. Let $|\psi_3\rangle \in \mathcal{H}_\perp^{(2)} \cap \mathcal{H}^{(1)} = \mathcal{H}^{(3)}$ and $|\psi_2\rangle \in \mathcal{H}^{(2)}$ and consider $\langle\psi_2|\Pi(g)|\psi_3\rangle$, $g \in \mathcal{G}$. We now want to show that $\langle\psi_2|\Pi(g)|\psi_3\rangle = 0$, for all $g \in \mathcal{G}$, since every element in $\mathcal{H}^{(3)}$ is orthogonal to $\mathcal{H}^{(2)}$.

$$\langle\psi_2|\Pi(g)|\psi_3\rangle = \langle\Pi(g)^\dagger\psi_2|\psi_3\rangle = \langle\Pi(g)^{-1}\psi_2|\psi_3\rangle = 0, \tag{3.1.22}$$

since $\Pi(g)^{-1}$ is a representation for another group element and $\mathcal{H}^{(2)}$ is invariant. Thus $\mathcal{H}^{(3)}$ is invariant under Π. □

3.2 Direct Sum and Tensor Product

Even though we have used "direct sum of vector spaces" several times it is probably a good idea to give a proper definition. In general, the **direct sum of modules** is the result of combining several modules to form a larger module. If the module is a vector space, i.e. over a field, we say that we have a **direct sum of vector spaces**.

Definition 3.2.1. *Let $V^{(1)}$ and $V^{(2)}$ be two vector spaces over a field K. The **direct sum of the vector spaces** $V^{(1)}$ and $V^{(2)}$ is such that*

$$\begin{aligned} &i)\ \ (v^{(1)} \oplus v^{(2)}) + (w^{(1)} \oplus w^{(2)}) = (v^{(1)} + w^{(1)}) \oplus (v^{(2)} + w^{(2)}) \\ &ii)\ \ (av^{(1)} \oplus av^{(2)}) = a(v^{(1)} \oplus v^{(2)}), \end{aligned} \tag{3.2.1}$$

where $a \in K$, $v^{(1)}$, $w^{(1)} \in V^{(1)}$ and $v^{(2)}$, $w^{(2)} \in V^{(2)}$. The resulting vector field, i.e. the direct sum, is denoted $V^{(1)} \oplus V^{(2)}$. If our vector spaces $V^{(1)}$ and $V^{(2)}$ have an inner product, such as Hilbert spaces, we can bring this property to the direct sum of the vector spaces as well. Consider a finite number of Hilbert spaces $\mathcal{H}^{(1)}$, $\mathcal{H}^{(2)}$,..., $\mathcal{H}^{(n)}$. The direct sum of these Hilbert spaces can then form a Hilbert space if we define the inner product as

$$\begin{aligned} &\langle(x^{(1)} \oplus x^{(2)} \oplus ... \oplus x^{(n)}),(y^{(1)} \oplus y^{(2)} \oplus ... \oplus y^{(n)})\rangle = \\ &= \langle x^{(1)},y^{(1)}\rangle + \langle x^{(2)},y^{(2)}\rangle + ...\langle x^{(n)},y^{(n)}\rangle. \end{aligned} \tag{3.2.2}$$

This means that the terms in the summation are orthogonal subspaces to $\mathcal{H}^{(1)} \oplus \mathcal{H}^{(2)} \oplus ... \oplus \mathcal{H}^{(n)}$.

We have seen that a group $\mathcal{G}$ can have different representations, Π_i, acing on different vector spaces, V_i. A nice way to construct a new representation Π from the Π_i:s we have already found is to take the direct sum of these representations and let it act on the vector space V of higher dimension.

Definition 3.2.2. *Consider a group $\mathcal{G}$ with representations $\Pi_1, \Pi_2,..., \Pi_n$ acting on vector spaces V_1,V_2,..., V_n respectively. The* **direct sum** *of the representations $\Pi = \Pi_1 \oplus \Pi_2 \oplus ... \oplus \Pi_n$ then acts on the* **direct sum** *of the vector spaces $V = V_1 \oplus V_2 \oplus ... \oplus V_n$, as a representation, defined by*

$$\begin{aligned} \Pi(g)(\psi_1,\psi_2,...,\psi_n) &= \Pi_1(g) \oplus \Pi_2(g) \oplus ... \oplus \Pi_n(g)(\psi_1,\psi_2,...,\psi_n) = \\ &= (\Pi_1(g)\psi_1,\Pi_2(g)\psi_2,...,\Pi_n(g)\psi_n), \end{aligned} \tag{3.2.3}$$

for all $g \in \mathcal{G}$. Let $\mathfrak{g}$ be a Lie algebra with representations $\pi_1, \pi_2,..., \pi_m$ acting on vector spaces V_1,V_2,..., V_m respectively. The **direct sum** *of the representations $\pi = \pi_1 \oplus \pi_2 \oplus ... \oplus \pi_m$ then acts on the* **direct sum** *of the vector spaces $V = V_1 \oplus V_2 \oplus ... \oplus V_m$ as a representation, defined by*

$$\begin{aligned}\pi(g)(\psi_1,\psi_2,...,\psi_m) &= \pi_1(t) \oplus \pi_2(t) \oplus ... \oplus \pi_n(t)(\psi_1,\psi_2,...,\psi_m) = \\ &= (\pi_1(t)\psi_1,\pi_2(t)\psi_2,...,\pi_n(t)\psi_m),\end{aligned} \tag{3.2.4}$$

for all $t \in \mathfrak{g}$ and for all $\psi_i \in V_i$.

This means that if we choose an appropriate basis for V we can express $\Pi(g)$ as a block diagonalized matrix where the entries in the diagonal is the matrices for the $\Pi(g)_i$:s. If the $\Pi(g)_i$:s are reducible, we can continue to block diagonalize these matrices, but if they are irreducible this is not possible. Π can be expressed as a direct sum of irreducible representations, where the same representation may appear more than once,

$$\Pi = \Pi_1 \oplus \Pi_2 \oplus ... \oplus \Pi_m \tag{3.2.5}$$

with the corresponding vector spaces

$$V = V_1 \oplus V_2 \oplus ... \oplus V_m. \tag{3.2.6}$$

Proposition 3: A finite dimensional *completely reducible* representation is equivalent to a direct sum of *irreducible* representations.

Proof. First consider the case when the vector space V is irreducible. V is a direct sum with one term in the summation, namely V. The other case is when V is reducible. Then there exists at least two invariant subspaces W and U such that $V = W \oplus U$, $\dim(W)+\dim(U) = \dim(V)$, and that $W \cap U = \{0\}$. W and U must be completely reducible as well because every subspace to W, for example, is a subspace to V. Now we have two representations, not necessarily irreducible, for the Lie group or the Lie algebra acting on W and U. But if W and U are completely reducible then we can express, for example, W as a direct sum of two or more completely reducible representations and by induction $W = W_1 \oplus W_2 \oplus ... \oplus W_n$. Thus we can express $V = W_1 \oplus W_2 \oplus ... \oplus W_n \oplus U_1 \oplus U_2 \oplus ... \oplus U_m$ with the W_i:s and U_j:s irreducible. □

We have seen how the direct sum of vector spaces allows us to add representations acting on different vector spaces and to form new representations based on them. Another way to form new representations is to form the so called **tensor product**. The idea is to form a product of elements from two vector spaces V and W as $\boldsymbol{v} \otimes \boldsymbol{w}$, $\boldsymbol{v} \in V$ and $\boldsymbol{w} \in W$. What exactly do we mean by the "product", $\otimes$, of vectors?

Definition 3.2.3. *The* **tensor product** *V and W of the vector spaces, $V \otimes W$, is a vector space U together with a* bilinear *mapping $\Phi : V \otimes W \longrightarrow U$. In addition we have a* bilinear *mapping $\Psi : V \otimes W \longrightarrow X$, X is a vector space, such that there exists a* unique linear *mapping $\Psi' : U \longrightarrow X$.*

Observe that we have gone from a *bilinear* map from $V \otimes W$ to X to a *linear* operator from U to X. This is the main reason why we form the tensor product. An arbitrary vector of the form $\sum_{i,j} a_{i,j} \boldsymbol{v}_i \otimes \boldsymbol{w}_j$ can be expressed as a linear combination of the vectors $\boldsymbol{a}_i \otimes \boldsymbol{b}_j$, where $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$ form a basis for V and W, respectively. This can be seen from the fact that $\otimes$ is bilinear, so we can express the $\boldsymbol{v}_i$:s and $\boldsymbol{w}_j$:s in their respective bases. In other words, we can take $\boldsymbol{a}_i \otimes \boldsymbol{b}_j$, for all i,j, to be a basis for U and $\dim(U) = \dim(V) \times \dim(W)$.

Proposition 4: Let V and W be two finite dimensional vector spaces and $\hat{A}$ and $\hat{B}$ be two linear operators such that $\hat{A} : V \longrightarrow V$ and $\hat{B} : W \longrightarrow W$. Then we have a unique operator $\hat{A} \otimes \hat{B}$ acting on $V \otimes W$, $\hat{A} \otimes \hat{B} : V \otimes W \longrightarrow V \otimes W$, as

$$\hat{A} \otimes \hat{B}(\boldsymbol{v} \otimes \boldsymbol{w}) = (\hat{A}\boldsymbol{v}) \otimes (\hat{B}\boldsymbol{w}), \ \ \forall \boldsymbol{v} \in V, \forall \boldsymbol{w} \in W. \tag{3.2.7}$$

We will see later on how this works when we add angular momentum in section 4.5. The tensor product, as we will see, often arise in physics when we want to describe a system consisting of several particles as a unit, i.e. we want to work with the total system. A typical example is when we want to calculate the total energy of a system. Classically, we would add the constituent particle's energies but when it comes to quantum mechanics we have to find the eigenvalues of the total Hamiltonian, i.e. an operator that acts on the tensor product space. To get the total Hamiltonian we need the Hamiltonian of the individual particles and then take the tensor product of these operators. We know that a representation involves both a realization, e.g. matrices, and a vector space which our operators act on, so in order to find the representation of our total operator we need to define the tensor product of representations.

Definition 3.2.4. *Let Π_1 and Π_2 be two representation of $\mathcal{G}$ acting on V_1 and V_2, respectively. The* **tensor product** *$\Pi_1 \otimes \Pi_2$ is a representation of $\mathcal{G}$ acting on the vector space $V_1 \otimes V_2$ as*

$$\Pi_1(g) \otimes \Pi_2(g)(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2) = \Pi_1(g)\boldsymbol{v_1} \otimes \Pi_2(g)\boldsymbol{v}_2, \tag{3.2.8}$$

for all $\boldsymbol{v}_1 \in V_1, \boldsymbol{v}_2 \in V_2$ and for all $g \in \mathcal{G}$. Let $\mathfrak{g}$ be a Lie algebra with representations π_1 and π_2 acting on V_1 and V_2, respectively. The **tensor product** *$\pi_1 \otimes \pi_2$ is then a representation acting on $V_1 \otimes V_2$ as*

$$\pi_1(t) \otimes \pi_2(t)(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2) = \pi_1(t)\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_1 \otimes \pi_2(t)\boldsymbol{v}_2. \tag{3.2.9}$$

for all $\boldsymbol{v}_1 \in V_1, \boldsymbol{v}_2 \in V_2$ and for all $t \in \mathfrak{g}$.

As we saw in Proposition 1, there is a correspondence between the representations of a Lie group and its Lie algebra. Worth noting is that the expression (3.2.9) is not the same as (3.2.8). We motivate this with the following Proposition.

Proposition 5: Let Π_1, Π_2 be representations of the Lie group $\mathcal{G}$, acting on V_1,V_2, and π_1, π_2 be representations of the corresponding Lie algebra $\mathfrak{g}$, acting on V_1,V_2. $\pi_1(t) \otimes \pi_2(t)$ then acts on $V_1 \otimes V_2$ as

$$\pi_1(t) \otimes \pi_2(t) = \pi_1(t) \otimes \mathbb{1} + \mathbb{1} \otimes \pi_2(t) \tag{3.2.10}$$

for all $t \in \mathfrak{g}$.

Proof. Let us first consider the representations Π_1, Π_2 of the group. We have

$$\Pi_1(t) \otimes \Pi_2(t)(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2) = \Pi_1(e^{\lambda t}) \otimes \Pi_2(e^{\lambda t})(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2) = \Pi_1(e^{\lambda t})\boldsymbol{v}_1 \otimes \Pi_2(e^{\lambda t})\boldsymbol{v}_2. \quad (3.2.11)$$

Consider $\Pi_1(e^{\lambda t})\boldsymbol{v}_1$ and $\Pi_2(e^{\lambda t})\boldsymbol{v}_2$ as two smooth curves in V_1 and V_2, respectively, and denote them as $V_1(\lambda)$, $V_2(\lambda)$. λ is the curve parameter. The derivative of $V_1(\lambda) \otimes V_2(\lambda)$ then becomes

$$\begin{aligned} &\lim_{\Delta\lambda \to 0} \frac{V_1(\lambda + \Delta\lambda) \otimes V_2(\lambda + \Delta\lambda) - V_1(\lambda) \otimes V_2(\lambda)}{\Delta\lambda} = \\ &= \lim_{\Delta\lambda \to 0} \frac{V_1(\lambda + \Delta\lambda) \otimes V_2(\lambda + \Delta\lambda) - V_1(\lambda) \otimes V_2(\lambda + \Delta\lambda)}{\Delta\lambda} + \\ &\frac{V_1(\lambda) \otimes V_2(\lambda + \Delta\lambda) - V_1(\lambda) \otimes V_2(\lambda)}{\Delta\lambda} = \frac{dV_1(\lambda)}{d\lambda} \otimes V_2(\lambda) + V_1(\lambda) \otimes \frac{dV_2(\lambda)}{\lambda}, \end{aligned} \quad (3.2.12)$$

i.e. the usual product rule. To see how the representation $\pi_1 \otimes \pi_2$ becomes we have to differentiate (3.2.11) with respect to λ at the identity, $\lambda = 0$.

$$\begin{aligned} &\frac{d}{d\lambda}\left(\Pi_1(e^{\lambda t})\boldsymbol{v}_1 \otimes \Pi_2(e^{\lambda t})\boldsymbol{v}_2\right)\Big|_{\lambda=0} = \frac{d}{d\lambda}\Pi_1(e^{\lambda t})\Big|_{\lambda=0} \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_1 \otimes \frac{d}{d\lambda}\Pi_2(e^{\lambda t})\Big|_{\lambda=0} \boldsymbol{v}_2 = \\ &= \pi_1(t)\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_1 \otimes \pi_2(t)\boldsymbol{v}_2 = \pi_1(t) \otimes \mathbb{1} + \mathbb{1} \otimes \pi_2(t)(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2). \end{aligned} \quad (3.2.13)$$

Since $\boldsymbol{v}_1 \otimes \boldsymbol{v}_2$ is arbitrary (3.2.10) is true on $V_1 \otimes V_2$. □

At the end of chapter 4 we will see a concrete example of the tensor product when we learn how the tensor product of two irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$ can be decomposed into a direct sum of irreducible representations of the same algebra, and how this is applied in the addition of angular momenta in quantum mechanics.
Although there is much more to say and learn about representation theory, we will now leave this chapter. The basic concepts have been presented and with these we are ready to handle the physical interpretations of representations. There will be much more said about represein the forthcoming chapters, see for example section 6.5 and 4.4, that will show how to construct representations. The most important concept of this chapter is undoubtedly irreducible representations and as we shall see, have a very important physical meaning.

Chapter 4

A Physical Approach

In the previous chapters we studied the basic mathematical formalism of group theory and representation theory. The purpose of the present chapter is to provide examples of how these concepts arise naturally in the field of quantum mechanics. More specifically, we will first derive the *fundamental commutation relations of angular momentum* in quantum mechanics. These relations will reveal the connection between the Lie group $SU(2)$ and the set of rotation operators acting on the ket space of any given physical system. A concrete example of how the fundamental commutation relations may be realized will then be given through the study of spin 1/2 systems.

By the end of this chapter we will also have studied ladder operators, irreducible representations and the addition of angular momenta. But before we can advance that far we must take a closer look at rotations in $\mathbb{R}^3$.

4.1 Rotations in $\mathbb{R}^3$

Rotations about the same axis in $\mathbb{R}^3$ always commute with one another while *finite* rotations about different axes generally do not. Let $R_l(\phi)$ denote a rotation by the angle ϕ about an axis l passing through the origin of a Cartesian coordinate system xyz. Then an arbitrary rotation $R_l(\phi_1)$ followed by a rotation $R_l(\phi_2)$ about the same axis will always give the same result as a rotation $R_l(\phi_2)$ followed by a rotation $R_l(\phi_1)$, while, for instance, a rotation $R_x(\pi/2)$ followed by a rotation $R_z(\pi/2)$ will obviously give a different result than that of the same rotations performed in the reversed order.

Even though finite rotations generally fail to commute, *infinitesimal* rotations about any two axes in $\mathbb{R}^3$ always commute in a first order approximation. However, they *do not* commute in a second order approximation. This is a very important fact which will lead us almost directly to the fundamental commutation relations of angular momentum. As a first step we must therefore prove the statements just made.

Rotations are linear transformations and for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{x} \mapsto T(\boldsymbol{x})$ there exists a uniquely determined **standard matrix** A such that

$$T(\boldsymbol{x}) = A\boldsymbol{x}, \, \forall \boldsymbol{x} \in \mathbb{R}^n. \tag{4.1.1}$$

It can easily be proven that A is a real $m \times n$ matrix given by

$$A = [T(\boldsymbol{e}_1) \dots T(\boldsymbol{e}_n)], \tag{4.1.2}$$

where $\boldsymbol{e}_j$ $(j = 1, \ldots, n)$ is the jth column of the identity matrix $\mathbb{1}$ in $\mathbb{R}^n$ ([12], p. 99).

Hence, rotation matrices acting on $\mathbb{R}^3$ are 3×3 matrices with real entries. In fact, they are *orthogonal* matrices with determinant one (i.e. proper orthogonal matrices). One can also prove that the converse is true, namely that any real 3×3 orthogonal matrix with determinant one is a rotation matrix[13]. The set of all rotation matrices acting on $\mathbb{R}^3$ therefore constitutes the defining representation of the Lie group $SO(3)$. The general non-commutative nature of finite rotations in $\mathbb{R}^3$ is an expression of the fact that $SO(3)$ is a non-Abelian group.

For simplicity's sake we will denote the matrix for a rotation $R_l(\phi)$ in the same way as the rotation itself.

According to equation (4.1.2) the matrix for a rotation about the z-axis by the angle ϕ is given by

$$R_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4.1.3}$$

Note that this is the 3×3 analogy of the 2×2 rotation matrix in equation 2.3.1.

By replacing the trigonometric functions in equation (4.1.3) with their respective Taylor polynomials of order two and letting $\phi = \epsilon \to 0$ we obtain the second order approximation

$$R_z(\epsilon) = \begin{pmatrix} 1-\epsilon^2/2 & -\epsilon & 0 \\ \epsilon & 1-\epsilon^2/2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.1.4}$$

of the matrix for a rotation about the z-axis by the *infinitesimal* angle ϵ.

Through cyclic permutations of (x, y, z) equation (4.1.4) gives the second order approximations

$$R_x(\epsilon) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\epsilon^2/2 & -\epsilon \\ 0 & \epsilon & 1-\epsilon^2/2 \end{pmatrix} \tag{4.1.5}$$

and

$$R_y(\epsilon) = \begin{pmatrix} 1-\epsilon^2/2 & 0 & \epsilon \\ 0 & 1 & 0 \\ -\epsilon & 0 & 1-\epsilon^2/2 \end{pmatrix} \tag{4.1.6}$$

of the matrices for rotations about the x- and y-axis, respectively, by the infinitesimal angle ϵ.

Let us now consider an infinitesimal rotation about the x-axis by the angle ϵ_1 followed by another infinitesimal rotation about the y-axis by the angle ϵ_2. With the help of equation (4.1.5) and (4.1.6) we find that the second order approximation of the matrix for the total

rotation is given by

$$\begin{aligned} R_y(\epsilon_2)R_x(\epsilon_1) &= \begin{pmatrix} 1-\epsilon_2^2/2 & 0 & \epsilon_2 \\ 0 & 1 & 0 \\ -\epsilon_2 & 0 & 1-\epsilon_2^2/2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\epsilon_1^2/2 & -\epsilon_1 \\ 0 & \epsilon_1 & 1-\epsilon_1^2/2 \end{pmatrix} \\ &= \begin{pmatrix} 1-\epsilon_2^2/2 & \epsilon_1\epsilon_2 & (1-\epsilon_1^2/2)\epsilon_2 \\ 0 & 1-\epsilon_1^2/2 & -\epsilon_1 \\ -\epsilon_2 & (1-\epsilon_2^2/2)\epsilon_1 & (1-\epsilon_1^2/2)(1-\epsilon_2^2/2) \end{pmatrix} \\ &\approx \begin{pmatrix} 1-\epsilon_2^2/2 & \epsilon_1\epsilon_2 & \epsilon_2 \\ 0 & 1-\epsilon_1^2/2 & -\epsilon_1 \\ -\epsilon_2 & \epsilon_1 & 1-\epsilon_1^2/2-\epsilon_2^2/2 \end{pmatrix}. \end{aligned} \tag{4.1.7}$$

If we reverse the order in which the rotations are performed we instead obtain the matrix

$$\begin{aligned} R_x(\epsilon_1)R_y(\epsilon_2) &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\epsilon_1^2/2 & -\epsilon_1 \\ 0 & \epsilon_1 & 1-\epsilon_1^2/2 \end{pmatrix} \begin{pmatrix} 1-\epsilon_2^2/2 & 0 & \epsilon_2 \\ 0 & 1 & 0 \\ -\epsilon_2 & 0 & 1-\epsilon_2^2/2 \end{pmatrix} \\ &= \begin{pmatrix} 1-\epsilon_2^2/2 & 0 & \epsilon_2 \\ \epsilon_1\epsilon_2 & 1-\epsilon_1^2/2 & (1-\epsilon_2^2/2)(-\epsilon_1) \\ (1-\epsilon_1^2/2)(-\epsilon_2) & \epsilon_1 & (1-\epsilon_1^2/2)(1-\epsilon_2^2/2) \end{pmatrix} \\ &\approx \begin{pmatrix} 1-\epsilon_2^2/2 & 0 & \epsilon_2 \\ \epsilon_1\epsilon_2 & 1-\epsilon_1^2/2 & -\epsilon_1 \\ -\epsilon_2 & \epsilon_1 & 1-\epsilon_1^2/2-\epsilon_2^2/2 \end{pmatrix}. \end{aligned} \tag{4.1.8}$$

The matrices (4.1.7) and (4.1.8) are obviously not equal to each other, but they *are* equal if we set $\epsilon_1^2 = \epsilon_2^2 = \epsilon_1\epsilon_2 = 0$. This proves that rotations about the x- and y-axis always commute in a first order approximation, but generally not in a second order approximation (the exceptions being when $\epsilon_1 = 0$ and/or $\epsilon_2 = 0$). We will now show that this is true for *any* two axes in $\mathbb{R}^3$.

From a symmetry point of view it seems clear that the result above must also apply to rotations about the x- and z-axis or the y- and z-axis. This means that infinitesimal rotations about any two Cartesian coordinate axes always commute in a first but not in a second order approximation. Moreover, it can be shown that any rotation in $\mathbb{R}^3$ (finite or infinitesimal) can be carried out by performing three successive rotations about two arbitrary Cartesian coordinate axes ([14], p. 177). Such rotations are in fact a variation of the perhaps more familiar *Euler rotations*. If we for instance select the y- and z-axis as our rotation axes then for any given axis l and infinitesimal angle ϵ

$$R_l(\epsilon) = R_z(\epsilon_1)R_y(\epsilon_2)R_z(\epsilon_3) \tag{4.1.9}$$

for some values of $\epsilon_1, \epsilon_2, \epsilon_3 \in \mathbb{R}$.

Since infinitesimal rotations about Cartesian coordinate axes commute in a first order approximation it therefore follows that the same can be said about infinitesimal rotations about two arbitrary axes l and l' since for any infinitesimal angles $\epsilon, \epsilon' \in \mathbb{R}$ equation (4.1.9)

gives that

$$\begin{aligned} R_l(\epsilon)R_{l'}(\epsilon') &= [R_z(\epsilon_1)R_y(\epsilon_2)R_z(\epsilon_3)][R_{z'}(\epsilon_{1'})R_{y'}(\epsilon_{2'})R_{z'}(\epsilon_{3'})] \\ &= R_z(\epsilon_1)R_y(\epsilon_2)R_z(\epsilon_3)R_{z'}(\epsilon_{1'})R_{y'}(\epsilon_{2'})R_{z'}(\epsilon_{3'}) \\ &= R_z(\epsilon_1)R_y(\epsilon_2)R_{z'}(\epsilon_{1'})R_z(\epsilon_3)R_{y'}(\epsilon_{2'})R_{z'}(\epsilon_{3'}) \\ &\dots \\ &= R_{z'}(\epsilon_{1'})R_{y'}(\epsilon_{2'})R_{z'}(\epsilon_{3'})R_z(\epsilon_1)R_y(\epsilon_2)R_z(\epsilon_3) \\ &= [R_{z'}(\epsilon_{1'})R_{y'}(\epsilon_{2'})R_{z'}(\epsilon_{3'})][R_z(\epsilon_1)R_y(\epsilon_2)R_z(\epsilon_3)] \\ &= R_{l'}(\epsilon')R_l(\epsilon) \end{aligned} \tag{4.1.10}$$

for some infinitesimal values of $\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_{1'}, \epsilon_{2'}, \epsilon_{3'} \in \mathbb{R}$.

By closely examining equation (4.1.10) we can also see that infinitesimal rotations about arbitrary axes do not commute in a second order approximation since in general $R_z(\epsilon_3)R_{z'}(\epsilon_{1'}) \neq R_{z'}(\epsilon_{1'})R_z(\epsilon_3)$, $R_y(\epsilon_2)R_{z'}(\epsilon_{1'}) \neq R_{z'}(\epsilon_{1'})R_y(\epsilon_2)$ etc. in a second order approximation. We have thus proven our initial assertions.

Moreover, if we let $\epsilon_1 = \epsilon_2 = \epsilon$ then equation (4.1.4), (4.1.7) and (4.1.8) give the commutation relation

$$[R_x(\epsilon),\, R_y(\epsilon)] = \begin{pmatrix} 0 & -\epsilon^2 & 0 \\ \epsilon^2 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = R_z(\epsilon^2) - \mathbb{1}, \tag{4.1.11}$$

where we have ignored terms of order three and higher in ϵ.

Equation (4.1.11) is the sought after relation which will help us to derive the fundamental commutation relations of angular momentum in quantum mechanics in the next section. Through cyclic permutations of (x, y, z) we can also obtain the similar commutation relations

$$[R_z(\epsilon),\, R_x(\epsilon)] = R_y(\epsilon^2) - \mathbb{1} \tag{4.1.12}$$

$$[R_y(\epsilon),\, R_z(\epsilon)] = R_x(\epsilon^2) - \mathbb{1}. \tag{4.1.13}$$

In general, we have that

$$[R_i(\epsilon),\, R_j(\epsilon)] = \epsilon_{ijk}[R_k(\epsilon^2) - \mathbb{1}], \tag{4.1.14}$$

where $i, j, k \in \{1, 2, 3\}$, $R_1 = R_x, R_2 = R_y$ and $R_3 = R_z$.

4.2 Rotations in Quantum Mechanics

A rotation of a given physical system generally affects the probability distributions for the system's observables, such as its linear momentum and spin. We therefore expect the state ket of the system to be altered by the rotation (for a discussion about kets and ket spaces see appendix C).

Let us find the operators for rotations acting on the system's ket space $\mathcal{H}$. For every rotation matrix $R_l(\phi)$ acting on $\mathbb{R}^3$, where l is given by the unit direction vector $\hat{\boldsymbol{n}}$, we associate an operator $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ (D as in *Drehung*, which is the German word for *rotation*[1]) acting on $\mathcal{H}$ such that

$$|\psi\rangle_R = \hat{D}(\hat{\boldsymbol{n}}, \phi)|\psi\rangle, \tag{4.2.1}$$

[1]For historical reasons, many words in quantum mechanics are of German origin.

where $|\psi\rangle$ is the state ket before the rotation and $|\psi\rangle_R$ is the state ket after the rotation.

In order to derive an explicit expression for $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ we will first find the operator $\hat{D}(\hat{\boldsymbol{n}}, \mathrm{d}\phi)$ for an infinitesimal rotation by the angle $\mathrm{d}\phi$ about the same axis. Once we have found the latter operator we may compute $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ as $\lim_{N\to\infty}[\hat{D}(\hat{\boldsymbol{n}}, \phi/N)]^N$, just as we did for rotation matrices acting on $\mathbb{R}^2$ in section 2.3.

To obtain an *idea* about what $\hat{D}(\hat{\boldsymbol{n}}, \mathrm{d}\phi)$ might look like we will first examine some familiar operators for infinitesimal symmetry transformations in quantum mechanics. We know for instance that the operator $\hat{\mathscr{J}}(\mathrm{d}\boldsymbol{x})$ for an infinitesimal translation $\mathrm{d}\boldsymbol{x}$ in $\mathbb{R}^3$ is given by

$$\hat{\mathscr{J}}(\mathrm{d}\boldsymbol{x}) = \hat{\mathbb{1}} - i\frac{\hat{\boldsymbol{p}}\cdot \mathrm{d}\boldsymbol{x}}{\hbar}, \tag{4.2.2}$$

where $\hat{\mathbb{1}}$ is the identity operator and $\hat{\boldsymbol{p}}$ is the linear momentum operator ([14], p. 46).

We also know that the operator $\hat{\mathscr{U}}(t_0 + \mathrm{d}t, t_0)$ for an infinitesimal time translation $\mathrm{d}t$ at time $t = t_0$ is given by

$$\hat{\mathscr{U}}(t_0 + \mathrm{d}t, t_0) = \hat{\mathbb{1}} - i\frac{\hat{H}\,\mathrm{d}t}{\hbar}, \tag{4.2.3}$$

where $\hat{H}$ is the Hamiltonian operator ([14], p. 69).

In classical mechanics angular momentum is the generator[2] of rotations in the same way as linear momentum and the Hamiltonian are the generators of translations in space and time, respectively. By analogy it is therefore reasonable to assume that

$$\hat{D}(\hat{\boldsymbol{n}}, \mathrm{d}\phi) = \hat{\mathbb{1}} - i\frac{\hat{\boldsymbol{J}}\cdot\hat{\boldsymbol{n}}\,\mathrm{d}\phi}{\hbar}, \tag{4.2.4}$$

where $\hat{\boldsymbol{J}}$ is the angular momentum operator.

Unfortunately, we cannot prove that (4.2.4) is the correct expression for $\hat{D}(\hat{\boldsymbol{n}}, \mathrm{d}\phi)$. The reason for this is that in classical mechanics the angular momentum $\boldsymbol{J}$ of a particle with respect to a point P is defined as $\boldsymbol{J} = \boldsymbol{x}\times\boldsymbol{p}$, where $\boldsymbol{x}$ is the particle's position vector relative to P and $\boldsymbol{p}$ is the its linear momentum. With this definition of angular momentum one can prove that angular momentum is the generator of rotations in classical mechanics (see chapter 9 in [15]). However, such a proof does not hold in quantum mechanics since the angular momentum associated with the particle's spin has nothing to do with the particle's position or linear momentum.

Let us therefore *define* $\hat{\boldsymbol{J}}$ as the operator which gives $\hat{D}(\hat{\boldsymbol{n}}, \mathrm{d}\phi)$ the form in equation (4.2.4). Moreover, we postulate that the set of all rotation operators $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ acting on $\mathcal{H}$, together with the operation $\hat{D}(\hat{\boldsymbol{n}}_1, \phi_1)\star\hat{D}(\hat{\boldsymbol{n}}_2, \phi_2) = \hat{D}(\hat{\boldsymbol{n}}_1, \phi_1)\hat{D}(\hat{\boldsymbol{n}}_2, \phi_2)$, forms a group and that the mapping $R_l(\phi) \mapsto \hat{D}(\hat{\boldsymbol{n}}, \phi)$ is a group homomorphism ([14], p. 162).

As we mentioned above, we can now obtain an explicit expression for $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ by perform-

[2]Note that in physics the word *generator* is quite ambiguous. In this context we are neither referring to the generators of groups nor to the generators of Lie algebras. For an excellent disposition of generators in classical mechanics see chapter 9 in [15].

ing a succession of increasingly smaller rotations about the same axis:

$$\begin{aligned}\hat{D}(\hat{\boldsymbol{n}},\phi) &= \lim_{N\to\infty}\left[\hat{\mathbb{1}} - i\frac{\hat{\boldsymbol{J}}\cdot\hat{\boldsymbol{n}}}{\hbar}\frac{\phi}{N}\right]^N \\ &= \exp\left(-i\frac{\hat{\boldsymbol{J}}\cdot\hat{\boldsymbol{n}}}{\hbar}\phi\right) \qquad (4.2.5)\\ &= \hat{\mathbb{1}} - i\frac{\hat{\boldsymbol{J}}\cdot\hat{\boldsymbol{n}}}{\hbar}\phi - \frac{(\hat{\boldsymbol{J}}\cdot\hat{\boldsymbol{n}})^2}{2\hbar^2}\phi^2 + \dots \qquad (4.2.6)\end{aligned}$$

Let us now at last derive the fundamental commutation relations of angular momentum. We begin by noticing that the quantum mechanical analogy of equation (4.1.11) is

$$\hat{D}(\hat{\boldsymbol{x}},\epsilon)\hat{D}(\hat{\boldsymbol{y}},\epsilon) - \hat{D}(\hat{\boldsymbol{y}},\epsilon)\hat{D}(\hat{\boldsymbol{x}},\epsilon) = \hat{D}(\hat{\boldsymbol{z}},\epsilon^2) - \hat{\mathbb{1}}, \qquad (4.2.7)$$

where we have used the fact that $R_l(\phi) \mapsto \hat{D}(\hat{\boldsymbol{n}},\phi)$ is a group homomorphism.

By replacing the operators in (4.2.7) with their second order approximations, which are given by equation (4.2.6), we find that

$$\begin{aligned}&\left(\hat{\mathbb{1}} - i\frac{\hat{J}_x\epsilon}{\hbar} - \frac{\hat{J}_x^2\epsilon^2}{2\hbar^2}\right)\left(\hat{\mathbb{1}} - i\frac{\hat{J}_y\epsilon}{\hbar} - \frac{\hat{J}_y^2\epsilon^2}{2\hbar^2}\right)\\ &\qquad - \left(\hat{\mathbb{1}} - i\frac{\hat{J}_y\epsilon}{\hbar} - \frac{\hat{J}_y^2\epsilon^2}{2\hbar^2}\right)\left(\hat{\mathbb{1}} - i\frac{\hat{J}_x\epsilon}{\hbar} - \frac{\hat{J}_x^2\epsilon^2}{2\hbar^2}\right)\\ &\qquad = \hat{\mathbb{1}} - i\frac{\hat{J}_z\epsilon^2}{\hbar} - \hat{\mathbb{1}}\\ &\qquad = i\frac{\hat{J}_z\epsilon^2}{\hbar}, \qquad (4.2.8)\end{aligned}$$

where $\hat{J}_x, \hat{J}_y$ and $\hat{J}_z$ are the operators for the Cartesian components of the angular momentum $\boldsymbol{J}$.

If we simplify equation (4.2.8) and omit terms of order three and higher in ϵ then we obtain the commutation relation

$$[\hat{J}_x, \hat{J}_y] = i\hbar\hat{J}_z. \qquad (4.2.9)$$

By permuting (x, y, z) we finally obtain the **fundamental commutation relations of angular momentum**:

$$[\hat{J}_i, \hat{J}_j] = i\hbar\epsilon_{ijk}\hat{J}_k, \qquad (4.2.10)$$

where $\hat{J}_1 = \hat{J}_x, \hat{J}_2 = \hat{J}_x$ and $\hat{J}_3 = \hat{J}_z$.

Let us now find the promised connection between the rotation operators and the group $SU(2)$. We proved in section 2.3 that the generators of the defining representation $\pi(\mathfrak{su}(2))$ of the Lie algebra $\mathfrak{su}(2)$ are $i\sigma_1, i\sigma_2$ and $i\sigma_3$. Since $\pi(\mathfrak{su}(2))$ is a vector space over $\mathbb{R}$ this means that every element X in $\pi(\mathfrak{su}(2))$ can be expressed on the form

$$X = a_1(i\sigma_1) + a_2(i\sigma_2) + a_3(i\sigma_3) = ia_1\sigma_1 + ia_2\sigma_2 + ia_3\sigma_3,\ a_1, a_2, a_3 \in \mathbb{R}. \qquad (4.2.11)$$

Note that if we were to take linear combinations of $i\sigma_1, i\sigma_2$ and $i\sigma_3$ with complex coefficients then we would leave $\pi(\mathfrak{su}(2))$ and enter the the defining representation of the complexification $\mathfrak{sl}(2,\mathbb{C})$ of $\mathfrak{su}(2)$ (see equation 2.3.32 in section 2.3).

Moreover, we mentioned in section 2.3 that the commutation relations for the generators of $\pi(\mathfrak{su}(2))$ are

$$[i\sigma_i, i\sigma_j] = -2i\epsilon_{ijk}\sigma_k, \tag{4.2.12}$$

and that by exponentiating elements of $\pi(\mathfrak{su}(2))$ we obtain elements of the defining representation $\Pi(SU(2))$ of $SU(2)$:

$$\exp(ia_1\sigma_1 + ia_2\sigma_2 + ia_3\sigma_3) \in \Pi(SU(2)),\ \forall a_1, a_2, a_3 \in \mathbb{R}. \tag{4.2.13}$$

If we define the operators $\hat{J}'_1, \hat{J}'_2$ and $\hat{J}'_3$ according to

$$\hat{J}'_i \overset{def}{=} \frac{2}{\hbar}\hat{J}_i,\ i = 1, 2, 3, \tag{4.2.14}$$

then we can rewrite the commutation relations (4.2.10) as

$$[i\hat{J}'_i, i\hat{J}'_j] = -2i\epsilon_{ijk}\hat{J}'_k, \tag{4.2.15}$$

which is on the same form as (4.2.12).

This proves that $i\hat{J}'_1, i\hat{J}'_2$ and $i\hat{J}'_3$ are the generators of a representation of $\mathfrak{su}(2)$[3]. Naturally, the same can be said about $i\hat{J}_1, i\hat{J}_2$ and $i\hat{J}_3$ since they are real multiples of $i\hat{J}'_1, i\hat{J}'_2$ and $i\hat{J}'_3$, respectively.

Finally, we see in equation (4.2.5) that the rotation operators $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ are obtained through exponentiation of *real* multiples of

$$i\hat{\boldsymbol{J}} \cdot \hat{\boldsymbol{n}} = in_1\hat{J}_1 + in_2\hat{J}_2 + in_3\hat{J}_3, \tag{4.2.16}$$

where $\hat{\boldsymbol{n}} = (n_1, n_2, n_3)$.

In other words, the rotation operators are obtained through exponentiation of *real* linear combinations of the generators $i\hat{J}_1, i\hat{J}_2$ and $i\hat{J}_3$ of a representation of $\mathfrak{su}(2)$. Moreover, $SU(2)$ is an exponential group, which means that the exponential map covers all of $SU(2)$.

We have thus proven the following statement:

> The rotation operators acting on the ket space of a physical system, together with the ket space itself, form a representation of the group $SU(2)$.

In the next section we will exemplify the results above through the study of spin 1/2 systems.

4.3 Spin 1/2 Systems

Let n be the dimension of the ket space of a physical system. The lowest value of n for which the commutation relations (4.2.10) can be realized is $n = 2$ ([14], p. 163). This just happens to be the dimension of the ket space of a spin 1/2 system with no other degrees of freedom. For the sake of simplicity, let us therefore have a look at such a system.

Let $\hat{S}_x, \hat{S}_y$ and $\hat{S}_z$ denote the operators for the components of the spin in the x-, y- and z-direction, respectively. Moreover, let $|\uparrow\rangle$ and $|\downarrow\rangle$ be normalized eigenkets of $\hat{S}_z$ with the corresponding eigenvalues $\pm\hbar/2$, where $\uparrow$ and $\downarrow$ refer to spin up and spin down.

[3] At least from a mathematicians point of view. As we mentioned in section 2.3 a physicist considers $\{\hat{J}'_i\}$, rather than $\{i\hat{J}_i\}$, to be the generators of a representation of $\mathfrak{su}(2)$.

These eigenkets form a basis for the ket space $\mathcal{H}$ and it can be shown ([14], p. 163) that the spin operators can be expressed as

$$\hat{S}_x = \frac{\hbar}{2}\Big(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big) \tag{4.3.1}$$

$$\hat{S}_y = i\frac{\hbar}{2}\Big(-|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big) \tag{4.3.2}$$

$$\hat{S}_z = \frac{\hbar}{2}\Big(|\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|\Big), \tag{4.3.3}$$

where $|\uparrow\rangle\langle\downarrow|$ denotes the *outer product* of $|\uparrow\rangle$ and $\langle\downarrow|$ (a definition of the outer product of a ket and a bra is given in appendix C).

Since spin is an intrinsic *angular momentum* the spin operators satisfy the fundamental commutation relations of angular momentum. We can prove this explicitly by inserting the expressions (4.3.1)–(4.3.3) into equation (4.2.10). For instance:

$$\begin{aligned}
[\hat{S}_x, \hat{S}_y] &= \hat{S}_x\hat{S}_y - \hat{S}_y\hat{S}_x \\
&= \left[\frac{\hbar}{2}\Big(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big)\right]\left[i\frac{\hbar}{2}\Big(-|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big)\right] \\
&\quad - \left[i\frac{\hbar}{2}\Big(-|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big)\right]\left[\frac{\hbar}{2}\Big(|\uparrow\rangle\langle\downarrow| + |\downarrow\rangle\langle\uparrow|\Big)\right] \\
&= i\frac{\hbar^2}{2}\Big(|\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow|\Big) \\
&= i\hbar\hat{S}_z,
\end{aligned} \tag{4.3.4}$$

where we have used the orthonormality of $|\uparrow\rangle$ and $|\downarrow\rangle$.

Moreover, the operator $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ for a finite rotation by the angle ϕ about the axis l through the origin given by the unit direction vector $\hat{\boldsymbol{n}}$ is, according to equation (4.2.5), given by

$$\hat{D}(\hat{\boldsymbol{n}}, \phi) = \exp\left(-i\frac{\hat{\boldsymbol{S}}\cdot\hat{\boldsymbol{n}}}{\hbar}\phi\right). \tag{4.3.5}$$

It is interesting to observe the effect of $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ on an arbitrary ket in $\mathcal{H}$. Any ket $|\psi\rangle$ in $\mathcal{H}$ can be expressed as a linear combination of the base kets:

$$|\psi\rangle = |\uparrow\rangle\langle\uparrow|\psi\rangle + |\downarrow\rangle\langle\downarrow|\psi\rangle. \tag{4.3.6}$$

If we for instance let $\hat{D}(\hat{\boldsymbol{z}}, \phi)$ act on $|\psi\rangle$ then the result will be

$$\begin{aligned}
\hat{D}(\hat{\boldsymbol{z}}, \phi)|\psi\rangle &= \exp\left(-i\frac{\hat{S}_z\phi}{\hbar}\right)|\psi\rangle \\
&= \exp\left(-i\frac{\hat{S}_z\phi}{\hbar}\right)\Big(|\uparrow\rangle\langle\uparrow|\psi\rangle + |\downarrow\rangle\langle\downarrow|\psi\rangle\Big) \\
&= \left(\hat{\mathbb{1}} - i\frac{\hat{S}_z\phi}{\hbar} - \dots\right)\Big(|\uparrow\rangle\langle\uparrow|\psi\rangle + |\downarrow\rangle\langle\downarrow|\psi\rangle\Big) \\
&\dots \\
&= e^{-i\phi/2}|\uparrow\rangle\langle\uparrow|\psi\rangle + e^{i\phi/2}|\downarrow\rangle\langle\downarrow|\psi\rangle,
\end{aligned} \tag{4.3.7}$$

where we have used the fact that $|\uparrow\rangle$ and $|\downarrow\rangle$ are eigenkets of the operator $\hat{S}_z$ corresponding to the eigenvalues $\pm\hbar/2$.

Note that *any* axis l in $\mathbb{R}^3$ can be chosen as the z-axis of a Cartesian coordinate system. The physical significance of the half-angle $\phi/2$ in equation (4.3.7) is therefore that a spin 1/2 system has to be rotated by the angle 4π, rather than 2π, about any given axis before the state ket returns to its original value. Hence, for every rotation matrix $R_l(\phi)$ acting on $\mathbb{R}^3$ there are *two* corresponding rotation operators acting on $\mathcal{H}$, namely $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ and $\hat{D}(\hat{\boldsymbol{n}}, \phi+2\pi)$. This is an expression of the fact that for every element in the underlying set of $SO(3)$ there are two corresponding elements in the underlying set of $SU(2)$. We therefore say that $SU(2)$ is the *double cover* of $SO(3)$ (this subject is discussed further in section 6.7.2.).

Let us now introduce some new terminology that will facilitate calculations involving state kets of spin 1/2 systems.

4.3.1 Spinors

The **two-component spinor** associated with a given ket in the ket space $\mathcal{H}$ of a spin 1/2 system is the coordinate column vector of the ket with respect to the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$ for $\mathcal{H}$ (spinors are discussed further in example 6.5.2 in section 6.5). With the help of equation (4.3.6) we therefore find that the spinor corresponding to an arbitrary ket $|\psi\rangle$ in $\mathcal{H}$ is given by

$$|\psi\rangle = |\uparrow\rangle\langle\uparrow|\psi\rangle + |\downarrow\rangle\langle\downarrow|\psi\rangle \mapsto \begin{pmatrix} \langle\uparrow|\psi\rangle \\ \langle\downarrow|\psi\rangle \end{pmatrix}. \tag{4.3.8}$$

Moreover, every bra $\langle\psi|$ in the bra space $\mathcal{H}^*$ (the star denotes the dual space to $\mathcal{H}$ – see appendix C and D for an explanation) is associated with the Hermitian adjoint of the spinor corresponding to the ket $|\psi\rangle$:

$$\langle\psi| = \langle\psi|\uparrow\rangle\langle\uparrow| + \langle\psi|\downarrow\rangle\langle\downarrow| \mapsto \begin{pmatrix} \langle\uparrow|\psi\rangle \\ \langle\downarrow|\psi\rangle \end{pmatrix}^\dagger = \Big(\langle\psi|\uparrow\rangle, \langle\psi|\downarrow\rangle\Big). \tag{4.3.9}$$

Let us now calculate the matrix elements for the matrices of the spin operators $\hat{S}_x, \hat{S}_y$ and $\hat{S}_z$ with respect to the basis $\{|\uparrow\rangle, |\downarrow\rangle\}$. With the help of equation (4.3.1)–(4.3.3) we find that

$$\langle\uparrow|\hat{S}_x|\uparrow\rangle = 0, \quad \langle\uparrow|\hat{S}_x|\downarrow\rangle = \hbar/2, \quad \langle\downarrow|\hat{S}_x|\uparrow\rangle = \hbar/2, \quad \langle\downarrow|\hat{S}_x|\downarrow\rangle = 0 \tag{4.3.10}$$

$$\langle\uparrow|\hat{S}_y|\uparrow\rangle = 0, \quad \langle\uparrow|\hat{S}_y|\downarrow\rangle = -i\hbar/2, \quad \langle\downarrow|\hat{S}_y|\uparrow\rangle = i\hbar/2, \quad \langle\downarrow|\hat{S}_y|\downarrow\rangle = 0 \tag{4.3.11}$$

$$\langle\uparrow|\hat{S}_z|\uparrow\rangle = \hbar/2, \quad \langle\uparrow|\hat{S}_z|\downarrow\rangle = 0, \quad \langle\downarrow|\hat{S}_z|\uparrow\rangle = 0, \quad \langle\downarrow|\hat{S}_z|\downarrow\rangle = -\hbar/2. \tag{4.3.12}$$

According to example 6.5.2 the matrices of the spin operators are therefore given by

$$S_x = \frac{\hbar}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar}{2}\sigma_1 \tag{4.3.13}$$

$$S_y = \frac{\hbar}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{\hbar}{2}\sigma_2 \tag{4.3.14}$$

$$S_z = \frac{\hbar}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar}{2}\sigma_3. \tag{4.3.15}$$

This shows in a very straightforward manner that $i\hat{S}_x, i\hat{S}_y$ and $i\hat{S}_z$ are the generators of a representation of $\mathfrak{su}(2)$ and that the rotation operators, along with the ket space on which they act, form a representation of $SU(2)$. More specifically, it shows that the matrices of $i\hat{S}_x, i\hat{S}_y$ and $i\hat{S}_z$ are the generators of the *defining* representation of $\mathfrak{su}(2)$. The matrices of the rotation operators, together with the spinor space $\mathbb{C}^2$ on which they act, therefore form the *defining* representation of $SU(2)$.

To show this explicitly we can calculate the matrix $D(\hat{\boldsymbol{n}}, \phi)$ of an arbitrary rotation operator $\hat{D}(\hat{\boldsymbol{n}}, \phi)$ with the help of equation (4.3.5) and (4.3.13)–(4.3.15):

$$D(\hat{\boldsymbol{n}}, \phi) = \exp\left(-i\frac{\boldsymbol{S}\cdot\hat{\boldsymbol{n}}}{\hbar}\phi\right) = \exp\left(-i\frac{\boldsymbol{\sigma}\cdot\hat{\boldsymbol{n}}}{2}\phi\right), \tag{4.3.16}$$

where $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the so-called **Pauli vector** (which is actually not a vector at all since its elements are matrices).

By expanding the exponential function in (4.3.16) we find that

$$\begin{aligned}&\exp\left(-i\frac{\boldsymbol{\sigma}\cdot\hat{\boldsymbol{n}}}{2}\phi\right)\\ &= \begin{pmatrix}\cos(\phi/2) - in_z\sin(\phi/2) & (-in_x - n_y)\sin(\phi/2)\\ (-in_x + n_y)\sin(\phi/2) & \cos(\phi/2) + in_z\sin(\phi/2)\end{pmatrix}.\end{aligned} \tag{4.3.17}$$

It can readily be shown that the matrices given by (4.3.17) are unitary matrices with determinant one, i.e. elements of the underlying set of the defining representation of $SU(2)$.

4.4 Construction of Irreducible Representations

In the preceding section we defined the $|\uparrow\rangle$ and $|\downarrow\rangle$ to be eigenkets to $\hat{S}_z$ and used them as a basis for the 2-dimensional ket space for spin 1/2. We are now ready to develop a general method to find a basis for a ket space of any spin, constituted by eigenkets to one of our angular momentum operators.

We start by, in addition to our three operators $\hat{J}_x$, $\hat{J}_y$ and $\hat{J}_z$, defining

$$\hat{\boldsymbol{J}}^{\boldsymbol{2}} \overset{def}{=} \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2. \tag{4.4.1}$$

This new operator commutes with all $\hat{J}_i$'s, i.e.

$$[\hat{\boldsymbol{J}}^{\boldsymbol{2}}, \hat{J}_i] = 0, \quad i = x, y, z. \tag{4.4.2}$$

An operator which commutes with all other operators in the algebra is called a **Casimir operator**. These are of great importance and we will return to their general role in section 5.5. That our $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ really is a Casimir operator can be seen by

$$\begin{aligned}[\hat{\boldsymbol{J}}^{\boldsymbol{2}}, \hat{J}_z] &= (\hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2)\hat{J}_z - \hat{J}_z(\hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2)\\ &= \hat{J}_x^2\hat{J}_z + \hat{J}_z^2\hat{J}_z - \hat{J}_z\hat{J}_x^2 - \hat{J}_z\hat{J}_y^2\\ &= \hat{J}_x[\hat{J}_x, \hat{J}_z] + \hat{J}_x\hat{J}_z\hat{J}_x + [\hat{J}_x, \hat{J}_z]\hat{J}_x - \hat{J}_x\hat{J}_z\hat{J}_x\\ &\qquad + \hat{J}_y[\hat{J}_y, \hat{J}_z] + \hat{J}_y\hat{J}_z\hat{J}_y + [\hat{J}_y, \hat{J}_z]\hat{J}_y - \hat{J}_y\hat{J}_z\hat{J}_y\\ &= \hat{J}_x(-i\hbar\hat{J}_y) + (-i\hbar\hat{J}_y)\hat{J}_x + \hat{J}_y i\hbar\hat{J}_x + i\hbar\hat{J}_x\hat{J}_y = 0,\end{aligned} \tag{4.4.3}$$

which can be done analogously for $\hat{J}_x$ and $\hat{J}_y$ through permutations. Since $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ commutes with any of the $\hat{J}_i$'s we can choose a basis for the Hilbert space such that it diagonalize both $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ and the $\hat{J}_i$ of choice. By convention we choose $\hat{J}_z$ for this and hence the basis consists of simultaneous eigenstates to $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ and $\hat{J}_z$. We label these eigenstates with the corresponding eigenvalues to $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ and $\hat{J}_z$ respectively by

$$|a,b\rangle : \quad \hat{\boldsymbol{J}}^{\boldsymbol{2}}|a,b\rangle = a|a,b\rangle, \quad \hat{J}_z|a,b\rangle = b|a,b\rangle \tag{4.4.4}$$

where $a, b \in \mathbb{R}$. Note that since all $\hat{J}_i$'s are Hermitian this also applies for $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$. Our goal now is to find all states in this basis which will prove to constitute an irreducible representation for the $\mathfrak{su}(2)$ algebra. To our help we define two new operators

$$\hat{J}_+ \overset{def}{=} \hat{J}_x + i\hat{J}_y, \qquad \hat{J}_- \overset{def}{=} \hat{J}_x - i\hat{J}_y. \tag{4.4.5}$$

These are non-Hermitian operators satisfying the commutation relations

$$[\hat{J}_z, \hat{J}_\pm] = \pm\hbar\hat{J}_\pm \tag{4.4.6a}$$

$$[\hat{J}_+, \hat{J}_-] = 2\hbar\hat{J}_z \tag{4.4.6b}$$

and by (4.4.2)

$$[\hat{\boldsymbol{J}}^{\boldsymbol{2}}, \hat{J}_\pm] = 0. \tag{4.4.6c}$$

Also note that since $\hat{J}_x$ and $\hat{J}_y$ are Hermitian we have

$$\hat{J}_+^\dagger = \hat{J}_- \iff \hat{J}_-^\dagger = \hat{J}_+. \tag{4.4.7}$$

The observant reader may protest that the complex linear combination of the $\hat{J}_i$'s takes us out of the $\mathfrak{su}(2)$ since this algebra was over the field $\mathbb{R}$. That is indeed true and we are now in the $\mathfrak{sl}(2,\mathbb{C})$ algebra. But let us not worry about this for the moment and trust the assertion that the result will hold for $\mathfrak{su}(2)$ as well.

$\hat{J}_\pm$ are called **ladder operators** and the reason is their action on the eigenstates (4.4.4). $\hat{J}_+|a,b\rangle$ is still an eigenstate to $\hat{J}_z$ since

$$\hat{J}_z\hat{J}_+|a,b\rangle = ([\hat{J}_z,\hat{J}_+] + \hat{J}_+\hat{J}_z)|a,b\rangle \overset{(4.4.6a)}{=} (b+\hbar)\hat{J}_+|a,b\rangle \tag{4.4.8}$$

and the eigenvalue is $(b+\hbar)$. A similar calculation gives

$$\hat{J}_z\hat{J}_-|a,b\rangle = (b-\hbar)\hat{J}_-|a,b\rangle, \tag{4.4.9}$$

so the action of $\hat{J}_+$ ($\hat{J}_-$) is apparently to increase (decrease) the eigenvalue of $\hat{J}_z$ and we thereby have the proportionality relations

$$\hat{J}_\pm|a,b\rangle = c_\pm|a,b\pm\hbar\rangle. \tag{4.4.10}$$

The proportionality constants $c_\pm$ are to be determined later on.

We can now walk up and down amongst the eigenvalues with these operators with $(\hat{J}_\pm)^n|a,b\rangle \propto |a,b\pm n\hbar\rangle$ where $n \in \mathbb{N}_0$, but are there limits? We start answering this question by proving that the dimension of the Hilbert space is finite, i.e. there is a finite set of eigenstates, and that the eigenvalues to $\hat{J}_z$ is bounded. To accomplish this we study the eigenvalue a of $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$. Since $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ commutes with both $\hat{J}_\pm$ (see equation (4.4.6c)) a has to be the

same for all $|a, b \pm n\hbar\rangle$. Now, by exploiting the commutation relations we can rewrite $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ in several ways. In particular we have

$$\hat{\boldsymbol{J}}^{\boldsymbol{2}} = \hat{J}_z^2 + \hat{J}_-\hat{J}_+ + \hbar\hat{J}_z \tag{4.4.11a}$$

$$\hat{\boldsymbol{J}}^{\boldsymbol{2}} = \hat{J}_z^2 + \hat{J}_+\hat{J}_- - \hbar\hat{J}_z \tag{4.4.11b}$$

since

$$\hat{J}_\pm\hat{J}_\mp = \hat{J}_x^2 + \hat{J}_y^2 \pm i(\hat{J}_y\hat{J}_x - \hat{J}_x\hat{J}_y). \tag{4.4.12}$$

By the use of equations (4.4.11a) and (4.4.11b) we find that

$$\hat{\boldsymbol{J}}^{\boldsymbol{2}} - \hat{J}_z^2 = \frac{1}{2}(\hat{J}_+\hat{J}_- + \hat{J}_-\hat{J}_+) \tag{4.4.13}$$

and calculating the expectation value for $\hat{\boldsymbol{J}}^{\boldsymbol{2}} - \hat{J}_z^2$ gives

$$(a - b^2) = \langle a, b|\hat{\boldsymbol{J}}^{\boldsymbol{2}} - \hat{J}_z^2|a, b\rangle = \langle a, b|\tfrac{1}{2}(\hat{J}_+\hat{J}_- + \hat{J}_-\hat{J}_+)|a, b\rangle \geqslant 0. \tag{4.4.14}$$

This is greater or equal to zero since by (4.4.7) the right hand side is just half the squared norm of $\hat{J}_-|a, b\rangle + \hat{J}_+|a, b\rangle$, which of course is positive semi-definite. This yields

$$a \geqslant b^2 \tag{4.4.15}$$

and the eigenvalues b are therefore proved to be bounded and thereby it exists a $b_{\max}$ and a $b_{\min}$ with the properties

$$\hat{J}_+|a, b_{\max}\rangle = 0 \tag{4.4.16a}$$

$$\hat{J}_-|a, b_{\min}\rangle = 0. \tag{4.4.16b}$$

To find the value of these eigenvalues we make use of equations (4.4.11a) and (4.4.11b) once more and act with $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ on the states with $b_{\max}$ and $b_{\min}$ respectively:

$$\hat{\boldsymbol{J}}^{\boldsymbol{2}}|a, b_{\max}\rangle = (\hat{J}_z^2 + \hat{J}_-\hat{J}_+ + \hbar\hat{J}_z)|a, b_{\max}\rangle$$

$$\Rightarrow a = b_{\max}(b_{\max} + \hbar) \tag{i}$$

$$\hat{\boldsymbol{J}}^{\boldsymbol{2}}|a, b_{\min}\rangle = (\hat{J}_z^2 + \hat{J}_+\hat{J}_- - \hbar\hat{J}_z)|a, b_{\min}\rangle$$

$$\Rightarrow a = b_{\min}(b_{\min} - \hbar) \tag{ii}$$

$$\xRightarrow{(i),(ii)} \quad b_{\min}(b_{\min} - \hbar) = b_{\max}(b_{\max} + \hbar) \tag{4.4.17}$$

which have the solutions $b_{\min} = b_{\max} + \hbar$ and $b_{\min} = -b_{\max}$. But $b_{\min} < b_{\max}$ so we must have

$$b_{\min} = -b_{\max} \tag{4.4.18}$$

and

$$a = b_{\max}(b_{\max} + \hbar). \tag{4.4.19}$$

Now, by climbing upwards from $b_{\min}$ with the repeated use of $\hat{J}_+$ we find that

$$b_{\max} = b_{\min} + n\hbar, \quad n \in \mathbb{N}. \tag{4.4.20}$$

By convention and for convenience we define j to be $j = b_{\max}/\hbar$, so that

$$j = \frac{n}{2} \tag{4.4.21}$$

being a integer or half-integer since $-j$ and j are separated by n. For any eigenvalue b we must have

$$b_{\min} = -j\hbar \leqslant b = m\hbar \leqslant j\hbar = b_{\max} \tag{4.4.22}$$

with m restricted to the $2j+1$ values $-j, -j+1, \ldots, j-1, j$.

Since a and b is completely determined by j and m ($a = \hbar^2 j(j+1), b = m\hbar$) we can change the labels of the eigenkets $|a, b\rangle \to |j, m\rangle$, which are the standard labels of the angular momentum states.

We now turn to the normalization constants in (4.4.10). Assume that $|j, m\rangle$ is normalized and $\hat{J}_\pm |j, m\rangle = c^\pm_{jm} |j, m \pm 1\rangle$. Calculating the squared norms yields

$$\begin{aligned}
\langle j, m | \hat{J}_+^\dagger \hat{J}_+ | j, m \rangle &\overset{(4.4.7)}{\overset{(4.4.11)}{=}} \langle j, m | \hat{\boldsymbol{J}}^{\boldsymbol{2}} - \hat{J}_z^2 - \hbar \hat{J}_z | j, m \rangle = \hbar^2 [j(j+1) - m(m+1)] \\
&= \\
&\langle j, m+1 | (c^+_{jm})^* c^+_{jm} | j, m+1 \rangle = |c^+_{jm}|^2 \\
&\Rightarrow |c^+_{jm}| = \hbar \sqrt{j(j+1) - m(m+1)},
\end{aligned} \tag{4.4.23}$$

i.e. c^+_{jm} is determined up to an arbitrary phase. We choose this phase equal to 1. A similar calculation for $\hat{J}_-$ gives us

$$|c^-_{jm}| = \hbar \sqrt{j(j+1) + m(m-1)}. \tag{4.4.24}$$

We sum up what we have achieved so far. Based entirely on the commutation relations of infinitesimal rotations we have found the simultaneous eigenstates for $\hat{J}_z$ and $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ for each integer or half-integer value j. All orthogonal to each other ($\hat{J}_z$ and $\hat{\boldsymbol{J}}^{\boldsymbol{2}}$ being Hermitian) these states form bases in $(2j+1)$-dimensional Hilbert spaces. With the $\hat{J}_i$'s corresponding to the generators of the Lie algebra $\mathfrak{sl}(2, \mathbb{C})$ these operators constitute $(2j+1)$-dimensional representations of this algebra. In fact, they are *irreducible* representations since there is no subspace of the Hilbert space that maps onto itself under all the $\hat{J}_i$'s.

Now that we have found the eigenstates we can return to real linear combinations of $\hat{J}_i$'s, i.e. to $\mathfrak{su}(2)$, and still have a basis. Although derived in $\mathfrak{sl}(2, \mathbb{C})$ they are still eigenkets to $\hat{J}_z$ and hence form a basis for the module corresponding to the $\mathfrak{su}(2)$ realization. This representation is also irreducible which should be apparent by rewriting the $\hat{J}_x$ and $\hat{J}_y$ as $\frac{1}{2}(\hat{J}_+ \pm \hat{J}_-)$.

This procedure to construct irreducible representations by the use of ladder operators may be generalized and will prove itself to be very useful further on. This will be treated in chapter 5.

4.4.1 Matrix Representation of the Ladder Operators

By use of the formula learned in the example 12 of chapter 3 we can find the matrix representations of the ladder operators

$$\langle j', m'|\hat{J}_z|j, m\rangle = \hbar m \delta_{m'm} \tag{4.4.25}$$

$$\langle j', m'|\hat{J}_\pm|j, m\rangle = \hbar\sqrt{j(j+1) \mp m(m \pm 1)}\delta_{j'j}\delta_{m'(m\pm1)}. \tag{4.4.26}$$

As an example we explicitly calculate these matrices for the case $j = 1/2$:

$$J_z = \begin{pmatrix} \frac{1}{2}\hbar & 0 \\ 0 & -\frac{1}{2}\hbar \end{pmatrix}, \qquad J_+ = \begin{pmatrix} 0 & \hbar \\ 0 & 0 \end{pmatrix}, \qquad J_- = \begin{pmatrix} 0 & 0 \\ \hbar & 0 \end{pmatrix}, \tag{4.4.27}$$

and compare them with the matrices obtained from the Pauli matrices and definitions (4.4.5)

$$J_z = \frac{\hbar}{2}\sigma_z = \begin{pmatrix} \frac{1}{2}\hbar & 0 \\ 0 & -\frac{1}{2}\hbar \end{pmatrix}, \tag{4.4.28}$$

$$J_\pm = \frac{\hbar}{2}\left[\sigma_x \pm i\sigma_y\right] = \frac{\hbar}{2}\left[\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \pm i\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\right] = \begin{cases} \begin{pmatrix} 0 & \hbar \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & 0 \\ \hbar & 0 \end{pmatrix}. \end{cases} \tag{4.4.29}$$

To our content they are the same. We now consider these matrices and note that J_z, as expected, is diagonal and that $J_\pm$ are strictly upper respectively lower triangular. This is not a coincidence but a general property of the ladder construction. This will be seen in the next chapter where a generalization and a systematic use of the ladder operators will be developed.

As a consolidation before digging further into the theory of Lie algebras we are now ready for an example which will illustrate many parts of what we have learned so far.

4.5 Summarizing Example: Addition of Angular Momenta

A common example to illustrate how representations work is the addition of angular momentum. The example is good because it contains many of the key terms we have gone through, such as irreducible representations, tensor products and ladder operators. What we want to do in this section is to find the representation for the total angular momentum of two particles. This will turn out to be a reducible representation.

Consider two particles with spin $S^{(1)}$ and $S^{(2)}$, respectively. Let $|\Psi_m^{(1)}\rangle$ and $|\Psi_n^{(2)}\rangle$ be the ket-basis for the two particles in their respective Hilbert space $\mathcal{H}^{(1)}$, $\mathcal{H}^{(2)}$. These are eigenstates to $\hat{S}_z^{(1)}$ and $\hat{S}_z^{(2)}$, respectively. In each of the Hilbert spaces we have our spin operators, $\hat{\boldsymbol{S}}^{2(1)}$, $\hat{S}_z^{(1)}$ and $\hat{\boldsymbol{S}}^{2(2)}$, $\hat{S}_z^{(2)}$. The state kets for the total system is now in the tensor product space $\mathcal{H} = \mathcal{H}^{(1)} \otimes \mathcal{H}^{(2)}$ with the basis $|\Psi_{m'}\rangle = |\Psi_m^{(1)}\rangle \otimes |\Psi_n^{(2)}\rangle$, $m = -S^{(1)}, -S^{(1)}+1,..., S^{(1)}$ and $n = -S^{(2)}, -S^{(2)}+1,..., S^{(2)}$. For example, the state $|\Psi_{S^{(1)}}^{(1)}\rangle \otimes |\Psi_{S^{(2)}}^{(2)}\rangle$ corresponds to when the two particles z-component of the spin is $S^{(1)}\hbar$ and $S^{(2)}\hbar$. In order to add the spin

we need to find the corresponding operators on the ket space, Hilbert space $\mathcal{H}$, for the total system, i.e. $\hat{\boldsymbol{S}}^2$, $\hat{S}_z$.

Consider the infinitesimal rotation operators $\hat{D}_1(d\phi)$, $\hat{D}_2(d\phi)$ that acts on $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$, respectively, and form

$$\begin{aligned}\hat{D}_1(d\phi)\otimes\hat{D}_2(d\phi) &= \left(\mathbb{1}-\frac{i\hat{S}_z^{(1)}d\phi}{\hbar}\right)\otimes\left(\mathbb{1}-\frac{i\hat{S}_z^{(2)}d\phi}{\hbar}\right)= \\ &= \mathbb{1}-\mathbb{1}\otimes\frac{i\hat{S}_z^{(2)}d\phi}{\hbar}-\frac{i\hat{S}_z^{(1)}d\phi}{\hbar}\otimes\mathbb{1}=\mathbb{1}-\frac{i(\hat{S}_z^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_z^{(2)})d\phi}{\hbar}\end{aligned} \tag{4.5.1}$$

where we have omitted the quadratic term $d\phi^2$. The same thing can be done for the x- and y-components, so, the total spin operator, i.e. the operator that acts on $\mathcal{H}$, is therefore $\hat{S}=\hat{S}^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}^{(2)}$. The ladder operators

$$\begin{aligned}\hat{S}_\pm &= \hat{S}_x\pm i\hat{S}_y=\hat{S}_x^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_x^{(2)}\pm i(\hat{S}_y^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_y^{(2)})= \\ &= \hat{S}_x^{(1)}\otimes\mathbb{1}\pm i\hat{S}_y^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_x^{(2)}\pm i\mathbb{1}\otimes\hat{S}_y^{(2)}=\hat{S}_\pm^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_\pm^{(2)}\end{aligned} \tag{4.5.2}$$

will help us to create another base element in $\mathcal{H}$. In analogy to what we did in the preceding section 4.4 we now want to find all the eigenstates and eigenvalues to $\hat{\boldsymbol{S}}^2$ and $\hat{S}_z$. Let us begin with $|\Psi_{S^{(1)}}\rangle\otimes|\Psi_{S^{(2)}}\rangle$ and see if this state is an eigenstate,

$$\begin{aligned}&\hat{S}_z|\Psi_{S^{(1)}}^{(1)}\rangle\otimes|\Psi_{S^{(2)}}^{(2)}\rangle=(\hat{S}_z^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_z^{(2)})|\Psi_{S^{(1)}}^{(1)}\rangle\otimes|\Psi_{S^{(2)}}^{(2)}\rangle= \\ &=\hat{S}_z^{(1)}|\Psi_{S^{(1)}}^{(1)}\rangle\otimes|\Psi_{S^{(2)}}^{(2)}\rangle+|\Psi_{S^{(1)}}^{(1)}\rangle\otimes\hat{S}_z^{(2)}|\Psi_{S^{(2)}}^{(2)}\rangle=(S_z^{(1)}+S_z^{(2)})|\Psi_{S^{(1)}}^{(1)}\rangle\otimes|\Psi_{S^{(2)}}^{(2)}\rangle.\end{aligned} \tag{4.5.3}$$

Thus, the z-component for the total angular momentum is just the sum of the constituent particles z-components. This state must correspond to spin $S^{(1)}+S^{(2)}$, since $S^{(1)}$ and $S^{(2)}$ are the highest eigenvalues of $\hat{S}_z^{(1)}$ and $\hat{S}_z^{(2)}$ respectively. We can continue by calculating the eigenvalue for $\hat{S}_-|\Psi_{S^{(1)}}^{(1)}\rangle\otimes|\Psi_{S^{(2)}}^{(2)}\rangle$ with respect to $\hat{S}_z$. This will give us the eigenvalue $S^{(1)}+S^{(2)}-1$. The eigenvalue $S^{(1)}+S^{(2)}-1$ can be otained in two ways; the sum of $m=S^{(1)}-1$ and $n=S^{(2)}$ or the sum of $m=S^{(1)}$ and $n=S^{(2)}-1$. A state with $\hat{S}_z$-eigenvalue $S^{(1)}+S^{(2)}-1$ can obviously be in a spin $S^{(1)}+S^{(2)}$ state but it can also be in a spin $S^{(1)}+S^{(2)}-1$ state. $\hat{S}_-|\Psi_{S^{(1)}}^{(1)}\rangle\otimes|\Psi_{S^{(2)}}^{(2)}\rangle$ will be in the spin $S^{(1)}+S^{(2)}$ state and by finding its orthogonal complement we have the state corresponding to spin $S^{(1)}+S^{(2)}-1$ The latter state will be the highest state of the spin $S^{(1)}+S^{(2)}-1$ states. We can confirm that these states have different spin if we calculate the $\hat{\boldsymbol{S}}^2$-eigenvalues. By repeating this procedure and making use of ladder operators we get a breakdown of the Hilbert space in which each subspace corresponds to a spin value. These values are: $|S^{(1)}-S^{(2)}|$, $|S^{(1)}-S^{(2)}|+1$,..., $S^{(1)}+S^{(2)}$. The subspace corresponding to spin S has dimension $2S+1$. The Hilbert space $\mathcal{H}=\mathcal{H}^{(1)}\otimes\mathcal{H}^{(2)}$ can be expressed as a *direct sum* of these subspaces, since these are orthogonal,

$$\mathcal{H}=\mathcal{H}^{(|S^{(1)}-S^{(2)}|)}\oplus\mathcal{H}^{(|S^{(1)}-S^{(2)}+1|)}\oplus...\oplus\mathcal{H}^{(S^{(1)}+S^{(2)})} \tag{4.5.4}$$

and the total spin operator $\hat{S}_z=\hat{S}_z^{(1)}\otimes\mathbb{1}+\mathbb{1}\otimes\hat{S}_z^{(2)}$ can now be expressed as a *direct sum* of operators

$$\hat{S}_z=\hat{S}_z^{(|S^{(1)}-S^{(2)}|)}\oplus\hat{S}_z^{(|S^{(1)}-S^{(2)}+1|)}\oplus...\oplus\hat{S}_z^{(S^{(1)}+S^{(2)})}, \tag{4.5.5}$$

where the operators in the product acts on its corresponding subspace. In fact, this is a part of a reducible representation for $\mathfrak{su}(2)$ and together with ladder operators we have a *reducible* representation expressed as a sum of *irreducible* representations. Let us call the representation for $\mathfrak{su}(2)$ π,

$$\pi = \pi^{S^{(1)}} \otimes \pi^{S^{(2)}} = \pi^{(|S^{(1)}-S^{(2)}|)} \oplus \pi^{(|S^{(1)}-S^{(2)}|+1)} \oplus ... \oplus \pi^{(S^{(1)}+S^{(2)})}, \tag{4.5.6}$$

where the $\pi^{(i)}$:s are the irreducible representations of $\mathfrak{su}(2)$ acting on the vector spaces $\mathcal{H}^{(i)}$.

We have skipped some steps in the calculations but it is not these that we are interested in but rather the results observed in terms of representation theory. We started from two vector spaces with a set of operators corresponding to these spaces. These operators $\hat{S}_z^{(1)}$, $\hat{S}_z^{(2)}$, $\hat{S}_\pm^{(1)}$, $\hat{S}_\pm^{(2)}$ form irreducible representations of the Lie algebra $\mathfrak{su}(2)$ acting on a Hilbert space of dimension $2S^{(1)}+1$ and $2S^{(2)}+1$. We cannot find any invariant subspaces of $\mathcal{H}^{(1)}$ or $\mathcal{H}^{(2)}$. When we act with the total spin operator $\hat{S}_z$ and the ladder operators $\hat{S}_\pm$ on $\mathcal{H}$ it will "split" into operators acting on their corresponding vector space. We can apply the $\hat{S}_\pm$ to an element in $\mathcal{H}$ as many times as we want but it will never leave its subspace.

One interesting thing we have learned on the way is how to form the tensor product between two irreducible representations of $\mathfrak{sl}(2,\mathbb{C})$: We have seen that irreducible representations of $\mathfrak{su}(2)$ are characterized by a their spin j, and the dimension of the module of such a representation is a $(2j+1)$-dimensional vector space. Such a representation is often denoted simply by either the dimension of its module in bold, or simply as $[j]$. So for example

$$[0] = \mathbf{1} \tag{4.5.7}$$
$$[1/2] = \mathbf{2} \tag{4.5.8}$$
$$[1] = \mathbf{3}.$$

Equation (4.5.4) tells us the action of the tensor product on such representations. It states that

$$[j] \otimes [j'] = \bigoplus_{i=|j-j'|}^{j+j'} [i]. \tag{4.5.9}$$

So we have, for example:

$$[0] \otimes [1/2] = \bigoplus_{i=1/2}^{1/2} [i] = [1/2] \Longleftrightarrow \mathbf{1} \otimes \mathbf{2} = \mathbf{2} \tag{4.5.10}$$

$$[1/2] \otimes [1/2] = \bigoplus_{i=0}^{1} [i] = [0] \oplus [1] \Longleftrightarrow \mathbf{2} \otimes \mathbf{2} = \mathbf{1} \oplus \mathbf{3}. \tag{4.5.11}$$

But perhaps the most important example of the tensor product between two representations of $\mathfrak{sl}(2,\mathbb{C})$ is $[j]\otimes[1/2]$, since it represents the composition of a system of spin j with one of spin 1/2, which could represent the sum of the orbital angular momentum of an electron in an atom and the electron spin. In this case we have

$$[j] \otimes [1/2] = \bigoplus_{i=j-1/2}^{j+1/2} [i] = [j-1/2] \oplus [j+1/2] \Longleftrightarrow (\mathbf{2j+1}) \otimes \mathbf{2} = \mathbf{2j} \oplus (\mathbf{2j+2}). \tag{4.5.12}$$

To summarize, in this chapter the study of rotations naturally gave rise to the $\mathfrak{su}(2)$ commutations relations exposing a connection between the physical world and group theory. In analogy with the Hamiltonian and the momentum operators we defined the angular momentum operators as the ones generating rotations. Not surprisingly, satisfying the $\mathfrak{su}(2)$ their 2-dimensional representation proved to be the Pauli matrices. To find the module of any of their representations we defined the ladder operators and by climbing up- and downwards with these we found a basis for each $(2j+1)$-dimensional module. We also calculated the matrices of the ladder operators. As a conclusion we exemplified most of what have been done so far in the addition of angular momenta.

In the following chapter we will turn a bit more to the mathematics and delve deeper into Lie algebras.

Chapter 5

Lie Algebra

We are now ready to expand our knowledge of Lie algebras and put things we have done so far in a wider context. Basically everything we did in section 4.4 will return in this chapter and we will see how to classify Lie algebras. Above all, we will see how we can generalize step operators and our previously used "$\hat{J}_z$" operators. Important concepts such as roots, weights and the Cartan-matrix will be introduced and the chapter ends with an instructive example about $\mathfrak{sl}(3,\mathbb{C})$ where most of the formal theory will be demonstrated. The purpose of this chapter is to immerse ourselves in Lie algebras and present the formal structure of a semi-simple Lie algebra and its properties.

5.1 Structure Theory

Let us begin by defining two basic concepts.

Definition 5.1.1. *A* **subalgebra** $\mathfrak{h}$ *to a Lie algebra* $\mathfrak{g}$ *is a subspace of* $\mathfrak{g}$ *that is closed under the Lie bracket.*

Definition 5.1.2. *An* **ideal** *is a subalgebra* $\mathfrak{h}$ *of a Lie algebra* $\mathfrak{g}$ *with the following property. If* $t \in \mathfrak{h}$ *and* $h \in \mathfrak{g}$*, then* $[t,h] \in \mathfrak{h}$ *for every* h *and every* t*, i.e. it is an invariant subalgebra. An algebra with no non-trivial ideals, i.e. the 0 element and the algebra itself, is a* **simple algebra**. *A Lie algebra with no non-trivial Abelian ideals is a* **semi-simple Lie algebra**.

This is the Lie algebra's equivalent to a simple group. One characteristic feature of semi-simple Lie algebras is that they can be expressed as a direct sum of simple ideals each of which can be considered as a simple Lie algebra. Examples of semi-simple Lie algebras are $\mathfrak{so}(1,3)$, $\mathfrak{sl}(2,\mathbb{C})$, $\mathfrak{sl}(3,\mathbb{C})$, $\mathfrak{so}(2)$, $\mathfrak{so}(3)$, $\mathfrak{su}(2)$ and $\mathfrak{su}(3)$.

We start with a semi-simple Lie algebra $\mathfrak{g}$ of dimension n and we will often refer to $\mathfrak{g}$ as a vector space. $\mathfrak{g}$ consists of the linearly independent elements $t_1, t_2, ..., t_n$ so that

$$\mathfrak{g} = \text{span}_{\mathbb{C}}\{t_1, t_2, ..., t_n\} \tag{5.1.1}$$

and these elements thus represent a basis for $\mathfrak{g}$ but of course, we are free to make a change of basis. By a suitable choice of basis we may find the following characteristics when one

considers the adjoint representation of $\mathfrak{g}$. There exists a subalgebra $\mathfrak{C} = \text{span}_{\mathbb{C}}\{h_1, h_2, ..., h_r\}$ of dimension r, $h_i \in \mathfrak{g}$, such that

$$\begin{aligned} &\text{ad}_{h_i}(h_k) = [h_i, h_k] = 0, \quad \text{and} \\ &\text{ad}_{h_i}(t_j) = [h_i, t_j] = \beta_i^{(j)} t_j, \quad \forall i, k = 1, 2, ..., r, \quad \forall j = r+1, ..., n. \end{aligned} \tag{5.1.2}$$

This means that every element t_j, that is not in $\mathfrak{C}$, is an eigenvector to ad_{h_i} with eigenvalue $\beta_i^{(j)}$ for all $i = 1, 2, ..., r$. For every t_j we have r eigenvalues $\beta_1^{(j)}, \beta_2^{(j)}, ..., \beta_r^{(j)}$ and these eigenvalues can be thought of as components in a r-dimensional vector $\boldsymbol{\alpha}_j = (\beta_1^{(j)}, \beta_2^{(j)}, ..., \beta_r^{(j)})$ called a **root**. Thus, we can associate a root to each of the elements t_j that is not in $\mathfrak{C}$. Actually, we can allow t_j to be in $\mathfrak{C}$ but this will only mean that its root is 0 since it is a linear combination of the elements $h_1, h_2, ..., h_r$. A similar way to define α_j is as a linear map, $\alpha_j : \mathfrak{C} \longrightarrow \mathbb{R}$, that gives us an eigenvalue, i.e. $\alpha^{(j)}(h_i) = \beta_i^{(j)}$. A linear map that takes a vector space into a real number is a functional and these constitute the dual space $\mathfrak{C}^*$. So, we have divided the vector space $\mathfrak{g}$ into two subspaces: $\mathfrak{C}$ and its complement,

$$\mathfrak{g} = \mathfrak{C} \oplus \left(\bigoplus_{j=r+1}^{n} \mathbb{C} t_j \right). \tag{5.1.3}$$

The subalgebra $\mathfrak{C}$ is called the **Cartan subalgebra** and we restrict ourselves to the case where the elements $h_1, h_2, ..., h_r$ are Hermitian. Nothing is said about t_j though. The next step is to organize the t_j's, i.e. the complement to $\mathfrak{C}$. The only thing we know about these vectors is that they are eigenvectors to $\mathfrak{C}$, but otherwise unknown,

$$[h_i, t_j] = \beta_i^{(j)} t_j. \tag{5.1.4}$$

Now, take the Hermitian conjugate on both sides

$$\begin{gathered} [h_i, t_j]^\dagger = (h_i t_j)^\dagger - (t_j h_i)^\dagger = t_j^\dagger h_i - h_i t_j^\dagger = -[h_i, t_j^\dagger] = \beta_i^{(j)} t_j^\dagger \Rightarrow \\ [h_i, t_j^\dagger] = -\beta_i^{(j)} t_j^\dagger. \end{gathered} \tag{5.1.5}$$

and we see that $t_j^\dagger$ is an eigenvector with the negative root of t_j. This means that for every element t_j there exists an element $t_j^\dagger$ and this implies that $n - r$ must be even. We have $(n-r)/2$ t_j vectors and $(n-r)/2$ $t_j^\dagger$ vectors. We denote these as e_j and f_j respectively and these are called **root vectors**[1]. Let us call the subspace $\text{span}_{\mathbb{C}}\{e_1, e_2, ..., e_{(n-r)/2}\} = \eta_+$ and the subspace $\text{span}_{\mathbb{C}}\{f_1, f_2, ..., f_{(n-r)/2}\} = \eta_-$ where the choice of the indices $+, -$ will be justified later. To summarize, our algebra can thus be written as

$$\begin{gathered} \mathfrak{g} = \eta_- \oplus \mathfrak{C} \oplus \eta_+ \\ \mathfrak{C} = \bigoplus_{i=1}^{r} \mathbb{C} h_i, \qquad \eta_+ = \bigoplus_{i=1}^{(n-r)/2} \mathbb{C} e_i, \qquad \eta_- = \bigoplus_{i=1}^{(n-r)/2} \mathbb{C} f_i \\ [h_i, h_k] = 0 \qquad [h_i, e_j] = \beta_i^{(j)} e_j \qquad [h_i, f_j] = -\beta_i^{(j)} f_j. \end{gathered} \tag{5.1.6}$$

[1] We have to be careful not to confuse *roots* with *root vectors*.

This is known as **the triangular decomposition**. Since η_-, $\mathfrak{C}$, and η_+ are subspaces, the direct sum in (5.1.6) shall be interpreted as a direct sum of *vector spaces*, not Lie algebras. If it was a direct sum of Lie algebras, η_-, $\mathfrak{C}$, and η_+ would all commute.

Before we move on, let us say something about the Cartan subalgebra $\mathfrak{C}$. We said that through a suitable choice of basis for the Lie algebra there exists a Cartan algebra consisting of commuting vectors h_i. In practice when we want to find the Cartan algebra we find the maximal commuting elements in the Lie algebra. This seems familiar from quantum mechanics when we want to find eigenstates to as many operators as possible. Observable quantities in quantum mechanics are Hermitian operators, just as the elements in our Cartan algebra. We can thus think of the Cartan algebra as the set of operators, corresponding to physical quantities, where all the operators can be diagonalized simultaneously. The physical interpretation of an eigenvector to several Hermitian operators is that we have a state with known values of all the corresponding physical quantities. An algebra consisting of r commuting vectors is an algebra of **rank** r and we can thus label our physical states with r numbers. We can make an observation about the basis for $\mathfrak{C}$: our choice of basis to $\mathfrak{C}$ does not affect the obtained eigenvectors e_i, f_i. That is, if we make a change of basis from the $\{h_1, h_2, ..., h_r\}$ basis used above to another basis $\{h_1', h_2', ..., h_r'\}$ we obtain the same eigenvectors e_i, f_i, but with different eigenvalues.

We have seen that every element in $\mathfrak{g}$ has an associated vector in an r-dimensional vector space. This set of vectors is denoted Σ, and the vectors are expressed with respect to the $\{h_1, h_2, ..., h_r\}$ basis, i.e. if we change basis of $\mathfrak{C}$ the components of the roots changes. This is called the h-basis. If we have $n - r$ non-zero roots in a r-dimensional vector space in general all these vectors can not be linearly independent. We can choose a basis for the root space as follows: arrange the components of the roots for a given order of the Cartan vectors $h_1, h_2, ..., h_r$. A root with the first component positive is said to be a **positive root** and similarly, a root with its first non-zero component negative is said to be a **negative root**. Moreover, a positive root that can not be written as a linear combination of the other positive roots is said to be a **simple root**. This is not the only definition[2] of a positive and negative root and we will come back to this in section 5.3. Let r vectors of η_+ be simple roots and take these as a basis for the root space, i.e. $\Sigma = \text{span}_{\mathbb{R}}(\alpha_1, \alpha_2, ..., \alpha_r)$. This means that every other vector in the root space can be expressed as a linear combination of these. But, it does not mean that every linear combination of the simple roots is a root. This will soon be seen in (5.1.10) and (5.1.11). It seems that there is a correlation between the root vectors in η_+ and in η_-. Let us examine this by first looking at the commutator between two vectors in η_+, say e_1 and e_2. Consider

$$[e_1, e_2] \overset{def}{=} N_{1,2} e_k \tag{5.1.7}$$

where k will be specified. It is not obvious a priori that e_k will be one of the generators of η_+, so for now you can regard e_k simply as an arbitrary element of the Lie algebra. $N_{1,2}$ is a constant that we will determine later. We are now interested in how e_k commutes with the rest of the algebra, especially $\mathfrak{C}$.

$$\begin{aligned}[h_i, e_k] = [h_i, [e_1, e_2]] = /\text{Jacobi identity}/ = -[e_1, [e_2, h_i]] - [e_2, [h_i, e_1]] = [e_1, \beta_i^{(2)} e_2] \\ - [e_2, \beta_i^{(1)} e_1] = (\beta_i^{(1)} + \beta_i^{(2)}) N_{1,2} e_k.\end{aligned} \tag{5.1.8}$$

[2] Actually, the definitions of positive and negative roots are arbitrary but this is a natural way to order the roots since there always is a root of opposite sign.

This means that the commutator is an eigenvector to ad_{h_i} with an eigenvalue equal to the sum of the eigenvalues of elements in the commutator. Since the commutator (5.1.7) is another element in η_+ we simply denote it with the index 3, that is $k = 3$. Note that we do not use the word root here. The reason is that we do not know if e_3 is equal to zero or not. If it is equal to zero, its eigenvalues do *not* constitute a root, and if it is not equal to zero then its eigenvalues do constitute a root. Consider a trivial case

$$[h_i, [e_j, e_j]] = 2\beta_i^{(j)}[e_j, e_j] = 0, \tag{5.1.9}$$

which indicates that $2\beta_i^{(j)}$ is not a root. This is not a proof since for example $[e_{j-1}, e_{j+1}]$ would also give us $2\beta_i^{(j)}$. Nevertheless, it is true and this implies that $[e_{j-1}, e_{j+1}] = 0$ for all j. Actually, ± 1 is the only multiple of a root that gives us another root and the difference between two simple roots is *not* a root. Let $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ be two simple roots and assume that $\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2$ is a root. Then it is either positive or negative. If it is equal to zero, then α_1 and α_2 can not both be simple roots. If positive, we can express $\boldsymbol{\alpha}_1$ as

$$\boldsymbol{\alpha}_1 = (\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2) + \boldsymbol{\alpha}_2 \tag{5.1.10}$$

and if negative we can express $\boldsymbol{\alpha}_2$ as

$$\boldsymbol{\alpha}_2 = (\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1) + \boldsymbol{\alpha}_1. \tag{5.1.11}$$

Since $(\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2)$ is positive in the first case and $(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)$ is positive in the second case we get a contradiction. $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ can not be expressed as linear combinations of other positive roots since they are assumed simple. This implies that $[e_i, f_j] = 0$ if $i \neq j$ and $i, j = 1, 2, ..., r$. That is, corresponds to a simple root. So to summarize,

$$\begin{gathered}
\text{If } [e_k, e_l] = N_{k,l}e_{k+l} \neq 0 \text{ and } [h_i, N_{k,l}e_{k+l}] = \beta_i^{(k+l)}N_{k,l}e_{k+l} = (\beta_i^{(k)} + \beta_i^{(l)})N_{k,l}e_{k+l} \Rightarrow \\
\boldsymbol{\alpha}_{k+l} \text{ is a root.} \\
\text{If } [e_k, e_l] = N_{k,l}e_{k+l} = 0 \text{ and } [h_i, N_{k,l}e_{k+l}] = \beta_i^{(k+l)}N_{k,l}e_{k+l} = (\beta_i^{(k)} + \beta_i^{(l)})N_{k,l}e_{k+l} \Rightarrow \\
\boldsymbol{\alpha}_{k+l} \text{ is } \textit{not} \text{ a root.}
\end{gathered} \tag{5.1.12}$$

We can do the exact same calculations for f_j. Next, let us consider the case when $\boldsymbol{\alpha} = 0$. This happens when we take the commutator between two elements of the form e_j and f_j and then take the multicommutator with h_j.

$$[h_i, [e_j, f_j]] = (\beta_i^{(j)} - \beta_i^{(j)})[e_j, f_j] = 0, \tag{5.1.13}$$

i.e. $[e_j, f_j]$ commutes with h_i which means that $[e_j, f_j] \in \mathfrak{C}$. We can therefore express $[e_j, f_j]$ as a linear combination of the h_i's.

$$[e_j, f_j] = a^i h_i, \quad a^i \in \mathbb{R} \tag{5.1.14}$$

where we use the Einstein summation convention. a^i is to be determined later on. What we have found so far is thus

$$\begin{aligned}
&[h_i, e_j] = \beta_i^{(j)} e_j \\
&[h_i, f_j] = -\beta_i^{(j)} f_j \\
&[e_k, e_l] = N_{k,\,l} e_{k+l} \\
&[h_i, [e_k, e_l]] = (\beta_i^{(k)} + \beta_i^{(l)})[e_k, e_l] \quad \text{if } (\boldsymbol{\alpha}_k + \boldsymbol{\alpha}_l) \in \Sigma \\
&[h_i, [e_k, e_l]] = 0, \qquad\qquad\qquad\quad \text{if } (\boldsymbol{\alpha}_k + \boldsymbol{\alpha}_l) \notin \Sigma \\
&[e_j, f_j] = a^i h_i \qquad\qquad\qquad\quad a^i \in \mathbb{R} \\
&[e_j, f_k] = 0 \qquad\qquad\qquad\qquad \text{if } \boldsymbol{\alpha}_j, \boldsymbol{\alpha}_k \text{ are simple roots}
\end{aligned} \tag{5.1.15}$$

where Σ denotes the set of roots. We can generalize (5.1.7) by considering the multicommutator

$$[e_{i_1}, [e_{i_2}, [e_{i_3} ..., [e_{i_{m-1}}, e_{i_m}]...]]], \quad m = 1, 2, ..., r \tag{5.1.16}$$

with possible root

$$\boldsymbol{\alpha}_{i_1} + \boldsymbol{\alpha}_{i_2} + \boldsymbol{\alpha}_{i_3} + ... + \boldsymbol{\alpha}_{i_{m-1}} + \boldsymbol{\alpha}_{i_m}, \quad m = 1, 2, ..., r. \tag{5.1.17}$$

By commuting elements in η_+ corresponding to our simple roots we obtain all other generators e_i in η_+. This can be seen from (5.1.17). The roots $\boldsymbol{\alpha}_i$ form a basis for Σ, thus we can express every root corresponding to a root vector, $e_k \in \eta_+$, as a linear combination of the $\boldsymbol{\alpha}_i$'s. This linear combination corresponds to a multicommutator of the form (5.1.16).

Since the dimension of our algebra is finite we can not for example construct a new element of the form

$$\underbrace{[t_j, [t_j, [t_j ..., [t_j}_{k \text{ times}}, t_l]...]]], \quad t_j, t_l \in \eta_+ \cup \eta_-, \tag{5.1.18}$$

by taking an "infinite commutator". For convenience, label the t's with its root. There must be two positive integers p, q for every pair of root vectors $t_{\boldsymbol{\alpha}_1}, t_{\boldsymbol{\alpha}_2}$ such that

$$\underbrace{[t_{\boldsymbol{\alpha}_1}, [t_{\boldsymbol{\alpha}_1}, [t_{\boldsymbol{\alpha}_1} ..., [t_{\boldsymbol{\alpha}_1}}_{q+1 \text{ times}}, t_{\boldsymbol{\alpha}_2}]...]]] = 0, \quad \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \Sigma \tag{5.1.19}$$

and

$$\underbrace{[t_{-\boldsymbol{\alpha}_1}, [t_{-\boldsymbol{\alpha}_1}, [t_{-\boldsymbol{\alpha}_1} ..., [t_{-\boldsymbol{\alpha}_1}}_{p+1 \text{ times}}, t_{\boldsymbol{\alpha}_2}]...]]] = 0, \quad \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \Sigma \tag{5.1.20}$$

with the corresponding eigenvalues $\boldsymbol{\alpha}_2 + (q+1)\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2 - (p+1)\boldsymbol{\alpha}_1$ respectively. This means that we will get a finite number of generators in our "chain" of generators. The still unknown constants $N_{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2}$ from

$$[e_{\boldsymbol{\alpha}_1}, e_{\boldsymbol{\alpha}_2}] = N_{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2} e_{\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2} \tag{5.1.21}$$

satisfying

$$N_{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2} = -N^*_{-\boldsymbol{\alpha}_2, -\boldsymbol{\alpha}_1} = -N_{-\boldsymbol{\alpha}_2, -\boldsymbol{\alpha}_1} \tag{5.1.22}$$

will now be determined in order to find p and q. What follows is a fairly intense calculation that will give us an expression for $p - q$. Equation (5.1.22) follows from the fact that the

structure constants are real, see ([2], p.125). First of all we need some properties of $N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2}$, that will be used, and in order to obtain these we need to consider a representation π. Multiply the equation

$$[\pi(t_{\boldsymbol{\alpha}}),\pi(t_{-\boldsymbol{\alpha}})] = a^i\pi(h_i) \tag{5.1.23}$$

with $\pi(h_j)$ on both sides. For sake of readability we drop the $\pi(\cdot)$'s in this specific equation, remembering that we are still dealing with a representation:

$$\begin{aligned}
[t_{\boldsymbol{\alpha}},t_{-\boldsymbol{\alpha}}]h_j &= t_{\boldsymbol{\alpha}}t_{-\boldsymbol{\alpha}}h_j - t_{-\boldsymbol{\alpha}}t_{\boldsymbol{\alpha}}h_j = \\
&= \ /\text{add and subtract } t_{-\boldsymbol{\alpha}}h_jt_{\boldsymbol{\alpha}}\ / = \\
&= t_{\boldsymbol{\alpha}}t_{-\boldsymbol{\alpha}}h_j + t_{-\boldsymbol{\alpha}}h_jt_{\boldsymbol{\alpha}} - t_{-\boldsymbol{\alpha}}h_jt_{\boldsymbol{\alpha}} - t_{-\boldsymbol{\alpha}}t_{\boldsymbol{\alpha}}h_j = \\
&= t_{\boldsymbol{\alpha}}t_{-\boldsymbol{\alpha}}h_j - t_{-\boldsymbol{\alpha}}h_jt_{\boldsymbol{\alpha}} + t_{-\boldsymbol{\alpha}}[h_j,t_{\boldsymbol{\alpha}}] = \\
&= t_{\boldsymbol{\alpha}}t_{-\boldsymbol{\alpha}}h_j - t_{-\boldsymbol{\alpha}}h_jt_{\boldsymbol{\alpha}} + \boldsymbol{\alpha}_jt_{-\boldsymbol{\alpha}}t_{\boldsymbol{\alpha}} = \\
&= a^ih_ih_j.
\end{aligned} \tag{5.1.24}$$

Take the trace of the second last line and use that the trace is invariant under cyclic permutations, i.e.

$$\boldsymbol{\alpha}_j\mathrm{Tr}\left(\pi(t_{-\boldsymbol{\alpha}})\pi(t_{\boldsymbol{\alpha}})\right) = a^i\mathrm{Tr}\left(\pi(h_i)\pi(h_j)\right). \tag{5.1.25}$$

$\mathrm{Tr}(\pi(t_{-\boldsymbol{\alpha}})\pi(t_{\boldsymbol{\alpha}})$ can be chosen equal to one, for every $t \in \eta_+ \cup \eta_-$, since we can redefine $t_{\boldsymbol{\alpha}}$ with a constant. Define $\mathrm{Tr}(\pi(h^i)\pi(h^j))$ as

$$\begin{aligned}
&\mathrm{Tr}(\pi(h_i)\pi(h_j)) \overset{def}{=} g_{ij} \quad \Rightarrow \\
&\boldsymbol{\alpha}_j \overset{(5.1.25)}{=} a^ig_{ij}.
\end{aligned} \tag{5.1.26}$$

This is actually a metric for the vector space spanned by the simple roots and the scalar product thus becomes $\langle\boldsymbol{\alpha}_i,\boldsymbol{\alpha}_j\rangle = g_{ij}a^ia^j$. A metric can be seen as a function that provides a definition of distance in spaces and thus appears in scalar product. For a rigorous treatment of metrics, see appendix B.5. We need an identity. Let $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2$,$\boldsymbol{\alpha}_3$ be three roots satisfying

$$\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \boldsymbol{\alpha}_3 = 0. \tag{5.1.27}$$

Apply the Jacobi identity to $[\pi(t_{\boldsymbol{\alpha}_1}),[\pi(t_{\boldsymbol{\alpha}_2}),\pi(t_{\boldsymbol{\alpha}_3})]]$ corresponding to the three roots,

$$(a^iN_{\boldsymbol{\alpha}_2,\,\boldsymbol{\alpha}_3} + b^iN_{\boldsymbol{\alpha}_3,\,\boldsymbol{\alpha}_1} + c^iN_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2})\pi(h_i) = 0, \tag{5.1.28}$$

where a^i,b^i,c^i are the coefficients from (5.1.14). Multiply this with $\pi(h_j)$ and take the trace, just as we did above,

$$\begin{aligned}
&(a^iN_{\boldsymbol{\alpha}_2,\,\boldsymbol{\alpha}_3} + b^iN_{\boldsymbol{\alpha}_3,\,\boldsymbol{\alpha}_1} + c^iN_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2})\mathrm{Tr}(\pi(h_i)\pi(h_j)) = \\
&\overset{(5.1.26)}{=} (\boldsymbol{\alpha}_1)_iN_{\boldsymbol{\alpha}_2,\,\boldsymbol{\alpha}_3} + (\boldsymbol{\alpha}_2)_iN_{\boldsymbol{\alpha}_3,\,\boldsymbol{\alpha}_1} + (\boldsymbol{\alpha}_3)_iN_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2} = 0 \Rightarrow \\
&/\text{use (5.1.27)}/ \\
&N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2} = N_{\boldsymbol{\alpha}_2,\,-\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2} = N_{-\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_2,\,\boldsymbol{\alpha}_1}.
\end{aligned} \tag{5.1.29}$$

Now when we have (5.1.29) we can consider the Jacobi identity for the root vectors corresponding to $\boldsymbol{\alpha}_1$, $-\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2 + k\boldsymbol{\alpha}_1$, i.e. $t_{\boldsymbol{\alpha}_1},t_{-\boldsymbol{\alpha}_1},t_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1}$, $-p \le k \le q$

$$\begin{aligned}
&[t_{\boldsymbol{\alpha}_1},[t_{-\boldsymbol{\alpha}_1},t_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1}]] + [t_{-\boldsymbol{\alpha}_1},[t_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1},t_{\boldsymbol{\alpha}_1}]] = -[t_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1},[t_{\boldsymbol{\alpha}_1},t_{-\boldsymbol{\alpha}_1}]] = \\
&= -[t_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1},a^i_{\alpha_1}h_i] = a^i_{\alpha_1}(\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1)_it_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1},
\end{aligned} \tag{5.1.30}$$

where we have used (5.1.14). The subscript α_1 on a^i is to indicate the relation 5.1.26 between α_1 and a. Evaluation of the left hand side yields

$$a^i_{\alpha_1}(\boldsymbol{\alpha}_2 + k\boldsymbol{\alpha}_1)_i = N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2+(k-1)\boldsymbol{\alpha}_1} N_{-\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1} + N_{-\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2+(k+1)\boldsymbol{\alpha}_1} N_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_1} \tag{5.1.31}$$

where we have dropped the factor $t_{\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1}$. With the use of (5.1.22) and (5.1.29) and after some trickery we get

$$a^i_{\alpha_1}(\boldsymbol{\alpha}_2 + k\boldsymbol{\alpha}_1)_i = N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1} N_{-\boldsymbol{\alpha}_1,\,-\boldsymbol{\alpha}_2-k\boldsymbol{\alpha}_1} - N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2+(k-1)\boldsymbol{\alpha}_1} N_{-\boldsymbol{\alpha}_1,\,-\boldsymbol{\alpha}_2-(k-1)\boldsymbol{\alpha}_1}. \tag{5.1.32}$$

Here it is convenient to introduce

$$\mathcal{F}(k) \stackrel{def}{=} N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2+k\boldsymbol{\alpha}_1} N_{-\boldsymbol{\alpha}_1,\,-\boldsymbol{\alpha}_2-k\boldsymbol{\alpha}_1} \tag{5.1.33}$$

which makes that (5.1.32) now becomes

$$a^i_{\alpha_1}(\boldsymbol{\alpha}_2 + k\boldsymbol{\alpha}_1)_i = \mathcal{F}(k) - \mathcal{F}(k-1). \tag{5.1.34}$$

To solve this recursion equation we start with the premise that $\mathcal{F}(q) = 0$. This will give us

$$\mathcal{F}(q-1) = -a^i_{\alpha_1}(\boldsymbol{\alpha}_2 + q\boldsymbol{\alpha}_1)_i. \tag{5.1.35}$$

If we continue in this way we obtain

$$\begin{aligned}
\mathcal{F}(k) &= -a^i_{\alpha_1}\left((q-k)\boldsymbol{\alpha}_2 + \sum_{i=0}^{q-k-1}(q-i)\boldsymbol{\alpha}_1\right)_i = \\
&= -a^i_{\alpha_1}\left((q-k)\boldsymbol{\alpha}_2 + \left((q-k)q - (q-k)\frac{(q-k-1)}{2}\right)\boldsymbol{\alpha}_1\right)_i = \\
&= a^i_{\alpha_1}(k-q)\left(\boldsymbol{\alpha}_2 + \left(\frac{q+k+1}{2}\right)\boldsymbol{\alpha}_1\right)_i.
\end{aligned} \tag{5.1.36}$$

Once again with the help of (5.1.29) and the fact that $N_{-\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2-p\boldsymbol{\alpha}_1} = 0$ we obtain

$$\begin{aligned}
&N_{-\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2-p\boldsymbol{\alpha}_1} N_{\boldsymbol{\alpha}_1,\,-\boldsymbol{\alpha}_2+p\boldsymbol{\alpha}_1} = N_{-\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2-(p+1)\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_1} N_{\boldsymbol{\alpha}_1,\,-\boldsymbol{\alpha}_2+(p+1)\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_1} = \\
&= N_{-\boldsymbol{\alpha}_2+p\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_1} N_{\boldsymbol{\alpha}_2-(p+1)\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_1} = N_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2-(p+1)\boldsymbol{\alpha}_1} N_{-\boldsymbol{\alpha}_1,\,-\boldsymbol{\alpha}_2+(p+1)\boldsymbol{\alpha}_1} = 0 \Rightarrow \\
&\mathcal{F}(-p-1) = 0.
\end{aligned} \tag{5.1.37}$$

By inserting this into (5.1.36) we obtain

$$\begin{aligned}
&a^i_{\alpha_1}(p+q+1)\left(\boldsymbol{\alpha}_2 + \frac{q-p}{2}\boldsymbol{\alpha}_1\right)_i = 0 \quad \Rightarrow \\
&a^i_{\alpha_1}\left(\boldsymbol{\alpha}_2 + \frac{q-p}{2}\boldsymbol{\alpha}_1\right)_i = 0.
\end{aligned} \tag{5.1.38}$$

Recall, both p and q are positive integers. If we rearrange (5.1.38) we get

$$p - q = 2\frac{a^i_{\alpha_1}(\boldsymbol{\alpha}_2)_i}{a^i_{\alpha_1}(\boldsymbol{\alpha}_1)_i}, \tag{5.1.39}$$

which is the result we were looking after[3]. Finally, set $k = 0$ in (5.1.36) and use (5.1.22),

$$\begin{aligned} N^2_{\boldsymbol{\alpha}_1,\,\boldsymbol{\alpha}_2} &= a^i q \left(\boldsymbol{\alpha}_2 + \left(\frac{q+1}{2}\right)\boldsymbol{\alpha}_1\right)_i = q\left(a^i(\boldsymbol{\alpha}_2)_i + \left(\frac{q+1}{2}\right)a^i(\boldsymbol{\alpha}_1)_i\right) = \\ &\overset{(5.1.39)}{=} q\left(\left(\frac{p-q}{2}\right)a^i(\boldsymbol{\alpha}_1)_i + \left(\frac{q+1}{2}\right)a^i(\boldsymbol{\alpha}_1)_i\right) = \frac{q}{2}(p+1)a^i(\boldsymbol{\alpha}_1)_i. \end{aligned} \tag{5.1.40}$$

The above calculations for p and q are valid under the assumption that $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_1$ are two roots. Suppose that $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are two *simple* roots and recall that the difference between two simple roots is not a root. This means that $p = 0$ for every commutator of the form (5.1.20) whenever the two root vectors corresponds to simple roots. It is not true that $p = 0$ in general. For example; let $\boldsymbol{\alpha_1}$ and $\boldsymbol{\alpha_1}$ be two simple roots and take

$$[e_{\boldsymbol{\alpha_1}}, e_{\boldsymbol{\alpha_2}}] = N_{\boldsymbol{\alpha_1},\,\boldsymbol{\alpha_2}} e_{\boldsymbol{\alpha_1}+\boldsymbol{\alpha_2}} \tag{5.1.41}$$

and then

$$[f_{\boldsymbol{\alpha_1}}, e_{\boldsymbol{\alpha_1}+\boldsymbol{\alpha_2}}] = N_{-\boldsymbol{\alpha_1},\,\boldsymbol{\alpha_1}+\boldsymbol{\alpha_2}} e_{\boldsymbol{\alpha_2}}. \tag{5.1.42}$$

Thus p can not be equal to zero for the commutator between $f_{\boldsymbol{\alpha_1}}$ and $e_{\boldsymbol{\alpha_1}+\boldsymbol{\alpha_2}}$. Observe that $\boldsymbol{\alpha_1} + \boldsymbol{\alpha_2}$ is not a simple root.

5.1.1 Cartan-Killing form

As mentioned in the begining of the chapter the roots can be seen as elements α_j in the dual space $\mathfrak{C}^*$. This means that (5.1.4) becomes

$$[h_i, t_j] = \alpha_j(h_i)t_j \tag{5.1.43}$$

and we will use this rather than $\boldsymbol{\alpha}$ from now on. The only difference is that we think of the components, $(\boldsymbol{\alpha}_j)_i = \alpha_j(h_i)$, as the result of the mapping $\alpha_j : \mathfrak{C} \longrightarrow \mathbb{R}$ and we can therefore think of α as a vector in $\mathfrak{C}^*$.

The bilinear symmetric **Cartan-Killing form** is defined as

$$(a, b) \overset{def}{=} \mathrm{Tr}\ (\mathrm{ad}_a \mathrm{ad}_b), \quad a, b \in \mathfrak{g}, \tag{5.1.44}$$

which is a representation independent definition. We will see in (5.1.57) how we can get an expression for (a, b) when $a, b \in \mathfrak{C}$. In section 5.3, we will explicitly obtain the matrices for the adjoint representation. Bilinear and symmetric implies that

$$\begin{aligned} &(a, b) = (b, a) \qquad a, b \in \mathfrak{g} \\ &(a, b + c) = (a, b) + (a, c), \quad a, b, c \in \mathfrak{g} \\ &(\lambda a, b) = (a, \lambda b) = \lambda(a, b) \quad \lambda \in \mathbb{C}. \end{aligned} \tag{5.1.45}$$

It seems that the Cartan-Killing form is some sort of scalar product on the Lie algebra but that is not correct to be precise. One may for example find an element $a \in \mathfrak{g}$ such that $(a, a) < 0$. But there is a scalar product on the root space Σ as in (5.1.26) denoted $\langle \cdot, \cdot \rangle$.

[3] This result will be presented in section 5.1.1 in a different notation, see (5.1.50).

What we now want to do is express this scalar product between elements in Σ in terms of the Cartan-Killing form.

There is a connection between $\mathfrak{C}$ and $\mathfrak{C}^*$ for most of the semi-simple Lie algebras, e.g. the Lie algebras we have mentioned in the beginning of the chapter, see ([3], p. 19). For any root α_i there exists a unique element $h_{\alpha_i} \in \mathfrak{C}$ such that

$$\alpha_i(h') = (h_{\alpha_i}, h') \tag{5.1.46}$$

for every element $h' \in \mathfrak{C}$. Or, since every root vector t_i in $\mathfrak{g}$ has a root we may say that there is an element h_{α_i} associated to t_i through (5.1.46). To find h_{α_i} given α_i we start with the right hand side of (5.1.46) by expanding h' in terms of the basis $\{h_j\}$ of $\mathfrak{C}$, i.e. $h' = \sum_{j=1}^{r} b^j h_j$ and in the same manner, $h_{\alpha_i} = \sum_{l=1}^{r} c^l h_l$, where c^l are the coefficients that we are to determine. By replacing h' on the left hand side of (5.1.46) we obtain

$$\sum_{j=1}^{r} b^j \alpha_i(h_j) = \sum_{j=1}^{r}\sum_{l=1}^{r} b^j c^l (h_l, h_j), \tag{5.1.47}$$

and since (5.1.46) is valid for every h' the coefficients with respect to b^j must be equal, that is

$$\alpha_i(h_j) = \sum_{l=1}^{r} c^l (h_l, h_j). \tag{5.1.48}$$

This will give us r equations for every α_i. We will see this explicitly in section 5.3. With this connection between the Lie algebra and the root space we can restate the definition of the scalar product on the root space as

$$\langle \alpha_1, \alpha_2 \rangle \overset{def}{=} (h_{\alpha_1}, h_{\alpha_2}), \tag{5.1.49}$$

where $h_{\alpha_1}, h_{\alpha_2}$ are the corresponding Cartan elements according to (5.1.46).

With this definition we can express (5.1.39) as

$$p - q = 2\frac{\langle \alpha_1, \alpha_2 \rangle}{\langle \alpha_1, \alpha_1 \rangle}. \tag{5.1.50}$$

Another property of the Cartan-Killing form in addition to bilinearity and symmetry is **invariance**

$$(a, [b, c]) = ([a, b], c), \quad a, b, c \in \mathfrak{g} \tag{5.1.51}$$

To see this, use the invariance of the trace under cyclic permutations and (3.1.3),

$$\begin{aligned}(a, [b, c]) = \mathrm{Tr}(\mathrm{ad}_a \mathrm{ad}_{[b,c]}) = \mathrm{Tr}(\mathrm{ad}_a([\mathrm{ad}_b, \mathrm{ad}_c])) = \mathrm{Tr}(\mathrm{ad}_a\mathrm{ad}_b\mathrm{ad}_c - \mathrm{ad}_a\mathrm{ad}_c\mathrm{ad}_b) = \\ = \mathrm{Tr}([\mathrm{ad}_a, \mathrm{ad}_b]\mathrm{ad}_c - \mathrm{ad}_a\mathrm{ad}_c\mathrm{ad}_b + \mathrm{ad}_b\mathrm{ad}_a\mathrm{ad}_c) = \mathrm{Tr}([\mathrm{ad}_a, \mathrm{ad}_b]\mathrm{ad}_c) = ([a, b], c).\end{aligned} \tag{5.1.52}$$

We can use this to get an expression for the coefficients $a^i_{\alpha_j}$ in (5.1.14). Let $h \in \mathfrak{C}$ and $e_j \in \eta_+$, $f_j \in \eta_-$. Consider

$$\begin{aligned}([e_j, f_j], h) \overset{(5.1.51)}{=} (e_j, [f_j, h]) = \alpha_j(h)(e_j, f_j) \Rightarrow \\ (\frac{[e_j, f_j]}{(e_j, f_j)}, h) = \alpha_j(h)\end{aligned} \tag{5.1.53}$$

and compare this to (5.1.46). This will give us

$$\frac{[e_j, f_j]}{(e_j, f_j)} = h_{\alpha_j} \quad \Rightarrow \quad [e_j, f_j] = (e_j, f_j) h_{\alpha_j}, \tag{5.1.54}$$

i.e. $a^i_{\alpha_j} h_i = (e_j, f_j) h_{\alpha_j}$.

We also have something similar to orthogonal subspaces. If $\mathfrak{h}$ and $\mathfrak{i}$ are two ideals to a Lie algebra $\mathfrak{g}$ such that $\mathfrak{h} \cap \mathfrak{i} = \{0\}$ then,

$$(a, b) = 0 \quad \forall a \in \mathfrak{h}, \forall b \in \mathfrak{i}. \tag{5.1.55}$$

This follows from the invariance (5.1.51). Let w be any element in $\mathfrak{g}$ and note that $[a, b] = 0$.

$$\begin{aligned} (\mathrm{ad}_a \mathrm{ad}_b) w = \mathrm{ad}_a [b, w] = [a, [b, w]] = [[a, b], w] = 0 \\ /w \text{ is arbitrary}/ \Rightarrow \\ \mathrm{ad}_a \mathrm{ad}_b = 0 \end{aligned} \tag{5.1.56}$$

The matrix representation for ad_{h_i} can be obtained once we have fixed a basis in $\mathfrak{g}$, see (3.1.21). Let this basis be $h_1, h_2, ..., h_r, e_1, e_2, ..., f_1, f_2, ...$, i.e. the one we have used before. The matrix representation for $\mathrm{ad}(h_i)$ becomes

$$\mathrm{ad}(h_i) = \begin{pmatrix} 0 & & & & & & \\ & \ddots & & & & & \\ & & \alpha_1(h_i) & & & & \\ & & & \ddots & & & \\ & & & & -\alpha_1(h_i) & & \\ & & & & & \ddots & \end{pmatrix}. \tag{5.1.57}$$

This means that (h_i, h_j) can be calculated as

$$(h_i, h_j) = \sum_{k \in \Sigma} \alpha_k(h_i) \alpha_k(h_j). \tag{5.1.58}$$

5.1.2 Constructing New Roots

The procedure to form new roots is to take linear combinations of our simple roots. Let us start with two simple roots α_1 and α_2. The first step is to form $\alpha_1 + \alpha_2$ and ask ourselves if this is a root. Consider the two integers p, q from (5.1.50). We know that $p = 0$ since the difference between two simple roots is not a root. So, if $\alpha_1 + \alpha_2$ should be a root then

$$\begin{aligned} -q = 2\frac{\langle \alpha_1, \alpha_2 \rangle}{\langle \alpha_2, \alpha_2 \rangle} < 0 \Rightarrow \\ \langle \alpha_1, \alpha_2 \rangle < 0. \end{aligned} \tag{5.1.59}$$

Suppose that $q = k$. We continue to do add α_2 to α_1 and obtain after k times the root $\alpha_j = \alpha_1 + k\alpha_2$. To see whether $\alpha_j + \alpha_2$ is a root or not we have to check if

$$q = p - 2\frac{\langle \alpha_j, \alpha_2 \rangle}{\langle \alpha_2, \alpha_2 \rangle} > 0. \tag{5.1.60}$$

For α_j, p is known since we just can go back to were we started our "chain" of roots.

Given two roots α_1, α_2 and a scalar product we should be able to calculate the angle between α_1, α_2. Take p_1, q_1 and p_2, q_2 to be the corresponding integers to α_1, α_2, respectively, according to (5.1.19) and (5.1.20), i.e.

$$\begin{aligned} p_1 - q_1 &= 2\frac{\langle \alpha_2, \alpha_1 \rangle}{\langle \alpha_2, \alpha_2 \rangle} = n_1 \\ p_2 - q_2 &= 2\frac{\langle \alpha_1, \alpha_2 \rangle}{\langle \alpha_1, \alpha_1 \rangle} = n_2. \end{aligned} \tag{5.1.61}$$

n_1, n_2 are just for convenience. Consider

$$\begin{aligned} n_1 n_2 &= 4\frac{\langle \alpha_2, \alpha_1 \rangle \langle \alpha_1, \alpha_2 \rangle}{\langle \alpha_2, \alpha_2 \rangle \langle \alpha_1, \alpha_1 \rangle} \Rightarrow \\ \langle \alpha_2, \alpha_1 \rangle^2 &= \frac{1}{4} n_1 n_2 \langle \alpha_1, \alpha_1 \rangle \langle \alpha_2, \alpha_2 \rangle \end{aligned} \tag{5.1.62}$$

which gives us

$$\cos^2(\theta) = \frac{1}{4} n_1 n_2 \tag{5.1.63}$$

where θ is the angle between the roots. The ratio

$$\frac{n_1}{n_2} = \frac{\langle \alpha_2, \alpha_2 \rangle}{\langle \alpha_1, \alpha_1 \rangle} \geq 0 \tag{5.1.64}$$

tells us that n_1 and n_2 must have the same sign. We can use Cauchy-Schwartz inequality in (5.1.62),

$$\langle \alpha_1, \alpha_2 \rangle \leq (\langle \alpha_1, \alpha_1 \rangle)^{1/2} (\langle \alpha_2, \alpha_2 \rangle)^{1/2} \tag{5.1.65}$$

to get

$$n_1 n_2 \leq 4. \tag{5.1.66}$$

The equality is valid when α_1 is proportional to α_2, i.e. if $\alpha_1 = \pm\alpha_2$ but this will not give us any new roots. This means that

$$|n_{1,2}| = 0, 1, 2, 3. \tag{5.1.67}$$

since $n_{1,2}$ are integers with the same sign. Note that $n_1 = 0 \Leftrightarrow n_2 = 0$. One important conclusion of this comes from when we consider the case $p = 0$, $q = -n$ for some α. This implies that we can at most form a chain composed of four roots.

5.1.3 Weights

Instead of looking at the adjoint representation of $\mathfrak{g}$, let us look at an arbitrary representation π with module V of finite dimension n. Since a representation must preserve the commuting rules we still have that

$$\begin{aligned}
&[\pi(h_i), \pi(e_j)] = \beta_i^{(j)} \pi(e_j) \\
&[\pi(h_i), \pi(f_j)] = -\beta_i^{(j)} \pi(f_j) \\
&[\pi(e_k), \pi(e_l)] = N_{k,l} \pi(e_{k+l}) \\
&[\pi(h_i), [\pi(e_k), \pi(e_l)]] = (\beta_i^{(k)} + \beta_i^{(l)})[\pi(e_k), \pi(e_l)] \quad \text{if } (\boldsymbol{\alpha}_k + \boldsymbol{\alpha}_l) \in \Sigma \\
&[\pi(h_i), [\pi(e_k), \pi(e_l)]] = 0, \quad \text{if } (\boldsymbol{\alpha}_k + \boldsymbol{\alpha}_l) \notin \Sigma \\
&[\pi(e_j), \pi(f_j)] = a^i \pi(h_i) \quad a^i \in \mathbb{R} \\
&[\pi(e_j), \pi(f_k)] = 0 \quad \text{if } \boldsymbol{\alpha}_j, \boldsymbol{\alpha}_k \text{ are simple roots.}
\end{aligned} \tag{5.1.68}$$

Let $\{\psi_1, \psi_2, ..., \psi_n\}$ be a basis for V with the property

$$\pi(h_i)\psi_j = \lambda_i^{(j)} \psi_j, \quad j = 1, 2, ..., n. \tag{5.1.69}$$

That is, we choose the set of eigenvectors as basis. In analogy to what we have done for the adjoint representation we can associate a vector, $(\lambda_1^{(j)}, \lambda_2^{(j)}, ..., \lambda_n^{(j)})$ to every eigenvector ψ_j. These vectors are called **weights**. Thus, *the roots are the weights of the adjoint representation.* In the adjoint representation our eigenvectors are the e_{α_i}'s and f_{α_i}'s and the eigenvalue equation is of course $[h_i, e_{\alpha_i}] = \alpha_i(h_i) e_{\alpha_i}$. The weights can also be expressed as

$$\pi(h_i)\psi_j = M_j(h_i)\psi_j, \quad j = 1, 2, ..., n \tag{5.1.70}$$

where $M_j(h_i)$[4] belongs to the dual space $\mathfrak{C}^*$, just like for the adjoint representation. That is,

$$M_j : \mathfrak{C} \longrightarrow \mathbb{R}. \tag{5.1.71}$$

Since $M_j \in \mathfrak{C}^*$ we can use $\langle \cdot, \cdot \rangle$ as scalar product.

Let us now investigate the connection between roots and weights a bit more. Consider

$$\pi(h_i)\psi_j = M_j(h_i)\psi_j. \tag{5.1.72}$$

Now, take $\pi(e_{\alpha_k})\psi_j$, $e_{\alpha_k} \in \eta_+$, and let $\pi(h_i)$ act on it,

$$\begin{aligned}
&\pi(h_i)\pi(e_{\alpha_k})\psi_j = ([\pi(h_i), \pi(e_{\alpha_k})] + \pi(e_{\alpha_k})\pi(h_i))\psi_j = \\
&\overset{(5.1.43)}{=} \alpha_k(h_i)\pi(e_{\alpha_k})\psi_j + \pi(e_{\alpha_k})M_j(h_i)\psi_j = \\
&= (\alpha_k(h_i) + M_j(h_i))\pi(e_{\alpha_k})\psi_j.
\end{aligned} \tag{5.1.73}$$

This means that $\pi(e_{\alpha_j})\psi_j$ is a new eigenvector to $\pi(h_i)$ with eigenvalue $\alpha_k(h_i) + M_j(h_i)$, i.e. "the old eigenvalue plus the root component corresponding to e_{α_k}". What is interesting here is the term $[\pi(h_i), \pi(e_{\alpha_k})]$ because it gives us the root $\alpha_k(h_i)$ added to our former eigenvalue $M_j(h_i)$. This seems very familiar from section 4.4 where our roots had only one component,

[4] Observe that there is no summation over j above.

namely $\pm\hbar$, see (4.4.6a) and (4.4.8) What we have done so far is just a generalization of section 4.4. We can do the exact same calculations for $\pi(f_{\alpha_k})\psi_j$ as we did for $\pi(e_{\alpha_k})\psi_j$. The only difference is that we get $-\alpha_k(h_i) + M_j(h_i)$ for eigenvalue. Thus, the operators $\pi(e_{\alpha_k})$ and $\pi(f_{\alpha_k})$ are the ladder operators, hence the subscripts + and − in η_+, η_-. In section 4.4, where we did this explicitly for the $\mathfrak{sl}(2,\mathbb{C})$ algebra, the underlying theory behind it was not so obvious. The reason for this is that the dimension of the $\mathfrak{sl}(2,\mathbb{C})$ algebra is too small for the theory to "reveal itself". The operator we called $\hat{J}_z$ span the Cartan subalgebra and our step operators $\hat{J}_\pm$ corresponds to e and f. That is, a three-dimensional algebra. To summarize, we have seen that in general we can have several "step up" and "step down" operators together with a set of "$\hat{J}_z$'s" that commutes.

In analogy with (5.1.19) we can construct a so called "chain" of weights through

$$\pi(h_i)\pi(e_\alpha)^k\psi_j = (M_j(h_i) + k\alpha(h_i))\pi(e_\alpha)^k\psi_j \tag{5.1.74}$$

and

$$\pi(h_i)\pi(f_\alpha)^k\psi_j = (M_j(h_i) - n\alpha(h_i))\pi(f_\alpha)^k\psi_j, \tag{5.1.75}$$

where we have dropped the k in $e_{\alpha_k}, f_{\alpha_k}$. In the adjoint representation, the $\pi(e_\alpha)^k\psi_j$'s corresponded to (5.1.18) and since V is a finite dimensional vector space there must be two positive integers m, n for every j, such that

$$\begin{aligned}\pi(e_\alpha)^{m+1}\psi_j = 0\\ \pi(f_\alpha)^{n+1}\psi_j = 0\end{aligned} \tag{5.1.76}$$

We are now searching for an expression such as (5.1.50) for the two positive integers m, n. Let us start with an eigenvector ψ_0, such that

$$\pi(e_\alpha)\psi_0 = 0 \tag{5.1.77}$$

and let M_0 be its weight. Moreover, we define

$$\pi(f_\alpha)^k\psi_0 \overset{def}{=} \psi_{-k\alpha}. \tag{5.1.78}$$

α is any positive root and we label our eigenvectors with their weight as $\psi_{-k\alpha}$, i.e. $\psi_{-k\alpha}$ has weight $M_{-k\alpha} = M_0(h_i) - k\alpha(h_i)$. We can express (5.1.46) in terms of weights

$$M_i(h') = (h_{M_i}, h'), \tag{5.1.79}$$

remember; roots *are* weights. We will use this later. If we consider $\pi(h_i)\pi(e_\alpha)\psi_{-k\alpha}$ we see that

$$\pi(e_\alpha)\psi_{-k\alpha} \propto \psi_{-(k-1)\alpha}. \tag{5.1.80}$$

Two eigenvectors with the same eigenvalue differs by a constant, i.e.

$$\pi(e_\alpha)\psi_{-k\alpha} = r_k\psi_{-(k-1)\alpha}. \tag{5.1.81}$$

This constant can be determined if we use the assumption that $r_0 = 0$, see (5.1.77). For $\pi(f_\alpha)$, we have that $r = 1$ as a consequence of (5.1.78). Consider

$$\begin{aligned}\pi(e_\alpha)\psi_{-k\alpha} = r_k\psi_{-(k-1)\alpha} = \pi(e_\alpha)^k\pi(f_\alpha)^k\psi_{-(k-1)\alpha} =\\ = ([\pi(e_\alpha), \pi(f_\alpha)] + \pi(f_\alpha)\pi(e_\alpha))\psi_{-(k-1)\alpha},\end{aligned} \tag{5.1.82}$$

normalize such that

$$[\pi(e_\alpha), \pi(f_\alpha)] \overset{(5.1.54)}{=} (e_\alpha, f_\alpha)\pi(h_\alpha) = \pi(h_\alpha). \tag{5.1.83}$$

Equation (5.1.82) becomes

$$\begin{aligned} r_k\psi_{-(k-1)\alpha} &= (\pi(h_\alpha) + r_{k-1})\psi_{-(k-1)\alpha} = (M_{-(k-1)\alpha}(h_\alpha) + r_{k-1})\psi_{-(k-1)\alpha} = \\ &= (M_0(h_\alpha) - (k-1)\alpha(h_\alpha) + r_{k-1})\psi_{-(k-1)\alpha} \end{aligned} \tag{5.1.84}$$

so we need to solve the recursion equation

$$r_k = M_0(h_\alpha) - (k-1)\alpha(h_\alpha) + r_{k-1}, \quad r_0 = 0. \tag{5.1.85}$$

The solution is

$$\begin{aligned} &r_1 = M_0(h_\alpha) \\ &r_2 = 2M_0(h_\alpha) - \alpha(h_\alpha) \\ &\vdots \\ &r_k = kM_0(h_\alpha) - \sum_{l=1}^{k}(l-1)\alpha(h_\alpha) = kM_0(h_\alpha) - \frac{\alpha(h_\alpha)}{2}k(k-1). \end{aligned} \tag{5.1.86}$$

If we use $M_0(h_\alpha) = (h_0, h_\alpha) = \langle M_0, \alpha\rangle$ and $\alpha(h_\alpha) = (h_\alpha, h_\alpha) = \langle \alpha, \alpha\rangle$ we obtain

$$r_k = k\langle M_0, \alpha\rangle - \frac{\langle \alpha, \alpha\rangle k(k-1)}{2}. \tag{5.1.87}$$

In (5.1.86) we used that we have an upper limit of the chain of eigenvectors but we must also have a lower limit, i.e. $\pi(f_\alpha)\psi_n = 0$. To get the constraint on r, consider $\pi(e_\alpha)\pi(f_\alpha)\psi_n = 0$, which means that $r_{n+1} = 0$. This second "boundary condition" (5.1.87) yields

$$\begin{aligned} &(n+1)\langle M_0, \alpha\rangle - \frac{\langle \alpha, \alpha\rangle(n+1)n}{2} = 0 \Rightarrow \\ &n = \frac{2\langle M_0, \alpha\rangle - \langle \alpha, \alpha\rangle}{2\langle \alpha, \alpha\rangle} \pm \frac{\langle \alpha, \alpha\rangle + 2\langle M_0, \alpha\rangle}{2\langle \alpha, \alpha\rangle}. \end{aligned} \tag{5.1.88}$$

Since n is positive

$$n = \frac{2\langle M_0, \alpha\rangle}{\langle \alpha, \alpha\rangle}. \tag{5.1.89}$$

The weight M_0 is called **the highest weight** which we actually have seen in section 4.4 as the spin j. Note however that n in general will not be equal to the dimension of the irreducible representations since there might be several step operators.

To summarize: we started with the eigenvector ψ_0, called the **highest weight state**, satisfying (5.1.77), i.e. it is annihilated by any element from η_+. In analogy for the adjoint representation we wanted to construct chains of eigenvectors so that we can form the module of the representation. Since the module is of finite dimension we concluded that these chains can not be infinite and we obtained an expression for the length of the chains. Suppose that we had started with ψ_j instead of ψ_0, i.e. a state in the "middle of the chain". For this state

$m \neq 0$ unlike for ψ_0. But, the total length of the chain is the same whether we start from ψ_j or ψ_0 , so we can express (5.1.88) as

$$\begin{aligned} n+m &= \frac{2\langle M_0, \alpha\rangle}{\langle \alpha, \alpha\rangle} = \frac{2\langle M_j + m\alpha, \alpha\rangle}{\langle \alpha, \alpha\rangle} = \frac{2\langle M_j, \alpha\rangle}{\langle \alpha, \alpha\rangle} + 2m \Rightarrow \\ n-m &= \frac{2\langle M_j, \alpha\rangle}{\langle \alpha, \alpha\rangle}. \end{aligned} \tag{5.1.90}$$

Not very unexpectedly, this is exactly the same as (5.1.50). Note that the n in (5.1.90) is not the same n as in (5.1.88).

This is a good place for an example. Let us find the dimension of an irreducible representation of $\mathfrak{sl}(2,\mathbb{C})$. This is exactly what we did in section 4.4, but now we will derive it in a different way. Just to be clear, we want to find the n in (5.1.88). From section 4.4 we have the commuting relations

$$\begin{aligned} [\hat{J}_z, \hat{J}_\pm] &= \pm\hbar \hat{J}_\pm \\ [\hat{J}_+, \hat{J}_-] &= 2\hat{J}_z, \end{aligned} \tag{5.1.91}$$

i.e. $\mathfrak{sl}(2,\mathbb{C})$ is three dimensional. The two roots $\alpha(\hat{J}_z) = \pm\hbar$ has only one component each due to the fact that $\hat{J}_z$ spans the one-dimensional Cartan subalgebra. As highest weight we take $M_0 = M_j = j\hbar$, see (4.4.22), with the corresponding highest state ψ_j. To calculate

$$n = 2\frac{\langle M_j(\hat{J}_z), \alpha(\hat{J}_z)\rangle}{\langle \alpha(\hat{J}_z), \alpha(\hat{J}_z)\rangle} = 2\frac{(h_j, h_\alpha)}{(h_\alpha, h_\alpha)}, \tag{5.1.92}$$

where h_j, h_α are the elements associated to $M_j(\hat{J}_z), \alpha(\hat{J}_z)$ according to (5.1.79), we need to determine (h_j, h_α) and (h_α, h_α). From (5.1.48), the equations to determine h_j, h_α, we obtain for h_α

$$\begin{aligned} \hbar &= c(\hat{J}_z, \hat{J}_z) \Rightarrow \\ h_\alpha &= \frac{\hbar}{(\hat{J}_z, \hat{J}_z)}\hat{J}_z \end{aligned} \tag{5.1.93}$$

and for h_j we obtain

$$\begin{aligned} j\hbar &= c(\hat{J}_z, \hat{J}_z) \Rightarrow \\ h_j &= \frac{j\hbar}{(\hat{J}_z, \hat{J}_z)}\hat{J}_z. \end{aligned} \tag{5.1.94}$$

We can use (5.1.57) to determine $(\hat{J}_z, \hat{J}_z)$, but it is not necessary for determining n.

$$(\hat{J}_z, \hat{J}_z) = \sum_{k=1}^{2} (\alpha_k(\hat{J}_z))^2 = 2\hbar^2. \tag{5.1.95}$$

Finally, (5.1.92) becomes

$$n = 2\frac{(\frac{j\hbar}{(2\hbar^2)}\hat{J}_z, \frac{\hbar}{(2\hbar^2)}\hat{J}_z)}{\frac{\hbar^2}{(2\hbar^2)^2}(\hat{J}_z, \hat{J}_z)} = 2j. \tag{5.1.96}$$

which is in line with the result from section 4.4. Since $\hat{J}_-^k\psi_j$ is an eigenvector to $\hat{J}_z$ for every $0 \leq k \leq n$ this set can be chosen as a basis for the representation with dimension $2j+1$.

For $\mathfrak{sl}(2,\mathbb{C})$ we only have one chain of eigenvectors constituting the module since we only have two ladder operators. In general, however, the module consist of several chains. The principle is that you start with a highest weight state and then step downward with the step operators in η_- to form an irreducible representation. The procedure is very similar to what we did with the roots. For every weight M_j we check whether $M_j - \alpha_i$ and $M_j + \alpha_i$ are new roots or not and proceed in this way. The integers n_j, i.e.

$$n_j = \frac{2\langle M_0, \alpha_j\rangle}{\langle \alpha_j, \alpha_j\rangle}, \tag{5.1.97}$$

can be used to label our irreducible representations. For a rank r algebra, $j = 1, 2, ..., r$ corresponding to the r simple roots. Since the roots to the $\mathfrak{sl}(2,\mathbb{C})$ algebra are $\pm\hbar$, we will step down from the highest weight state, ψ_j, and subtract $\hbar$ to the weights in every step. This procedure can be illustrated in a weight-diagram such as figure 5.1

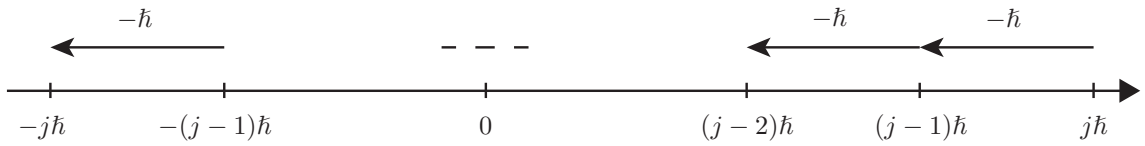


Figure 5.1: Weight diagram for a $\mathfrak{sl}(2,\mathbb{C})$ representation.

5.2 Chevally-Serre Presentation

To get a better grip of how to construct Lie algebras one can make use of a known algebra. The basic idea is to "make copies" of $\mathfrak{sl}(2,\mathbb{C})$ and use these as a basis for the new algebra. In this section the Cartan matrix will be introduced but otherwise nothing new. Rather a different way to look at the structure of a Lie algebra. First recall the $\mathfrak{sl}(2,\mathbb{C})$ algebra:

$$\begin{aligned} [e,f] &= h \\ [h,e] &= 2e \\ [h,f] &= -2f, \end{aligned} \tag{5.2.1}$$

i.e. $\mathfrak{sl}(2,\mathbb{C}) = \mathbb{C}f \oplus \mathbb{C}h \oplus \mathbb{C}e$. Obviously $h \in \mathfrak{C}$, $e \in \eta_+$ and $f \in \eta_-$. We also have the fundamental representation

$$\pi(h) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \pi(e) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \pi(f) = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}. \tag{5.2.2}$$

We proceed as follows. Take a set of r $\mathfrak{sl}(2,\mathbb{C})$ algebras and denote the elements as

$$\{f_i, h_i, e_i\}, \quad i = 1, 2, 3, ..., r \tag{5.2.3}$$

With this set of r $\mathfrak{sl}(2,\mathbb{C})$'s we can form a direct sum of vector spaces, not Lie algebras, as

$$\bigoplus_{i=1}^{r} \mathfrak{sl}(2,\mathbb{C})_i \tag{5.2.4}$$

which means that our new algebra has dimension $3r$ so far. Now we have to intertwine the $\mathfrak{sl}(2,\mathbb{C})$'s with each other through commutations. The $\mathfrak{sl}(2,\mathbb{C})$'s can not all commute since then we would just get a trivial algebra of dimension $3r$. New elements in an algebra can only be obtained through commutations. Let

$$\begin{aligned} [h_i, h_j] &= 0 \\ [h_i, e_j] &= A_{ji} e_j \\ [h_i, f_j] &= -A_{ji} f_j \\ [e_i, f_j] &= \delta_{ij} h_{ij}. \end{aligned} \tag{5.2.5}$$

Here we do *not* apply the summation convention. Clearly, h_i forms the Cartan algebra for our new algebra, $e_j \in \eta_+$, $f_j \in \eta_-$ and new elements can be constructed through commutations. Observe that the representation given in (5.2.2) do not apply to all of the $\mathfrak{sl}(2,\mathbb{C})$ when they are added. The matrix A_{ji} is called the **Cartan matrix**. This is an $r \times r$ matrix that contains all the information about the Lie algebra and we have actually been in contact with it in the previous section because

$$A_{ji} \overset{def}{=} 2\frac{\langle \alpha_i, \alpha_j \rangle}{\langle \alpha_j, \alpha_j \rangle}. \tag{5.2.6}$$

At first glance, the connection between (5.2.6) and (5.2.5) is not obvious except for the case $A_{ii} = 2$. If we take a simple case when $r = 2$ we have two $\mathfrak{sl}(2,\mathbb{C})$'s. Consider

$$[h_i, e_j] = \alpha_j(h_i)e_j = (h_{\alpha_j}, h_i)e_j = \langle \alpha_j, \alpha_i \rangle e_j. \tag{5.2.7}$$

Our rank two algebra consist so far of two Cartan elements and four ladder operators. A fifth ladder operator can be constructed $[e_1, e_2] = e_3$ and suppose that $[e_1, e_3] = 0$. Let us calculate $\langle \alpha_i, \alpha_i \rangle$, $i = 1, 2$. To do this we need to find A_{12}, A_{21}. Jacobi gives

$$\begin{aligned} [f_1,[e_1,e_3]] + [e_1,[e_3,f_1]] + [e_3,[f_1,e_1]] &= [e_1,[e_3,f_1]] + [e_3,h_1] = \\ = [e_1,[[e_1,e_2],f_1]] + [e_3,h_1] &= [e_1,[e_2,[f_1,e_1]]] + [e_3,h_1] = \\ = [e_1,[e_2,h_1]] + [e_3,h_1] &= -A_{21}e_3 - (A_{21}+A_{11})e_3 = \\ = -(2A_{21}+A_{11})e_3 &= 0 \end{aligned} \tag{5.2.8}$$

$$e_3 \neq 0 \Rightarrow A_{21} = -1.$$

Similar calculations gives $A_{12} = -1$ as well. Let α_1, α_2 be the roots to e_1 and e_2,

$$\langle \alpha_1, \alpha_1 \rangle = (\alpha_1)_i (\alpha_1)_j (g^{-1})^{ij} = (\alpha_1)^i (\alpha_1)^j g_{ij}$$

$$g_{ij} \overset{(5.1.26)}{=} \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \Rightarrow g^{ij} = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \tag{5.2.9}$$

$$\Rightarrow$$

$$\langle \alpha_1, \alpha_1 \rangle = \langle \alpha_2, \alpha_2 \rangle = 4\frac{2}{3} - 2\frac{1}{3} - 2\frac{1}{3} + \frac{2}{3} = 2.$$

Equation (5.2.7) can be expressed as

$$[h_i, e_j] = \langle \alpha_j, \alpha_i \rangle e_j = \frac{2}{\langle \alpha_j, \alpha_j \rangle} \langle \alpha_j, \alpha_i \rangle e_j = A_{ji} e_j. \tag{5.2.10}$$

So, for this example the Cartan matrix A becomes

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \tag{5.2.11}$$

which in fact is the Cartan matrix for $\mathfrak{sl}(3, \mathbb{C})$. This result can be generalized. A Cartan matrix A satisfies

$$\begin{aligned} &i)\ \ A_{ii} = 2\ \ \forall i \\ &ii)\ \ A_{ij} \in \mathbb{Z}_-\ \ \forall i \neq 0 \\ &iii)\ \ A_{ij} = 0 \Leftrightarrow A_{ji} = 0 \end{aligned} \tag{5.2.12}$$

If α_j are the roots to the e_j's, $j = 1, 2, 3, ..., r$ we can choose these as the simple roots to our algebra. As stated above (and in the previous sections), we can obtain more elements other than those of the $\mathfrak{sl}(2, \mathbb{C})$'s through commutations. Otherwise we would have to confine ourselves to a $3r$-dimensional algebra. Of course we have a constraint very similar to (5.1.50). Since α_j, α_i are simple roots $p = 0$. This means that if we would like to construct new elements by commuting e_j and e_i we have that, see (5.1.19),

$$\underbrace{[e_i, [e_i, [..., [e_i}_{q+1 \text{ times}}, e_j]...]]] = 0. \tag{5.2.13}$$

With

$$p - q = -q = A_{ji} \tag{5.2.14}$$

(compare to (5.1.50)) this can be expressed as

$$\begin{aligned} (\mathrm{ad}_{e_i})^{1+q}(e_j) = (\mathrm{ad}_{e_i})^{1-A_{ji}}(e_j) = 0 \\ (\mathrm{ad}_{f_i})^{1+q}(f_j) = (\mathrm{ad}_{f_i})^{1-A_{ji}}(f_j) = 0. \end{aligned} \tag{5.2.15}$$

These are known as the **Serre relations**. For the trivial case when we only have one $\mathfrak{sl}(2, \mathbb{C})$ $A_{ji} = 2$. That is, $1 - A_{ji} < 0$ which means that we do not get any new elements through commutation. Of course this is trivial since $[e, e] = 0$. The basis that we have worked with in this section is called the **Chevally basis**. With this basis, all the structure constants are integers. Suppose that we have a basis that is not the Chevally basis (the prime one) and we normalize our elements as

$$\begin{aligned} e'_{\alpha_i} &= \sqrt{\frac{2}{(e_{\alpha_i}, f_{\alpha_i})\langle \alpha_i, \alpha_i \rangle}} e_{\alpha_i} \\ f'_{\alpha_i} &= \sqrt{\frac{2}{(e_{\alpha_i}, f_{\alpha_i})\langle \alpha_i, \alpha_i \rangle}} f_{\alpha_i} \end{aligned} \tag{5.2.16}$$

we see that

$$[e'_{\alpha_i}, f'_{\alpha_i}] = \frac{2}{\langle \alpha_i, \alpha_i \rangle} h_{\alpha_i} \tag{5.2.17}$$

if we use (5.1.54). This makes it appropriate to define

$$h'_i = \frac{2}{\langle \alpha_i, \alpha_i \rangle} h_{\alpha_i} \tag{5.2.18}$$

such that

$$[e'_{\alpha_i}, f'_{\alpha_i}] = h'_i \tag{5.2.19}$$

and

$$\begin{aligned}[h'_i, e'_{\alpha_j}] &= \frac{2}{\langle \alpha_i, \alpha_i \rangle} \sqrt{\frac{2}{(e_{\alpha_j}, f_{\alpha_j}) \langle \alpha_j, \alpha_j \rangle}} [h_{\alpha_i}, e_{\alpha_j}] = \\ &= \frac{2}{\langle \alpha_i, \alpha_i \rangle} \sqrt{\frac{2}{(e_{\alpha_j}, f_{\alpha_j}) \langle \alpha_j, \alpha_j \rangle}} \alpha_j(h_{\alpha_i}) e_{\alpha_j} = \\ &= \frac{2}{\langle \alpha_i, \alpha_i \rangle} (h_{\alpha_j}, h_{\alpha_i}) e'_{\alpha_j} = \frac{2\langle \alpha_j, \alpha_i \rangle}{\langle \alpha_i, \alpha_i \rangle} e'_{\alpha_j} = A_{ji} e'_{\alpha_j}.\end{aligned} \tag{5.2.20}$$

Thus, we have the Chevally basis.

5.3 $\mathfrak{sl}(3, \mathbb{C})$

We will in this section investigate the $\mathfrak{sl}(3, \mathbb{C})$ algebra by finding its structure such as the Cartan subalgebra and the ladder operators. $\mathfrak{sl}(3, \mathbb{C})$ is an 8-dimensional algebra of rank 2 which means that we have a 2-dimensional Cartan subalgebra. One particular representation for this algebra is

$$\begin{aligned}
\pi(t_1) &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \pi(t_2) = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \pi(t_3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
\pi(t_4) &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \pi(t_5) = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \quad \pi(t_6) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
\pi(t_7) &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \quad \pi(t_8) = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix},
\end{aligned} \tag{5.3.1}$$

where $t_i \in \mathfrak{sl}(3, \mathbb{C})$. These are the **Gell-Mann matrices** which can be thought of as a generalization of the Pauli matrices. The main task is to find a suitable basis for $t \in \mathfrak{sl}(3, \mathbb{C})$ that allows us to find our Cartan subalgebra and ladder operators. If we take the commutator between $\pi(t_1)$ and $\pi(t_2)$ we get

$$[\pi(t_1), \pi(t_2)] = 2i\pi(t_3) \tag{5.3.2}$$

which seems familiar from the $\mathfrak{su}(2)$ algebra. Inspired by this we define

$$\begin{aligned}
x_1 &= \frac{1}{2} t_1 \quad x_2 = \frac{1}{2} t_2 \quad x_3 = \frac{1}{2} t_3 \\
y_4 &= \frac{1}{2} t_4 \quad y_5 = \frac{1}{2} t_5 \\
z_6 &= \frac{1}{2} t_6 \quad z_7 = \frac{1}{2} t_7 \quad z_8 = \frac{1}{\sqrt{3}} t_8,
\end{aligned} \tag{5.3.3}$$

and make the change of basis

$$\begin{aligned}
e_1 &= x_1 + i x_2 \quad e_2 = y_4 + i y_5 \quad e_3 = z_6 + i z_7 \\
f_1 &= x_1 - i x_2 \quad f_2 = y_4 - i y_5 \quad f_3 = z_6 - i z_7.
\end{aligned} \tag{5.3.4}$$

Expressed as matrices this becomes

$$\begin{aligned}
\pi(e_1) &= \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \pi(e_2) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \pi(e_3) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \\
\pi(f_1) &= \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \pi(f_2) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \pi(f_3) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \\
\pi(x_3) &= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \pi(z_8) = \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
\end{aligned} \tag{5.3.5}$$

If we take all the commutators we find that

$$\begin{aligned}
&[x_3, e_1] = e_1 \qquad [x_3, e_2] = \frac{1}{2}e_2 \qquad [x_3, e_3] = -\frac{1}{2}e_3 \\
&[x_3, f_1] = -f_1 \quad [x_3, f_2] = -\frac{1}{2}f_2 \quad [x_3, f_3] = \frac{1}{2}f_3 \\
&\\
&[z_8, e_1] = 0 \qquad [z_8, e_2] = e_2 \qquad [z_8, e_3] = e_3 \\
&[z_8, f_1] = 0 \qquad [z_8, f_2] = -f_2 \qquad [z_8, f_3] = -f_3 \\
&\\
&[e_1, e_3] = e_2 \qquad [f_1, f_3] = f_2 \qquad [x_3, z_8] = 0.
\end{aligned} \tag{5.3.6}$$

So, we have that $\mathfrak{C} = \text{span}_{\mathbb{C}}\{x_3, z_8\}$, $\eta_+ = \text{span}_{\mathbb{C}}\{e_1, e_2, e_3\}$ and $\eta_- = \text{span}_{\mathbb{C}}\{f_1, f_2, f_3\}$. Let us denote x_3, z_8 as h_1, h_2 respectively. Here we have another definition of a positive root. Instead of positive h_1 component, we define a positive root as a root with positive h_2 component. In the h_1, h_2 basis for the Cartan subalgebra we have the positive roots

$$\boldsymbol{\alpha}_1 = (1, 0) \quad \boldsymbol{\alpha}_2 = (1/2, 1) \quad \boldsymbol{\alpha}_3 = (-1/2, 1) \tag{5.3.7}$$

where we choose $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_3$ to be the simple roots. This means that

$$\boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_3. \tag{5.3.8}$$

Clearly, this is not expressed in the Chevally basis. From (5.3.6) we can obtain the matrix representation for ad_{h_1}

$$\text{ad}_{h_1} = \begin{pmatrix} 0 & & & & & & & \\ & 0 & & & & & & \\ & & 1 & & & & & \\ & & & \frac{1}{2} & & & & \\ & & & & -\frac{1}{2} & & & \\ & & & & & -1 & & \\ & & & & & & -\frac{1}{2} & \\ & & & & & & & \frac{1}{2} \end{pmatrix} \tag{5.3.9}$$

and for ad_{h_2}

$$\mathrm{ad}_{h_2} = \begin{pmatrix} 0 & & & & & & & \\ & 0 & & & & & & \\ & & 0 & & & & & \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & & & & 1 & & \\ & & & & & & -1 & \\ & & & & & & & -1 \end{pmatrix} \tag{5.3.10}$$

We can obtain the same algebra with the use of $\mathfrak{sl}(2,\mathbb{C})$ as we did in section 5.2 for $\mathfrak{sl}(3,\mathbb{C})$. Consider two $\mathfrak{sl}(2,\mathbb{C})$'s

$$\begin{aligned} &\{f_1', h_1', e_1'\} \\ &\{f_2', h_2', e_2'\}, \end{aligned} \tag{5.3.11}$$

where the primes are to distinguish this basis from the previously obtained from the Gell-Mann matrices. So far we have a 6-dimensional algebra, but $\mathfrak{sl}(3,\mathbb{C})$ is 8-dimensional, i.e. we have to construct more elements to our algebra. Suppose that we know that the rank of $\mathfrak{sl}(3,\mathbb{C})$ is 2 but not that it is 8-dimensional and we wonder how many new elements we can form. It is here that (5.2.15) comes in handy. We know from (5.2.11) that the $\mathfrak{sl}(3,\mathbb{C})'$s Cartan matrix is

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \tag{5.3.12}$$

So, the question is how many times we can commute the simple roots e_1, e_2 as

$$[e_1', [e_1'[..., [e_1', e_2']...]]]. \tag{5.3.13}$$

Equations (5.2.15) and (5.3.12) tells us that

$$(\mathrm{ad}_{e_1'})^{1-A_{12}}(e_2') = (\mathrm{ad}_{e_1'})^2(e_2') = [e_1', [e_1', e_2']] = 0. \tag{5.3.14}$$

Thus, $[e_1', e_2']$ is a new element but $[e_1', [e_1', e_2']] =$ is *not*. The calculation for f_1' and f_2' is completely analogous. The basis for $\mathfrak{sl}(3,\mathbb{C})$ is given by

$$\{f_1', f_2', f_3', h_1', h_2', e_1', e_2', e_3'\}. \tag{5.3.15}$$

Since we know the Cartan matrix, the roots are

$$\boldsymbol{\alpha}_1' = (2, -1) \quad \boldsymbol{\alpha}_2' = (-1, 2) \quad \boldsymbol{\alpha}_3' = (1, 1), \tag{5.3.16}$$

where $\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2'$ are the simple roots. These roots are clearly different from (5.3.7) due to the different basis for $\mathfrak{C}$. We saw in (5.2.16) one way to go from a particular basis in $\mathfrak{g}$ to a basis where all the structure constants are integers, i.e. the Chevally basis. Let us see how we can go from the basis given by (5.3.6) to the Chevally basis. We can find this change of basis in two ways, by direct computation or finding h_{α_i}. If we make an "ordinary" change of basis in $\mathfrak{C}$ we should be able to obtain the same roots as in (5.3.16) for the step operators defined in (5.3.4). Let the new basis of $\mathfrak{C}$, i.e. the Chevally basis denoted h_1', h_2', satisfy

$$\begin{aligned} &[h_1', e_1] = 2e_1 \quad [h_1', e_2] = -e_2 \\ &[h_2', e_1] = -e_1 \quad [h_2', e_2] = 2e_2. \end{aligned} \tag{5.3.17}$$

To find this basis we express h_1', h_2' in h_1, h_2 as

$$\begin{aligned} h_1' &= c_1 h_1 + c_2 h_2 \\ h_2' &= d_1 h_1 + d_2 h_2 \end{aligned} \tag{5.3.18}$$

and solve this using (5.3.17) and (5.3.7). This will give us

$$\begin{aligned} h_1' &= 2h_1 \\ h_2' &= -h_1 + \frac{3}{2} h_2. \end{aligned} \tag{5.3.19}$$

so that our new basis for $\mathfrak{g}$ becomes $\{f_1, f_2, f_3, h_1', h_2', e_1, e_2, e_3\}$. The matrix representation for h_1', h_2' becomes

$$\begin{aligned} h_1' &= 2 \cdot \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ h_2' &= -\frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \frac{3}{2} \cdot \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}. \end{aligned} \tag{5.3.20}$$

The same result is obtained if we find $h_{\alpha_{1,\,2}}$ from (5.1.48).

$$\begin{aligned} h_1' &= \frac{2}{\langle \alpha_1, \alpha_1 \rangle} h_{\alpha_1} \\ h_2' &= \frac{2}{\langle \alpha_2, \alpha_2 \rangle} h_{\alpha_2} \end{aligned} \tag{5.3.21}$$

Equation (5.1.48) gives for h_{α_1}

$$\begin{cases} 1 = 3c^1 \\ 0 = 4c^2 \end{cases} \Rightarrow h_{\alpha_1} = \frac{1}{3} h_1 \tag{5.3.22}$$

and for h_{α_2}

$$\begin{cases} -1/2 = 3c^1 \\ 1 = 4c^2 \end{cases} \Rightarrow h_{\alpha_2} = -\frac{1}{6} h_1 + \frac{1}{4} h_2. \tag{5.3.23}$$

Here we have used that $(h_1, h_1) = 3$, $(h_2, h_1) = (h_1, h_2) = 0$ and $(h_2, h_2) = 4$ which can be obtained from (5.1.58) or from (5.3.9), (5.3.10). Thus

$$\begin{aligned} h_1' &= \frac{2}{\langle \alpha_1, \alpha_1 \rangle} \frac{1}{3} h_1 \\ h_2' &= \frac{2}{\langle \alpha_2, \alpha_2 \rangle} \left(-\frac{1}{6} h_1 + \frac{1}{4} h_2 \right). \end{aligned} \tag{5.3.24}$$

With

$$\begin{aligned} \langle \alpha_1, \alpha_1 \rangle &= (h_{\alpha_1}, h_{\alpha_1}) = \frac{1}{9}(h_1, h_1) = \frac{1}{3} \\ \langle \alpha_2, \alpha_2 \rangle &= (h_{\alpha_2}, h_{\alpha_2}) = \left(-\frac{1}{6} h_1 + \frac{1}{4} h_2, \, -\frac{1}{6} h_1 + \frac{1}{4} h_2 \right) = \frac{1}{3} \end{aligned} \tag{5.3.25}$$

equation (5.3.24) becomes

$$\begin{aligned} h_1' &= 2h_1 \\ h_2' &= -h_1 + \frac{3}{2}h_2. \end{aligned} \tag{5.3.26}$$

There is an illustrative way to present the roots for an algebra through a **root diagram**, just as we did for the weights. For a rank 2 algebra, the roots can be thought of as 2-dimensional vectors in the plane. If we choose the x- and y-axis to correspond to the h_1 and h_2 component of the roots we obtain for $\mathfrak{sl}(3,\mathbb{C})$ the root diagram in figure 5.2. For convenience, we make

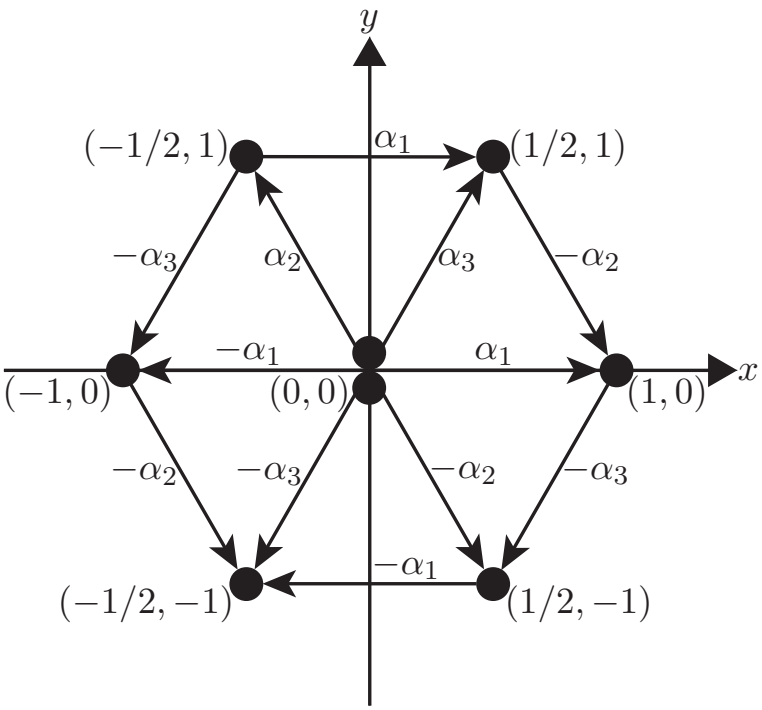


Figure 5.2: Root diagram for $\mathfrak{sl}(3,\mathbb{C})$. The two roots in the origin corresponds to the Cartan elements h_1, h_2.

the change $\alpha_3 \leftrightarrow \alpha_2$. The eight points correspond to the roots of which the two points in the origin are the roots to the Cartan elements. As highest weight for the adjoint representation we have α_3, i.e. its e_3 is annihilated when commuted with e_1 or e_2. The figure gives a good overview of all the roots and how they are related.

Since the roots are the weights of the adjoint representation, this is a good example of what we talked about in the end of section 5.1.3. We can start with the highest weight (root) state e_2 and step down with our step operators $\pi(f_1)$, $\pi(f_2)$ and $\pi(f_3)$[5] subtracting α_1 and α_2, respectively, from the highest weight. From the highest weight there are three ways to go, act with the $\pi(f_1)$ operator on e_3 to get e_2, corresponding to the weight $\alpha_2 = (-1/2, 1)$, act with $\pi(f_2)$ to get e_1, corresponding to the weight $\alpha_1 = (1, 0)$ or act with $\pi(f_3)$ to get a

[5]This step operator is actually unnecessary since its action is equivalent to a linear combination of $\pi(f_1), \pi(f_2)$.

Cartan element. Next, we repeat the procedure from the new states, i.e. e_1, e_2 and some linear Cartan element. Since $\pi(f_2)e_1 = [f_2, e_1] = 0$, there are only two possible ways to step away from e_1, to f_2 or to a Cartan element. If we continue in this way, we obtain the adjoint irreducible representation of $\mathfrak{sl}(3, \mathbb{C})$. This procedure might seem like the one we did when we first obtained all the step operators to $\mathfrak{sl}(3, \mathbb{C})$ but in "reverse". In that case we stared with our simple roots and stepped up instead of down, so we might wounder what the meaning of this example is. The answer is that this is the general procedure when we construct irreducible representations by the highest weight state and that is why we did this example.

Let us now leave the adjoint representation and begin to consider other representations for $\mathfrak{sl}(3, \mathbb{C})$. In order to do this we introduce the **weight lattice** defined by

$$\Lambda^i \alpha_j = \delta^i{}_j, \tag{5.3.27}$$

where Λ^i is a weight and $\delta^i{}_j$ is the Kronecker delta. The weight lattice Γ_Λ consist of all the linear combination of weights, i.e. $\Gamma_\Lambda = \sum_i \mathbb{Z}\Lambda_i$, and can be considered as the dual space to the root space. If we add our weights in the root diagram in figure 5.2 we obtain figure 5.3. The weights Λ_1, Λ_2 can be obtained by the conditions

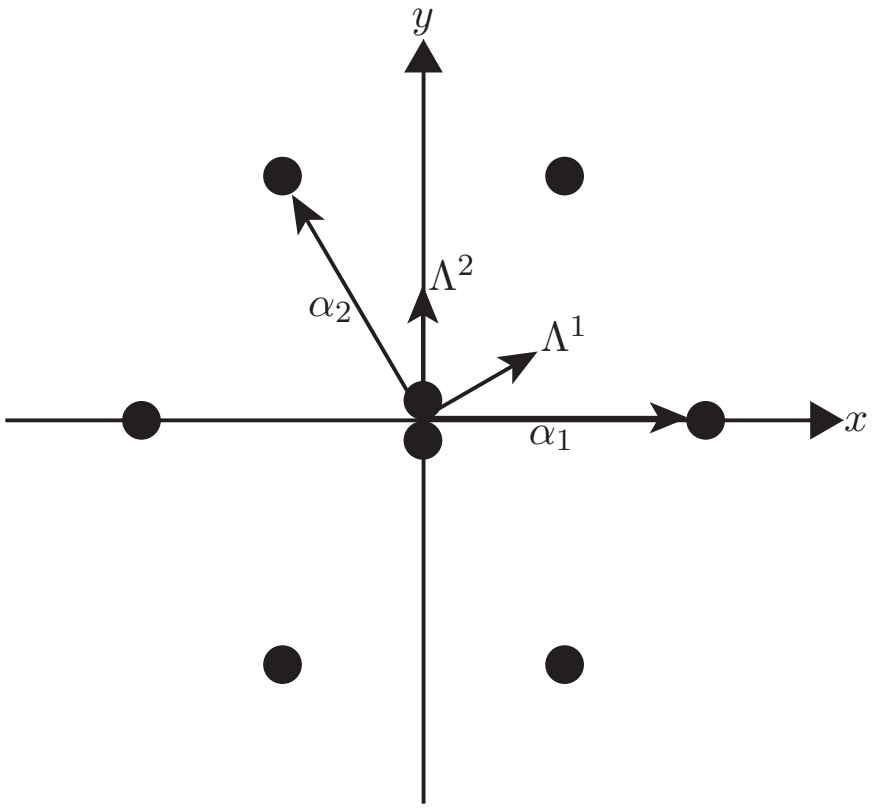


Figure 5.3: Root and weight diagram for $\mathfrak{sl}(3, \mathbb{C})$. The two weights Λ_1 and Λ_2 are orthogonal to the roots α_2 and α_1, respectively.

$$\begin{cases} \langle \Lambda_{1,2}, \alpha_{2,1} \rangle = \Lambda_{1,2} g^{-1} \alpha_{2,1} = 0 \\ \langle \Lambda_{1,2}, \alpha_{1,2} \rangle = \Lambda_{1,2} g^{-1} \alpha_{1,2} = 1 \end{cases} \tag{5.3.28}$$

where the metric inverse g^{-1} is

$$g^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 3/2 \end{pmatrix}. \tag{5.3.29}$$

If we solve (5.3.28) we get $\Lambda_1 = (1/2, 1/3)$ and $\Lambda_2 = (0, 2/3)$. Let us now start with Λ_2 as highest weight. To construct an irreducible representation, corresponding to this weight, we apply the step operators. There are two possible ways to step down: with α_1 or α_2. α_3 is not necessary since it is a linear combination of α_1 and α_1. To see whether $\Lambda_2 - \alpha_1$ and $\Lambda_2 - \alpha_2$ are weights or not we use (5.1.88) (the length of the weight chain),

$$\begin{aligned} n &= \frac{2\langle \Lambda_2, \alpha_1 \rangle}{\langle \alpha_1, \alpha_1 \rangle} = 0 \\ n &= \frac{2\langle \Lambda_2, \alpha_2 \rangle}{\langle \alpha_2, \alpha_2 \rangle} = 1 \end{aligned} \tag{5.3.30}$$

by the definition of Λ_2. Thus, $\Lambda_2 - \alpha_1$ is *not* a weight but $\Lambda_2 - \alpha_2$ is. This is not so unexpected considering how we chose our weights. From the weight $\Lambda_2 - \alpha_2 = (1/2, -1/3)$ we can not go any further with α_2, but

$$n = \frac{2\langle \Lambda_2 - \alpha_2, \alpha_1 \rangle}{\langle \alpha_1, \alpha_1 \rangle} = \frac{2\langle \Lambda_2, \alpha_1 \rangle}{\langle \alpha_1, \alpha_1 \rangle} - \frac{2\langle \alpha_2, \alpha_1 \rangle}{\langle \alpha_1, \alpha_1 \rangle} = 1, \tag{5.3.31}$$

i.e. $\Lambda_2 - \alpha_2 - \alpha_1 = (-1/2, -1/3)$ is a weight. In the above calculations we used that $\langle \alpha_1, \alpha_1 \rangle = \langle \alpha_2, \alpha_2 \rangle = 2$. These three weights form a triangle in the weight lattice as illustrated in figure 5.4. An equivalent representation is obtained if we start from Λ_2 and step down with α_3 and then up with α_1. This 3-dimensional representation, denoted $\bar{\mathbf{3}}$, is the *anti-fundamental* representation of $\mathfrak{sl}(3, \mathbb{C})$. The irreducible representation with highest weight Λ_1 is derived in exactly the same way as for Λ_2. This representation is called the *fundamental* representation[6] and is denoted as $\mathbf{3}$. The result is presented in figure 5.5. These representations together with the adjoint representation have a very nice physical significance that will bee discussed in chapter 8.

5.4 Dynkin Diagrams

The Cartan matrix encodes all information about the commutation relations, and thus the entire structure, of a semi-simple Lie algebra. If we know the Cartan matrix of a Lie algebra, we know all there is to know about the algebra, and a graphical way of presenting this information is in the form of a **Dynkin diagram**.

For a semi-simple Lie algebra of rank r with Cartan matrix A, the Dynkin diagram is drawn in the following way:

- Draw a node for every $i = 1,..., r$.

[6] Another common notation for the fundamental and anti-fundamental representation is $(1, 0)$ and $(0, 1)$, respectively. These are the integers n_j talked about in (5.1.97).

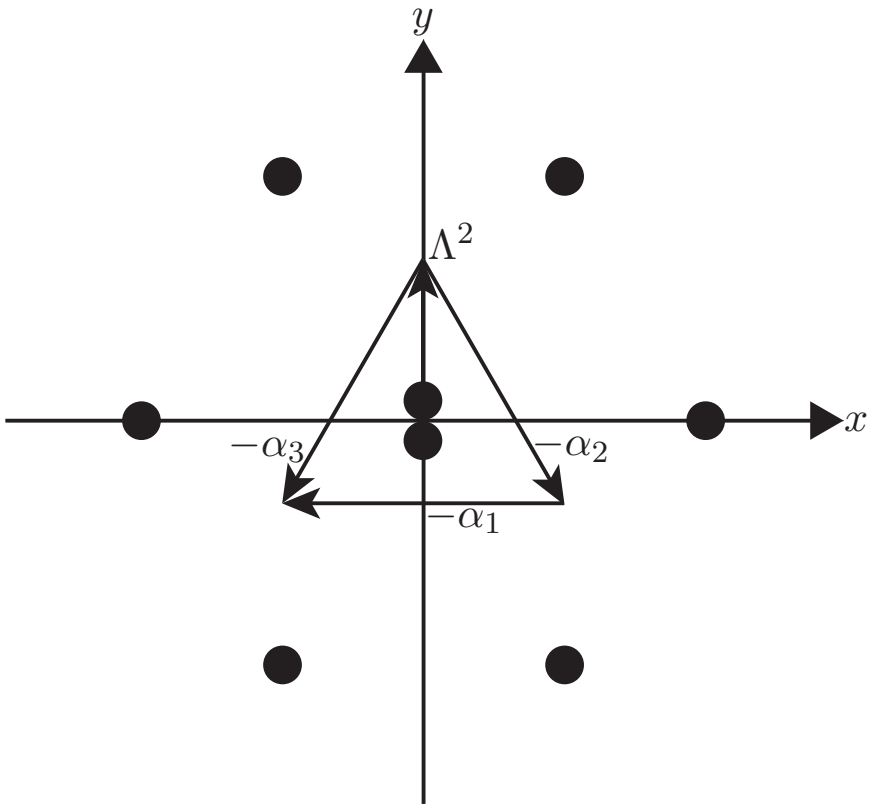


Figure 5.4: Weights for the 3-dimensional complex conjugate representation $\bar{\mathbf{3}}$ for $\mathfrak{sl}(3,\mathbb{C})$

- Draw a number of lines equal to $\max(|A_{ij}|, |A_{ji}|)$ between nodes i and j.
- Draw an arrow from j to i if $|A_{ij}| > |A_{ji}|$.

We will now look at a few examples that demonstrates how Dynkin diagrams are drawn. At the same time we introduce the mathematical names of the algebras which refer to their structure more than their representations. The algebras $\mathfrak{sl}(n,\mathbb{C})$ are in the mathematical literature called A_{n-1} and are thus indexed by their rank $n-1$. Other names will follow below.

Example 5.4.1. Dynkin diagram of $\mathfrak{sl}(2,\mathbb{C})$. From the commutation relations

$$\begin{aligned} [e,f] &= h \\ [h,e] &= 2e \\ [h,f] &= -2e, \end{aligned} \tag{5.4.1}$$

of $\mathfrak{sl}(2,\mathbb{C})$ (also called A_1), we see that its Cartan matrix is simply $A = 2$ (its rank is $r = 1$). Its Dynkin diagram is thus nothing more than an isolated node.

Example 5.4.2. Dynkin diagram of $\mathfrak{sl}(3,\mathbb{C})$. Recall that $\mathfrak{sl}(3,\mathbb{C})$ (also called A_2) has rank $r = 2$, with Cartan matrix

$$A = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}. \tag{5.4.2}$$

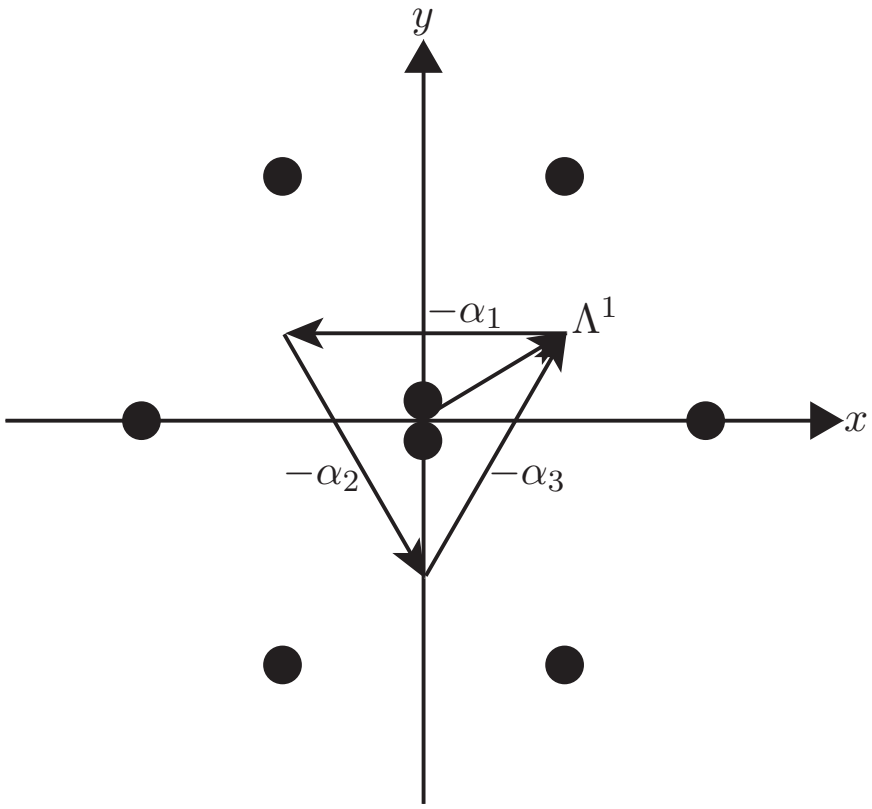


Figure 5.5: Weights for the 3-dimensional fundamental representation **3** for $\mathfrak{sl}(3,\mathbb{C})$

The number of nodes in the Dynkin diagram is equal to the rank of the Lie algebra, which means that the Dynkin diagram of $\mathfrak{sl}(3,\mathbb{C})$ will have 2 nodes, and the number of lines between these is equal to $\max(|A_{12}|,|A_{21}|) = \max(|-1|,|-1|) = 1$. The diagram is shown in figure 5.6.

Example 5.4.3. $\mathfrak{sl}(2,\mathbb{C})$ and $\mathfrak{sl}(3,\mathbb{C})$ are special cases of the more general group $\mathfrak{sl}(r+1,\mathbb{C})$ (also called simply A_r) of rank r, with the Cartan matrix

$$A = \begin{pmatrix} 2 & -1 & 0 & & & \\ -1 & 2 & -1 & & ... & \\ 0 & -1 & & & & \\ & & & 0 & -1 & 0 \\ & ... & & -1 & 2 & -1 \\ & & & 0 & -1 & 2 \end{pmatrix}. \tag{5.4.3}$$

Note that for every i, only the elements A_{ij} with $j = i$ or $j = i \pm 1$ will be non-zero. This means that if we arrange the nodes in the corresponding Dynkin diagram in a line, every node will be connected only to the two adjacent nodes, as illustrated in figure 5.6.

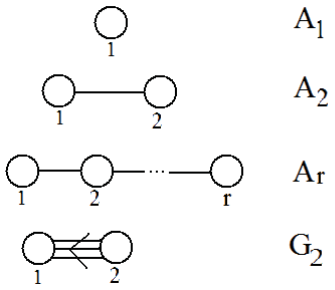


Figure 5.6: The Dynkin diagrams of A_1, A_2, A_r and G_2.

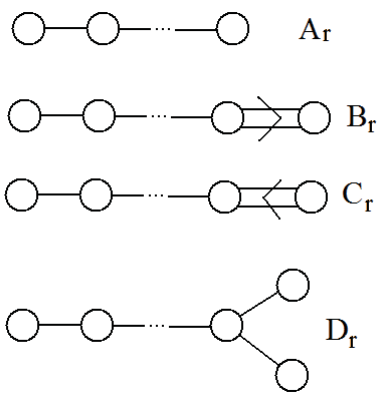


Figure 5.7: The Dynkin diagrams of the infinite families A_r, B_r, C_r and D_r.

Example 5.4.4. The Lie algebra G_2 has the Cartan matrix

$$A = \begin{pmatrix} 2 & -3 \\ -1 & 2 \end{pmatrix} \tag{5.4.4}$$

Since the Lie algebra has rank 2, the Dynkin diagram will consist of 2 nodes. Call them node 1 and node 2. $|A_{12}| > |A_{21}|$, and thus we should draw three lines ($\max(|A_{12}|, |A_{21}|) = 3$), with arrows pointing from 2 to 1. See figure 5.6.

The Dynkin diagram of a simple algebra will always be connected, and since semi-simple Lie algebras can be written as the direct sum of simple Lie algebras, these will in general consist of several disconnected components, where each component is just the Dynkin diagram of the corresponding simple algebra. One of the most prominent results in mathematics during the 20th century was the complete classification of simple Lie algebras over algebraically closed fields. It turns out that the finite simple Lie algebras over the complex numbers can be classified into 4 infinite families, A_r, B_r, C_r and D_r, with 5 exceptional cases, G_2, F_4, E_6, E_7 and E_8. This classification is neatly summarized in the Dynkin diagrams of which the diagrams of the infinite families are shown in figure 5.7, and the diagrams of the exceptional cases are shown in figure 5.8.

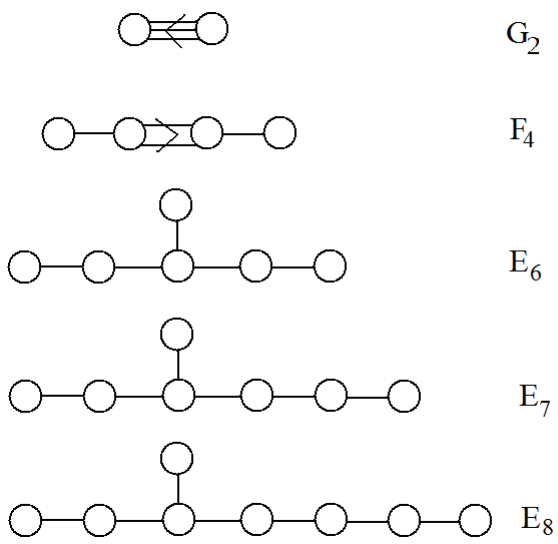


Figure 5.8: The Dynkin diagrams of the exceptional cases G_2, F_4, E_6, E_7 and E_8.

5.5 Casimir Operators

We end this chapter with a brief discussion about Casimir operators in Lie algebra representations. As mentioned in section 4.4, a Casimir operator $\hat{C}$ is a Lie algebra element that commutes with every other elements in the Lie algebra $\mathfrak{g}$, i.e.

$$[\hat{C}, t_i] = 0, \quad \forall t_i \in \mathfrak{g}. \tag{5.5.1}$$

This means that the Casimir operator is proportional to the unit matrix. Moreover, for a semi-simple Lie algebra the number of independent Casimir operators is equal to the rank of the algebra, ([16], p. 109). An example of a Casimir operator was given in section 4.4, i.e.

$$\hat{C} = \hat{\boldsymbol{J}}^2 = \hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2, \tag{5.5.2}$$

which is the Casimir operator for $\mathfrak{su}(2)$, called a **quadratic Casimir operator**, in a particular representation. The set of Casimir operators for a given Lie algebra is not unique. For example: if $\hat{C}_1$ and $\hat{C}_2$ are two Casimir operators then

$$\hat{C}' = a\hat{C}_1 + b\hat{C}_2 \tag{5.5.3}$$

is also a Casimir operator for every $a, b \in \mathbb{C}$. So far we have only talked about Casimir operators to the Lie algebra, but let us find an *Hermitian* invariant operator for an *unitary* group with *Hermitian* generators. Let $\hat{C}$ be an invariant operator (commutes with every $\hat{U}$), i.e.

$$\hat{C}\hat{U} = \hat{U}\hat{C}, \tag{5.5.4}$$

where $\hat{U}$ is an unitary operator in the group realisation. Take the Hermitian conjugate on both sides

$$\hat{U}^\dagger \hat{C}^\dagger = \hat{C}^\dagger \hat{U}^\dagger \tag{5.5.5}$$

and use that $\hat{U}^\dagger = \hat{U}^{-1}$. Since the inverse of every group element $\hat{U}$ covers the whole group, 5.5.5 is equivalent to

$$\hat{U}\hat{C}^\dagger = \hat{C}^\dagger\hat{U}. \tag{5.5.6}$$

Thus, $\hat{C}^\dagger$ is also an invariant operator. If we apply (5.5.3) to $\hat{C}$ and $\hat{C}^\dagger$ we get

$$\hat{C}' = \hat{C} + \hat{C}^\dagger \tag{5.5.7}$$

which obviously is Hermitian. Since the operator $\hat{C}'$ is an invariant operator to the group it follows that $\hat{C}'$ is a Casimir operator to the Lie algebra. Consider for example the infinitesimal group transformation $U = 1 + \lambda^a t_a$ where λ^a are the group parameters and t_a the generators for the group. If we act with $\hat{C}'$ and use that it is an invariant operator we get

$$\hat{C}'(1 + \lambda^a t_a) = (1 + \lambda^a t_a)\hat{C}' \Leftrightarrow \hat{C}' t_a \quad = t_a \hat{C}' \tag{5.5.8}$$

for every Lie algebra element t_a. We can now see why the action of a rotation on a spin state does not change the eigenvalue of $\hat{\boldsymbol{J}}^2$, i.e. the Casimir operator. Recall the rotation operators $\hat{D}$ in section 4.3.16, i.e. a group representation of $SU(2)$. A state $|\psi\rangle$ with spin j will change its spin under the action of $\hat{D}$ according to

$$\hat{\boldsymbol{J}}^2\hat{D}|\psi\rangle = \hat{D}\hat{\boldsymbol{J}}^2|\psi\rangle = \hbar j(j+1)\hat{D}|\psi\rangle. \tag{5.5.9}$$

Thus, the spin is invariant under $\hat{D}$.

There is no general method to construct a Casimir operator, so you have to consider the particular algebra in each case that you are interested of. For the $SU(n)$ groups, however, there exists a way, ([16], p.115), that is based on the fact that the Casimir operators must be polynomial of the Lie elements, i.e.

$$\hat{C}_i = \sum a_{i,jk\ldots}\pi(t_j)\pi(t_k)\ldots \ , \tag{5.5.10}$$

where $a_{i,jk\ldots}$ are constants and t the Lie algebra elements. Every semi-simple Lie algebra has a quadratic Casimir, [2], which is constructed as

$$\hat{C}_2 = \pi(t_i)\pi(t_j)g^{ij}, \tag{5.5.11}$$

where the subscript 2 indicates that it is a quadratic Casimir operator. t^i is an element of the dual basis for $\mathfrak{g}$ and $g_{ij} = (t_i, t_j)$, i.e. the Cartan-Killing form. For $SU(2)$

$$g_{ij} = \begin{pmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{pmatrix} \Rightarrow \quad g^{ij} = \begin{pmatrix} -1/2 & 0 & 0 \\ 0 & -1/2 & 0 \\ 0 & 0 & -1/2 \end{pmatrix}. \tag{5.5.12}$$

Thus, the Casimir operator for $SU(2)$ is

$$\hat{C} = -\frac{1}{2}(t_x^2 + t_y^2 + t_z^2) \tag{5.5.13}$$

which of course is equivalent to 5.5.2 up to a constant faktor. For a rank 2 algebra such as $\mathfrak{su}(3)$ there are two Casimir operators given by

$$\hat{C}_1 = \sum_{i=1}^{8}\pi(t_i), \quad \hat{C}_2 = \sum_{ijk} a_{i,jk}\pi(t_i)\pi(t_j)\pi(t_k) \tag{5.5.14}$$

where the $\pi(t_i)$:s are given by (5.3.1).

In (5.1.97) we mentioned a way to label our irreducible representations through the integers n. Another very common way to label irreducible representations is with the eigenvalues of the Casimir operators, since these are invariants. Thus, for a rank r Lie algebra the irreducible representations can be labeled with r numbers. For $SU(2)$, the irreducible representations are labeled with the spin j.

This chapter has given us a deeper insight about the structure of Lie algebras and many important results such as the commuting relations of the Lie algebra elements, 5.1.15, the Cartan-Killing form, weights and the Cartan matrix. The concepts of highest weight state and how to create irreducible representations has been presented and applied to $\mathfrak{sl}(3,\mathbb{C})$. The chapter is characterized by its quite mathematical presentation, even though the connection to physics has been made occasionally, which is intended to give a good foundation for applications in physics. We will come back to this chapter in section 8.7 where we will see how particles and irreducible representations are related.

Chapter 6

The Lorentz and Poincaré Groups

After the mathematical treatment of Lie algebras in the previous chapter, we will now begin a description of physical applications of this framework, starting with the symmetries of spacetime. In the beginning of the twentieth century Albert Einstein unified space and time in his theory of special relativity. This theory can be conveniently expressed in Minkowski spacetime. The symmetry operations we can perform on a physical object are quite a few. For instance, it can be rotated and translated in space as well as time. Such symmetries of spacetime where we have conserved quantities associated with temporal invariance, spatial invariance and rotational invariance are known as *external symmetries*.

To comprehend these notions through the usage of Lie algebra and representation theory, we will return to the Lorentz group which encodes special relativity. It was mentioned briefly in section 2.2. Moreover, we will discover its relation with the more general Poincaré group which also includes translation. By the derivation of the Poincaré algebra through actions of infinitesimal adjoint representations several physical applications will be validated. We will obtain new tools for characterizing elementary particles with quantum numbers and explore the properties of so-called projective representations with the formalism described in chapter 5. Proceeding, we will also investigate different manifestations of the Lorentz and Poincaré group as well as their topological structure. In addition we will see the first hints on how one can unify quantum mechanics with special relativity, finding our accustomed quantum mechanical operators in a wider context of Minkowski spacetime, through quantum field theory (QFT). Numerous books explore this theory thoroughly, see for instance [5], [17] and [6]. Especially chapter 2 of *The Quantum Theory of Fields: Vol I - Foundations* [5] has provided inspiration. This chapter with its introduction of external symmetries serves as a basis for the next two chapters which will mainly deal with *internal symmetries*.

Having said that, let us explore this topic and return to the realm of Lorentz transformations and their attributes.

6.1 Properties of the Lorentz Group

From special relativity we recall that a world without gravity can be thought of as four-dimensional Minkowski spacetime which we denote as $\mathbb{R}^{1,3}$, where 3 represents the three spatial dimensions and 1 the temporal one. We have that $\mathbb{R}^3_{x,\,y,\,z} \times \mathbb{R}_t = \mathbb{R}^{1,3}$. Let us set the speed of light $c = 1$. Note that we use Einstein's summation convention; that is repeated

indices are summed over with one subscript and the other one superscript or vice versa. Vectors in this space are called *4-vectors* $\boldsymbol{A} = (t,\, x,\, y,\, z)$ with coordinates A^μ, $\mu = 0, 1, 2, 3$ where $t = A^0$, $x = A^1$, $y = A^2$ and $z = A^3$. A **4-vector** $\boldsymbol{A}$ transforms as $\boldsymbol{A} \to \boldsymbol{A'} = \Lambda \boldsymbol{A}$ under a general *Lorentz transformation* Λ when the coordinate system is transformed by $\boldsymbol{x} \to \boldsymbol{x'} = \Lambda \boldsymbol{x}$ ([18], p. 35). These transformations are defined as:

Definition 6.1.1. *A* **Lorentz transformation** Λ *is a matrix representation of an element in the group* $SO(1,3)$*, i.e. a transformation such that it preserves the Minkowski norm through*

$$\eta_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma = \eta_{\rho\sigma} \tag{6.1.1}$$

where $\eta_{\mu\nu}$ *is the* **Minkowski metric tensor** *commonly known as the* **Minkowski metric***:*

$$\eta_{\mu\nu} \overset{def}{=} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6.1.2}$$

Occasionally, one defines the component which will act on the time component of a 4-vector as $+1$ instead and changes the sign of the other space-related components in (6.1.2). In general physicists use $\eta_{\mu\nu}$ to describe the entire Minkowski metric tensor as well as its components, while mathematicians use a different notation, careful to distinguish between the tensor and its components. The Minkowski inner product between two 4-vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ with components A^μ and B^μ is given by

$$\boldsymbol{A} \cdot \boldsymbol{B} = A_\mu B^\mu = \eta_{\mu\nu} A^\mu B^\nu \tag{6.1.3}$$

and

$$\boldsymbol{A}^2 = \boldsymbol{A} \cdot \boldsymbol{A} = A_\mu A^\mu = \eta_{\mu\nu} A^\mu A^\nu. \tag{6.1.4}$$

With the definition as in (6.1.2), the time direction will be negative when we lower the index of an arbitrary 4-vector.

It is possible to show the implications of equation (6.1.1) more clearly with a few calculations. Suppose $\Lambda \in SO(1,\, 3)$ and let $\boldsymbol{A'} = \Lambda \boldsymbol{A}$. Then according to the definition

$$\boldsymbol{A'}^2 = \boldsymbol{A'} \cdot \boldsymbol{A'} = (\Lambda \boldsymbol{A})^2 = \boldsymbol{A}^2. \tag{6.1.5}$$

In tensor language we can rewrite this to see the relations more evidently. Recall that $\boldsymbol{A'} = \Lambda \boldsymbol{A}$ in component form becomes $A'^\mu = \Lambda^\mu{}_\nu A^\nu$. Hence, in component form, equation (6.1.5) reads:

$$\begin{aligned} \boldsymbol{A'^2} &= \eta_{\mu\nu} A'^\mu A'^\nu = \eta_{\mu\nu} \left(\Lambda^\mu{}_\rho A^\rho\right) \left(\Lambda^\nu{}_\sigma A^\sigma\right) \\ &= \eta_{\mu\nu}\Lambda^\mu{}_\rho\Lambda^\nu{}_\sigma A^\rho A^\sigma = \eta_{\rho\sigma} A^\rho A^\sigma. \end{aligned} \tag{6.1.6}$$

Hereby, we see that the transformation Λ preserves the Minkowski norm.

We will now discuss the group properties of the set of all Lorentz transformations. A product of several Lorentz transformations is clearly another Lorentz transformation and the product is associative. The identity transformation is given by $\Lambda^\mu{}_\nu = \delta^\mu{}_\nu$ (i.e. the identity matrix)

and for every $\Lambda \in SO(3,1)$ there exists an inverse. If we for instance want to locate the inverse to $\Lambda^{\mu}{}_{\nu}$, $(\Lambda^{-1})^{\mu}{}_{\nu}$, we notice that we can restate the LHS of (6.1.1) as $\Lambda_{\nu\rho}\Lambda^{\nu}{}_{\sigma}$ since the metric lowers one of the indices and only has non-zero elements on the diagonal where $\mu = \nu$. Still, if we choose to raise the ρ index on the left hand side as well as the right hand side, we obtain $\Lambda_{\nu}{}^{\rho}\Lambda^{\nu}{}_{\sigma} = \delta^{\rho}{}_{\sigma}$ since

$$\eta^{\tau\rho}\eta_{\rho\mu} = \delta^{\tau}{}_{\mu}. \tag{6.1.7}$$

Yet, by multiplying the inverse of a group element with the group element in question yields the identity, i.e. $(\Lambda^{-1})^{\mu}{}_{\nu}\,\Lambda^{\nu}{}_{\sigma} = \delta^{\mu}{}_{\sigma}$. Hence, the inverse is given by

$$\left(\Lambda^{-1}\right)^{\mu}{}_{\nu} = \Lambda_{\nu}{}^{\mu}. \tag{6.1.8}$$

Let us investigate different subgroups of the Lorentz group. Equation (6.1.1) implies that $(\det\Lambda)^2 = 1$ and thus $\det\Lambda = \pm 1$. Transformations with $\det\Lambda = 1$ are referred to as **proper** and those with $\det\Lambda = -1$ as **improper**. A proper transformation preserves the orientation of an arbitrary 4-vector. Moreover, by evaluating the 0:th components of (6.1.1) through ordinary matrix multiplication, the following equation is obtained:

$$\begin{aligned} &-\left(\Lambda^{0}{}_{0}\right)^2 + \Lambda^{i}{}_{0}\Lambda^{i}{}_{0} = -1 \\ &\Leftrightarrow \left(\Lambda^{0}{}_{0}\right)^2 = \Lambda^{i}{}_{0}\Lambda^{i}{}_{0} + 1 \end{aligned} \tag{6.1.9}$$

where $i = 1, 2, 3$. This can easily be discerned by writing (6.1.1) in matrix form:

$$\Lambda^{T}\eta\Lambda = \eta. \tag{6.1.10}$$

Since $\Lambda^{i}{}_{0}\Lambda^{i}{}_{0}$ represents the norm of an ordinary 3-vector, it must be positive. Hence, we deduce an essential relation for $\Lambda^{0}{}_{0}$; it must satisfy either

$$\Lambda^{0}{}_{0} \leqslant -1 \text{ or } \Lambda^{0}{}_{0} \geqslant 1. \tag{6.1.11}$$

When a Lorentz transformation Λ satisfies $\Lambda^{0}{}_{0} \geqslant 1$ it is called **orthochronous** and it preserves the sign of the time-component of the 4-vector it acts on, and hence the direction of time. In physical applications we deal with proper, orthochronous Lorentz transformations which are representations of the group $SO(1,3)$ or more accurately $SO^{+}(1,3)$ where the $+$ says that the representation matrices are orthochronous. However it is customary to omit the plus sign.

Now, let us regard the Lorentz transform in a different light. We have previously discussed a definition which actually is just applicable for the so-called *homogeneous Lorentz group*. Instead, let us consider a **Poincaré transformation** or **inhomogeneous Lorentz transformation**. These transformations take a set of coordinates $\boldsymbol{x}$ to $\boldsymbol{x'}$ through

$$x^{\mu} \to x'^{\mu} = \Lambda^{\mu}{}_{\nu}x^{\nu} + a^{\mu} \tag{6.1.12}$$

where $\boldsymbol{a}$ is an arbitrary translation expressed as a 4-vector with components a^{μ} and Λ obeys (6.1.1). Infinitesimally we still have

$$\eta_{\mu\nu}\,\mathrm{d}x'^{\mu}\,\mathrm{d}x'^{\nu} = \eta_{\mu\nu}\,\mathrm{d}x^{\mu}\,\mathrm{d}x^{\nu} \tag{6.1.13}$$

which relates the coordinates in different inertial frames since $\boldsymbol{a}$ is constant.

Secondly, let us perform an additional transformation taking $\boldsymbol{x'}$ to $\boldsymbol{x''}$. This yields the following result:

$$\begin{aligned} x'^{\rho} \rightarrow x''^{\rho} &= \tilde{\Lambda}^{\rho}{}_{\sigma} x'^{\sigma} + \tilde{a}^{\rho} = \\ &= \tilde{\Lambda}^{\rho}{}_{\sigma} \left(\Lambda^{\sigma}{}_{\nu} x^{\nu} + a^{\sigma}\right) + \tilde{a}^{\rho} = \\ &= \tilde{\Lambda}^{\rho}{}_{\sigma} \Lambda^{\sigma}{}_{\nu} x^{\nu} + \tilde{\Lambda}^{\rho}{}_{\sigma} a^{\sigma} + \tilde{a}^{\rho}. \end{aligned} \tag{6.1.14}$$

The same result may thus be obtained directly from a transformation $\boldsymbol{x''} = \tilde{\Lambda}\Lambda\boldsymbol{x} + \tilde{\Lambda}\boldsymbol{a} + \tilde{\boldsymbol{a}}$. Abbreviated the transformation induced by a matrix Λ and a translational 4-vector $\boldsymbol{a}$ can be written as $T(\Lambda, \boldsymbol{a})$. From the reasoning above we deduce that different transformations $T(\tilde{\Lambda}, \tilde{\boldsymbol{a}})$ and $T(\Lambda, \boldsymbol{a})$ adhere to the composition rule

$$T(\tilde{\Lambda}, \tilde{\boldsymbol{a}})T(\Lambda, \boldsymbol{a}) = T(\tilde{\Lambda}\Lambda, \tilde{\Lambda}\boldsymbol{a} + \tilde{\boldsymbol{a}}). \tag{6.1.15}$$

The set of all transformations $T(\Lambda, \boldsymbol{a})$ is known as the **Poincaré group** or **the inhomogeneous Lorentz group**. The Poincaré group is often denoted $ISO(1, 3)$ and is composed of the proper orthochronous Lorentz group $SO(1, 3)$ and the translations $\boldsymbol{a}$ in $\mathbb{R}^4$ encoded by the Abelian group $\mathbb{R}^4$. In section 6.3 the nature of its composition and relations to the Lorentz group is developed further. With all translational vectors $\boldsymbol{a}, \tilde{\boldsymbol{a}}$ set to $\mathbf{0}$ the **homogeneous Lorentz group** is formed and the composition rule is modified accordingly

$$T(\tilde{\Lambda}, \mathbf{0})T(\Lambda, \mathbf{0}) = T(\tilde{\Lambda}\Lambda, \mathbf{0}). \tag{6.1.16}$$

For further remarks about tensors and some special relativity, please have a look at appendix B. The properties described above will guide us in our attempts to comprehend how special relativity can be connected to quantum mechanics. Now, let us take the first tentative steps towards unification.

6.2 From Minkowski Space to Hilbert Space

In order to construct any descriptions of particles' behavior, we must replace our coordinate transformations with transformations acting on fields, quantum mechanical states represented as rays in a Hilbert space. Recall from quantum mechanics that a ray is a one-dimensional Hilbert space. For additional information regarding Hilbert spaces, please look at appendix C.

If we observe a quantum mechanical system in a state represented by one particular ray and another observer looks at the same system but finds it in a different state then the probability for an event to occur must be the same for both observers. For instance, if a particle is moving in one inertial frame and is at rest in another, the probability for it to decay to another particle must be the same in both frames. Considering the coupling between these two observers Wigner stated in the 1930's that it is possible to go from the first ray $\mathscr{R}_1$ to the second ray $\mathscr{R}_2$ through

$$|\Psi_2\rangle = \hat{U}|\Psi_1\rangle \tag{6.2.1}$$

where $|\Psi_1\rangle$ and $|\Psi_2\rangle$ describes a state in $\mathscr{R}_1$ and $\mathscr{R}_2$, respectively, and $\hat{U}$ is a linear, unitary operator. Still, the theorem also states that this could be achieved by an antiunitary, antilinear

operator but those are quite uncommon in the description of physical symmetries. Clearly there is a trivial operator $\hat{U} = \hat{1}$ mapping a vector $|\Psi_1\rangle$ to $|\Psi_1\rangle$. Moreover, continuity demands that any symmetry transformation, such as a Lorentz transformation or the rotations from sections 4.2 and 4.3, which may be obtained through a continuous change of some parameters, such as velocities or angles, is represented by a linear, unitary operator. In addition, a symmetry transformation which is almost but not quite trivial can be represented by an operator infinitesimally close to the identity.

We have already explored some of the group properties of the set of induced transformations $T(\Lambda, \boldsymbol{a})$ which could be extended to any symmetry transformation T acting on rays. Closure is achieved since symmetry transformations taking a ray onto another and onto yet another one will always end up in a ray. Likewise, associativity is evident. If a symmetry transformation T_1 carries a ray $\mathscr{R}_1$ to $\mathscr{R}_2$ then there is an inverse transformation $(T_1)^{-1}$ which turn it back and an identity transformation $T = 1$. However, when we are dealing with the operators $\hat{U}$ we are acting on vectors $|\Psi\rangle$ rather than rays and as we recall from quantum mechanics two normed vectors can be in the same ray but may diverge up to a phase. This means that if we have a symmetry transformation T_1 which maps a ray $\mathscr{R}_i$ to $\mathscr{R}_j$, then the corresponding operator $\hat{U}(T_1)$ maps a vector $|\Psi_i\rangle$ in $\mathscr{R}_i$ to a vector $\hat{U}(T_1)|\Psi_i\rangle$ in ray $\mathscr{R}_j$ and if T_2 then takes $\mathscr{R}_j$ to $\mathscr{R}_k$, so does the corresponding operator $\hat{U}(T_2)$ then take $\hat{U}(T_1)|\Psi_i\rangle$ to $\hat{U}(T_2)\hat{U}(T_1)|\Psi_i\rangle$ in ray $\mathscr{R}_k$. Still, there is also the opportunity to map a vector $|\Psi_i\rangle$ to $\hat{U}(T_2T_1)|\Psi_i\rangle$ which is also in ray $\mathscr{R}_k$ and it is identical to $\hat{U}(T_2)\hat{U}(T_1)|\Psi_i\rangle$ up to a phase $\phi(T_1, T_2)$.

$$\hat{U}(T_1)\hat{U}(T_2)|\Psi_i\rangle = e^{i\phi(T_1, T_2)}\hat{U}(T_1T_2)|\Psi_i\rangle \tag{6.2.2}$$

It can be shown that the relationship in (6.2.2) is independent of $|\Psi_i\rangle$ if $|\Psi_i\rangle$ is not a superposition of two states with different spin. If that is not so, we really have the operator relationship [5]

$$\hat{U}(T_1)\hat{U}(T_2) = e^{i\phi(T_1, T_2)}\hat{U}(T_1T_2). \tag{6.2.3}$$

When the phase $\phi(T_1, T_2) = 0$, we can say that the operators $\hat{U}$ form a representation of the group of symmetry transformations. Otherwise, we refer to the $\hat{U}$:s as *projective representations* of the group of symmetry transformations. This topic will be explored in section 6.7. It is also possible to eliminate the phase factor through an appropriate enlargement of the Lorentz/Poincaré group.

As we associate a linear, unitary operator $\hat{U}$ to each combination $T(\Lambda, \boldsymbol{a})$, we are able to make *induced transformations* on the vectors in quantum mechanical Hilbert space. Abbreviated, one usually skips the T and simply write $\hat{U}(\Lambda, \boldsymbol{a})$.

$$\hat{U}(T(\Lambda, \boldsymbol{a})) = \hat{U}(\Lambda, \boldsymbol{a}) \tag{6.2.4}$$

With this in mind, we shall now derive the Lie algebra of the Poincaré group.

6.3 The Poincaré Algebra

Please note that this section requires some familiarity with tensor algebra and hence a look at appendix B might prove very helpful. As in previous sections, we can locate the Lie algebra if we make a Taylor expansion of the group elements near the identity. Correspondingly, let

us consider an infinitesimal Lorentz transformation away from the unit element in component form,

$$\Lambda^{\mu}{}_{\nu} = \delta^{\mu}{}_{\nu} + \omega^{\mu}{}_{\nu} \tag{6.3.1}$$

where $\delta^{\mu}{}_{\nu}$ is an infinitesimal displacement expressed as a matrix in component form and $\delta^{\mu}{}_{\nu}$ is the Kronecker delta. The Lorentz condition of equation (6.1.1) implies that

$$\begin{aligned} \eta_{\rho\sigma} &= \eta_{\mu\nu}\left(\delta^{\mu}{}_{\rho} + \omega^{\mu}{}_{\rho}\right)\left(\delta^{\nu}{}_{\sigma} + \omega^{\nu}{}_{\sigma}\right) = \\ &= \eta_{\mu\nu}\left(\delta^{\mu}{}_{\rho}\delta^{\nu}{}_{\sigma} + \delta^{\mu}{}_{\rho}\omega^{\nu}{}_{\sigma} + \delta^{\nu}{}_{\sigma}\omega^{\mu}{}_{\rho} + \mathcal{O}(\omega^2)\right) = \\ &= \eta_{\rho\sigma} + \eta_{\rho\nu}\omega^{\nu}{}_{\sigma} + \eta_{\mu\sigma}\omega^{\mu}{}_{\rho} + \mathcal{O}(\omega^2). \end{aligned} \tag{6.3.2}$$

By using the metric $\eta_{\rho\nu}$ to lower the first index of the ω and comparing sides of equation (6.3.2), we obtain the following property for the infinitesimal displacement

$$\omega_{\rho\sigma} = -\omega_{\sigma\rho}. \tag{6.3.3}$$

Therefore, $\omega_{\rho\sigma}$ is antisymmetric. Finally, we can investigate what happens to the operator $\hat{U}(\Lambda, \boldsymbol{a})$ which corresponds to an induced Poincaré transformation when Λ is an infinitesimal Lorentz transformation and $\boldsymbol{a} = \boldsymbol{\epsilon}$ where the norm of $\boldsymbol{\epsilon}$ is infinitesimal. For convenience, let us denote $\delta^{\mu}{}_{\nu}$ as $\mathbb{1}$ to avoid superfluous use of indices. Since the operator $\hat{U}(\mathbb{1}, 0)$ maps a ray onto itself it is proportional to the unit operator and might be set equal to it with a suitable phase choice. Hence, for infinitesimal Lorentz transformations, the operator $\hat{U}$ can be expressed as the one-dimensional unit operator $\hat{1}$ plus additional terms linear in ω and $\boldsymbol{\epsilon}$. By expanding $\hat{U}$ we obtain

$$\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon}) = \hat{1} + \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} - i\epsilon_{\rho}\hat{P}^{\rho} + \dots \tag{6.3.4}$$

where the $\dots$ indicate terms of higher order in ω and $\boldsymbol{\epsilon}$. Here the ω- and $\boldsymbol{\epsilon}$-independent operators $\hat{J}$ and $\hat{\boldsymbol{P}}$ are the generators of the Poincaré group. In order for $\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon})$ to be unitary, the operators $\hat{J}$ and $\hat{\boldsymbol{P}}$ must be Hermitian. As we will discover soon they correspond to familiar observables. To obtain Hermitian generators the convention is to omit the imaginary unit when defining the generators and leaving it as a constant to multiply with. The $1/2$ in equation (6.3.4) is another product of convention in order to simplify the commutation relations between $\hat{J}$ and $\hat{\boldsymbol{P}}$. Likewise the $\hat{1}$ is in reality the unit operator (i.e. "1"). Additionally, since we have already confirmed that $\omega_{\mu\nu}$ is antisymmetric in indices, this must also be the case for $\hat{J}^{\mu\nu}$. Otherwise, any symmetric part of $\hat{J}^{\mu\nu}$ would be subject to

$$\omega_{\rho\sigma}\hat{J}^{\rho\sigma} = \omega_{\rho\sigma}\hat{J}^{\sigma\rho} = -\omega_{\sigma\rho}\hat{J}^{\sigma\rho} = -\omega_{\rho\sigma}\hat{J}^{\rho\sigma} = 0 \tag{6.3.5}$$

and thus projected out when in contact with $\omega_{\mu\nu}$. Hence any presumed symmetric part of $\hat{J}^{\mu\nu}$ can be omitted and therefore we state that $\hat{J}^{\mu\nu}$ is antisymmetric.

$$\hat{J}^{\rho\sigma} = -\hat{J}^{\sigma\rho} \tag{6.3.6}$$

In fact the components 1,2,3 of $\hat{\boldsymbol{P}}$ correspond to momentum operators which generate translation in the x, y and z-directions whereas $\hat{P}^0$ generates a translation in time and is the Hamilton operator. Moreover, some of the components of $\hat{J}$ are the operators of angular

momentum in the three spatial directions. By finding the commutation relations between the generators we will be able to comprehend these statements in more detail.

As we recall from chapter 3 as well as the previous chapter, we can distinguish the commutation relations of the Lie algebra if we take the adjoint action of group elements on each other and the Lie algebra on itself. Remember that the adjoint action of the group on its Lie algebra preserves the commutation relations. Let us examine the adjoint action of the arbitrary operator $\hat{U}(\Lambda, \boldsymbol{a})$ on the operator $\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon})$ which is infinitesimally close to the unit operator.

$$\hat{U}(\mathbb{1} + \omega, \boldsymbol{a}) \to \hat{U}(\Lambda, \boldsymbol{a})\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon}) \left(\hat{U}(\Lambda, \boldsymbol{a})\right)^{-1} \tag{6.3.7}$$

To find the inverse of $\hat{U}(\Lambda, \boldsymbol{a})$ we use the composition rule according to equation (6.1.15) replacing T with $\hat{U}$ and using the fact

$$\left(\hat{U}(\Lambda, \boldsymbol{a})\right)^{-1} \hat{U}(\Lambda, \boldsymbol{a}) = \hat{U}(\mathbb{1}, \boldsymbol{0}) \tag{6.3.8}$$

due to the group axioms. This yields

$$\begin{aligned} &\hat{U}(\Lambda^{-1}, -\Lambda^{-1}\boldsymbol{a})\hat{U}(\Lambda, \boldsymbol{a}) = \hat{U}(\mathbb{1}, \boldsymbol{0}) \\ &\Rightarrow \left(\hat{U}(\Lambda, \boldsymbol{a})\right)^{-1} = \hat{U}(\Lambda^{-1}, -\Lambda^{-1}\boldsymbol{a}). \end{aligned} \tag{6.3.9}$$

Now using the composition rule (6.1.15) twice and (6.3.9) in equation (6.3.7), this gives us that

$$\begin{aligned} &\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon}) \to \hat{U}(\Lambda, \boldsymbol{a})\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon})\hat{U}(\Lambda^{-1}, -\Lambda^{-1}\boldsymbol{a}) = \\ &= \hat{U}(\mathbb{1} + \Lambda\omega\Lambda^{-1}, -\boldsymbol{a} - \Lambda\omega\Lambda^{-1}\boldsymbol{a} + \Lambda\boldsymbol{\epsilon} + \boldsymbol{a}) = \\ &= \hat{U}(\mathbb{1} + \Lambda\omega\Lambda^{-1}, -\Lambda\omega\Lambda^{-1}\boldsymbol{a} + \Lambda\boldsymbol{\epsilon}). \end{aligned} \tag{6.3.10}$$

To find the relations between the operators $\hat{J}$ and $\hat{\boldsymbol{P}}$ we expand both sides of equation (6.3.10) in the same manner as we did when we first established these operators in equation (6.3.4) (by expanding the operator $\hat{U}(\mathbb{1} + \omega, \boldsymbol{\epsilon})$ in terms of ω and $\boldsymbol{\epsilon}$). For the left-hand side, LHS, we get

$$\text{LHS} = \hat{U}(\Lambda, \boldsymbol{a}) \left(\frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} - i\epsilon_\rho\hat{P}^\rho\right) \hat{U}\left(\Lambda^{-1}, -\Lambda^{-1}\boldsymbol{a}\right) \tag{6.3.11}$$

where higher order terms in ω and $\boldsymbol{\epsilon}$ are absent. Equally the RHS of equation (6.3.10) can be expanded as

$$\text{RHS} = \frac{i}{2}\left(\Lambda\omega\Lambda^{-1}\right)_{\rho\sigma}\hat{J}^{\rho\sigma} - i\left(\Lambda\boldsymbol{\epsilon} - \Lambda\omega\Lambda^{-1}a\right)_\rho\hat{P}^\rho. \tag{6.3.12}$$

Note that the "$\mathbb{1}$" has been removed from both (6.3.11) and (6.3.12) since it appears on both sides unaltered. Well, since LHS = RHS in equation (6.3.10), this signifies that the coefficients in front of $\boldsymbol{\epsilon}$ and ω must be equal in both of equations (6.3.11) and (6.3.12). However, caution must be taken due to the fact that we are dealing with operators and symmetry vs. antisymmetry in indices matter so we cannot simply remove the $\boldsymbol{\epsilon}$ and ω and obtain the coefficients. Still there are neat ways around this problem. Let us start with the tricky one first, ω, and set the coefficients of equations (6.3.11) and (6.3.12) equal. Everything

which is not an operator can be shuffled around for convenience. Using equation (6.1.8) to account for the inverse Λ on the RHS we find in calculating coefficients that

$$\begin{aligned}&\omega_{\mu\nu}\left(\frac{i}{2}\hat{U}(\Lambda,\,\boldsymbol{a})\hat{J}^{\rho\sigma}\hat{U}\left(\Lambda^{-1},\,-\Lambda^{-1}\boldsymbol{a}\right)\right)=\\&=\omega_{\mu\nu}i\Lambda_{\rho}{}^{\mu}\Lambda_{\sigma}{}^{\nu}\left(\frac{1}{2}\hat{J}^{\rho\sigma}-\frac{1}{2}\left(a^{\rho}\hat{P}^{\sigma}+a^{\sigma}\hat{P}^{\rho}\right)\right).\end{aligned}\tag{6.3.13}$$

Here we encounter some difficulties. We have already affirmed the antisymmetric nature of $\hat{J}$ but nothing has been suggested about the properties of $\hat{\boldsymbol{P}}$ and $\boldsymbol{a}$ contracted with Λ and ω. Fortunately there is a way to decompose a tensor into a completely symmetric and completely antisymmetric part which we have used for $\hat{\boldsymbol{P}}$ contracted with Λ and ω above in equation (6.3.13). For a complete proof, please refer to appendix B. In two dimensions (where we will arrive after contracting Λ with $\hat{\boldsymbol{P}}$ and $\boldsymbol{a}$) we have for a tensor/matrix A that it can be decomposed according to

$$A_{\mu\nu}\overset{def}{=}A_{(\mu\nu)}+A_{[\mu\nu]}\tag{6.3.14}$$

where $A_{(\mu\nu)}$ is the completely symmetric part and $A_{[\mu\nu]}$ is the completely antisymmetric part. When performing operations on A these two parts never mix, and they are given by

$$A_{(\mu\nu)}\overset{def}{=}\frac{1}{2}\left(A_{\mu\nu}+A_{\nu\mu}\right)\;\;,\;\;A_{[\mu\nu]}\overset{def}{=}\frac{1}{2}\left(A_{\mu\nu}-A_{\nu\mu}\right).\tag{6.3.15}$$

Therefore, since ω is antisymmetric in indices and with the help of the reasoning from equation (6.3.5), we may establish that the symmetric part of the last $\boldsymbol{a},\hat{\boldsymbol{P}}$-expression of equation (6.3.12) is projected away when contracted with ω. Any symmetric part of Λ will also be eliminated when contracted with ω. Hence, only the antisymmetric part appears on the RHS of equation (6.3.13). Abbreviated it could be written as

$$\frac{1}{2}\left(a^{\sigma}\hat{P}^{\rho}-a^{\rho}\hat{P}^{\sigma}\right)=a^{[\rho}\hat{P}^{\sigma]}.\tag{6.3.16}$$

As a consequence we arrive at equation (6.3.13), which can be restated as

$$\begin{aligned}&\hat{U}(\Lambda,\,\boldsymbol{a})\hat{J}^{\mu\nu}\hat{U}\left(\Lambda^{-1},\,-\Lambda^{-1}\boldsymbol{a}\right)=\\&=\Lambda_{\rho}{}^{\mu}\Lambda_{\sigma}{}^{\nu}\left(\hat{J}^{\rho\sigma}-a^{\rho}\hat{P}^{\sigma}+a^{\sigma}\hat{P}^{\rho}\right).\end{aligned}\tag{6.3.17}$$

Thereafter, we follow the same procedure for the $\boldsymbol{\epsilon}$-coefficients of equations (6.3.11) and (6.3.12) in

$$\begin{aligned}&-i\epsilon_{\rho}\left(\hat{U}(\Lambda,\,\boldsymbol{a})\hat{P}^{\rho}\hat{U}\left(\Lambda^{-1},\,-\Lambda^{-1}\boldsymbol{a}\right)\right)=-i\epsilon_{\rho}\left(\Lambda_{\sigma}{}^{\rho}\hat{P}^{\rho}\right)\\&\Leftrightarrow\hat{U}(\Lambda,\,\boldsymbol{a})\hat{P}^{\rho}\hat{U}\left(\Lambda^{-1},\,-\Lambda^{-1}\boldsymbol{a}\right)=\Lambda_{\sigma}{}^{\rho}\hat{P}^{\sigma}.\end{aligned}\tag{6.3.18}$$

By performing these calculations, we have found the effect of the adjoint action of the linear unitary operator $\hat{U}(\Lambda,\,\boldsymbol{a})$ on the generators $\hat{J}$ and $\hat{\boldsymbol{P}}$ of the Poincaré group. If we linearize the operator $\hat{U}(\Lambda,\,\boldsymbol{a})$ so that it corresponds to an infinitesimal transformation away from the identity and take the adjoint action on our generators we are able to take the adjoint action on our Lie algebra with itself.

Once more let us set $\Lambda = \mathbb{1} + \omega$ and $\boldsymbol{a} = \boldsymbol{\epsilon}$ in the surrounding $\hat{U}(\Lambda, \boldsymbol{a})$ of equations (6.3.18) and (6.3.18). Note that these new ω and $\boldsymbol{\epsilon}$ are completely unrelated to the previous ones. The next issue to deal with is to locate the inverse of $\hat{U}$ since it appears in both (6.3.17) and (6.3.18) on the LHS with aid from equation (6.3.9). Replacing $\boldsymbol{a}$ with $\boldsymbol{\epsilon}$ is straightforward, and for Λ^{-1} we may find guidance offered by elementary calculus. Recall that in one-dimension

$$\frac{1}{1+\omega} = 1 - \omega + \dots \tag{6.3.19}$$

when $|\omega| < 1$ and where $\dots$ indicate terms of higher order. Correspondingly, we can deduce when dealing with an infinitesimal Lorentz transformation, that

$$\Lambda = \mathbb{1} + \omega \;\Rightarrow\; \Lambda^{-1} \approx \mathbb{1} - \omega. \tag{6.3.20}$$

With this in mind, equation (6.3.9) turns into

$$\hat{U}\left(\Lambda^{-1},\, -\Lambda^{-1}\boldsymbol{a}\right) = \hat{U}(\mathbb{1} - \omega,\, -\boldsymbol{\epsilon}) \tag{6.3.21}$$

where we have used the fact that

$$-\Lambda^{-1}\boldsymbol{a} \approx -(\mathbb{1} - \omega)\boldsymbol{\epsilon} \approx -\boldsymbol{\epsilon} \tag{6.3.22}$$

since both ω and $\boldsymbol{\epsilon}$ are infinitesimal. Next, let us express $\hat{U}(\mathbb{1} + \omega,\, \boldsymbol{\epsilon})$ and its inverse $\hat{U}(\mathbb{1} - \omega,\, -\boldsymbol{\epsilon})$ in terms of ω and $\boldsymbol{\epsilon}$ in the same fashion as we did in equation (6.3.4).

$$\begin{aligned} \hat{U}(\mathbb{1} + \omega,\, \boldsymbol{\epsilon}) &= \hat{1} + \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} - i\epsilon_\rho \hat{P}^\rho \\ \hat{U}(\mathbb{1} - \omega,\, -\boldsymbol{\epsilon}) &= \hat{1} - \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} + i\epsilon_\rho \hat{P}^\rho \end{aligned} \tag{6.3.23}$$

Notice that the expressions in (6.3.23) for $\hat{U}$ and its inverse only differ in signs for the coefficients of ω and $\boldsymbol{\epsilon}$. With the goal in sight let us expand the LHS of (6.3.17) by replacing $\hat{U}(\Lambda,\, \boldsymbol{a})$ with $\hat{U}(\mathbb{1}+\omega,\, \boldsymbol{\epsilon})$. Caution must be taken when dealing with operators and we cannot move them around arbitrary.

$$\begin{aligned} &\hat{U}(\mathbb{1} + \omega,\, \boldsymbol{\epsilon})\hat{J}^{\mu\nu}\hat{U}\,(\mathbb{1} - \omega,\, -\boldsymbol{\epsilon}) = \\ &= \left(\hat{1} + \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} - i\epsilon_\lambda \hat{P}^\lambda\right)\hat{J}^{\mu\nu}\left(\hat{1} - \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} + i\epsilon_\lambda \hat{P}^\lambda\right) = \\ &= \left(\hat{J}^{\mu\nu} + \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma}\hat{J}^{\mu\nu} - i\epsilon_\lambda \hat{P}^\lambda \hat{J}^{\mu\nu}\right)\left(\hat{1} - \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} + i\epsilon_\lambda \hat{P}^\lambda\right) = \\ &= /\text{Skipping all terms of higher order in } \omega \text{ and } \epsilon./ = \\ &= \hat{J}^{\mu\nu} + \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma}\hat{J}^{\mu\nu} - i\epsilon_\lambda \hat{P}^\lambda \hat{J}^{\mu\nu} - \frac{i}{2}\omega_{\rho\sigma}\hat{J}^{\mu\nu}\hat{J}^{\rho\sigma} + i\epsilon_\lambda \hat{J}^{\mu\nu}\hat{P}^\lambda = \\ &= /\text{Simplify the expression by introducing the commutator.}/ = \\ &= \hat{J}^{\mu\nu} + i\left[\frac{1}{2}\omega_{\rho\sigma}\hat{J}^{\rho\sigma} - \epsilon_\lambda \hat{P}^\lambda,\, \hat{J}^{\mu\nu}\right]. \end{aligned} \tag{6.3.24}$$

Let us proceed to the right-hand side of (6.3.17), omitting terms of higher order.

$$
\begin{aligned}
&\Lambda_\rho{}^\mu \Lambda_\sigma{}^\nu \left(\hat{J}^{\rho\sigma} - a^\rho \hat{P}^\sigma + a^\sigma \hat{P}^\rho \right) = \\
&= \left(\delta_\rho{}^\mu + \omega_\rho{}^\mu \right) \left(\delta_\sigma{}^\nu + \omega_\sigma{}^\nu \right) \left(\hat{J}^{\rho\sigma} - \epsilon^\rho \hat{P}^\sigma + \epsilon^\sigma \hat{P}^\rho \right) = \\
&= \left(\delta_\rho{}^\mu \delta_\sigma{}^\nu + \delta_\rho{}^\mu \omega_\sigma{}^\nu + \delta_\sigma{}^\nu \omega_\rho{}^\mu \right) \left(\hat{J}^{\rho\sigma} - \epsilon^\rho \hat{P}^\sigma + \epsilon^\sigma \hat{P}^\rho \right).
\end{aligned} \tag{6.3.25}
$$

This expression can easily be simplified since the result obtained by combining the two Kronecker deltas with the operator $\hat{J}$ can be abbreviated through

$$
\delta_\rho{}^\mu \delta_\sigma{}^\nu \hat{J}^{\rho\sigma} = \hat{J}^{\mu\nu} \tag{6.3.26}
$$

since the components of the identity matrix are only separated from zero on the diagonal. By making similar observations for the other components of equation (6.3.25), this yields that

$$
(6.3.25) = \hat{J}^{\mu\nu} - \epsilon^\mu \hat{P}^\nu + \epsilon^\nu \hat{P}^\mu + \omega_\rho{}^\mu \hat{J}^{\rho\nu} + \omega_\sigma{}^\nu \hat{J}^{\mu\sigma}. \tag{6.3.27}
$$

Finally setting the expanded LHS as of equation (6.3.24) and RHS as of equation (6.3.27) equal to one another, and realizing that $\hat{J}^{\mu\nu}$ appear on both sides, gives

$$
i \left[\frac{1}{2} \omega_{\rho\sigma} \hat{J}^{\rho\sigma} - \epsilon_\lambda \hat{P}^\lambda,\, \hat{J}^{\mu\nu} \right] = -\epsilon^\mu \hat{P}^\nu + \epsilon^\nu \hat{P}^\mu + \omega_\rho{}^\mu \hat{J}^{\rho\nu} + \omega_\sigma{}^\nu \hat{J}^{\mu\sigma}. \tag{6.3.28}
$$

Equivalently, if we examine and expand the last line of equation (6.3.18) in order to find the $\hat{\boldsymbol{P}}$-relations

$$
\hat{U}(\mathbb{1} + \omega,\, \boldsymbol{\epsilon}) \hat{P}^\rho \hat{U}\, (\mathbb{1} - \omega,\, -\boldsymbol{\epsilon}) = \Lambda_\sigma{}^\rho \hat{P}^\sigma \tag{6.3.29}
$$

and perform similar manipulations, we arrive at

$$
i \left[\frac{1}{2} \omega_{\rho\sigma} \hat{J}^{\rho\sigma} - \epsilon_\lambda \hat{P}^\lambda,\, \hat{P}^\mu \right] = \omega_\nu{}^\mu \hat{P}^\nu. \tag{6.3.30}
$$

At last there remains only to equate the coefficients in front of the components of $\omega_{\rho\sigma}$ and ϵ_ρ in equations (6.3.28) and (6.3.30). Let us look at the coefficients in front of the components of $\omega_{\rho\sigma}$ from the first equation, (6.3.28).

$$
\frac{i}{2} \omega_{\rho\sigma} \left(\hat{J}^{\rho\sigma} \hat{J}^{\mu\nu} - \hat{J}^{\mu\nu} \hat{J}^{\rho\sigma} \right) = \omega_\rho{}^\mu \hat{J}^{\rho\nu} + \omega_\sigma{}^\nu \hat{J}^{\mu\sigma}. \tag{6.3.31}
$$

This can be restated by introducing a commutator and a Minkowski metric tensor as

$$
\begin{aligned}
&\frac{i}{2} \omega_{\rho\sigma} \left[\hat{J}^{\rho\sigma},\, \hat{J}^{\mu\nu} \right] = \omega_{\rho\sigma} \eta^{\sigma\mu} \hat{J}^{\rho\nu} + \omega_{\sigma\rho} \eta^{\nu\rho} \hat{J}^{\mu\sigma} = \\
&= \omega_{\rho\sigma} \left(\eta^{\sigma\mu} \hat{J}^{\rho\nu} - \eta^{\nu\rho} \hat{J}^{\mu\sigma} \right) = \\
&= /\omega_{\rho\sigma} \text{ is antisymmetric in indices.}/ = \\
&= \frac{1}{2} \omega_{\rho\sigma} \left(\eta^{\sigma\mu} \hat{J}^{\rho\nu} - \eta^{\rho\mu} \hat{J}^{\sigma\nu} - \eta^{\nu\rho} \hat{J}^{\mu\sigma} + \eta^{\nu\sigma} \hat{J}^{\mu\rho} \right).
\end{aligned} \tag{6.3.32}
$$

By equating the coefficients in front of ϵ_ρ, we finally arrive at the commutation relations between the generators of the Lie algebra for the defining representation of the Poincaré group.

$$
\begin{aligned}
i\left[\hat{J}^{\rho\sigma},\, \hat{J}^{\mu\nu}\right] &= \eta^{\sigma\mu}\hat{J}^{\rho\nu} - \eta^{\rho\mu}\hat{J}^{\sigma\nu} - \eta^{\nu\rho}\hat{J}^{\mu\sigma} + \eta^{\nu\sigma}\hat{J}^{\mu\rho} \\
i\left[\hat{P}^{\rho},\, \hat{J}^{\mu\nu}\right] &= \eta^{\rho\mu}\hat{P}^{\nu} - \eta^{\rho\nu}\hat{P}^{\mu} \\
\left[\hat{P}^{\rho},\, \hat{P}^{\mu}\right] &= 0
\end{aligned}
\tag{6.3.33}
$$

These equations encode the **Lie algebra for the defining representation of the Poincaré group**, $\mathfrak{iso}(1,\,3)$. Notice that we obtained the generator $\hat{J}$ as a coefficient for a rotational matrix and that it will thus yield rotations. Moreover, $\hat{\boldsymbol{P}}$ was acquired from the infinitesimal translation $\boldsymbol{\epsilon}$ and is accordingly the generator for translations. With the Lorentz group we just describe rotations in spacetime and its Lie algebra is the first line of equation (6.3.33). The last line of (6.3.33) tells us that translations in time and space (time being the 0:th component of a 4-vector, the spatial directions the three others) commutes. Interestingly, on the second line of (6.3.33), we find through scrutiny that a translation followed by a rotation simply yields another translation.

From these commutation relations we can conclude some facts about the composition of the Poincaré group. Clearly the Lorentz group $SO(1,\,3)$ can be considered as a subgroup where we only account for the rotations. In addition, the Poincaré group describes translations in a 4-dimensional space. Since these translations are not affected by the Minkowski metric, the translational group in 4 dimensions $\mathbb{R}^4$ can be used. If the Poincaré group, $ISO(1,\,3)$, were a direct product of the Lorentz group and $\mathbb{R}^4$ there should not be any interaction between the generators for translations and the generators for rotations. Instead the Poincaré group is a semidirect product of the two.

$$
ISO(1,3) = SO(1,\,3) \rtimes \mathbb{R}^4. \tag{6.3.34}
$$

Another, slightly more evident way to see this on the group level is to recognize that in order to have a direct product both subgroups must be normal. By taking the adjoint action of the full Poincaré group on the translational group $\mathbb{R}^4$ we find upon using the composition rule for the symmetry transformations of the Poincaré group and the inverse of a unitary transformation and dropping the T:s and $\hat{U}$:s, respectively, for convenience that

$$
(\Lambda,\, \boldsymbol{a})(0,\, \boldsymbol{a})(\Lambda,\, \boldsymbol{a})^{-1} \overset{(6.3.9)}{=} (\Lambda,\, \boldsymbol{a})(0,\, \boldsymbol{a})(\Lambda^{-1},\, -\Lambda^{-1}\boldsymbol{a}) \overset{(6.1.15)}{=} (0,\, \Lambda\boldsymbol{a} + \boldsymbol{a}) \tag{6.3.35}
$$

which is another pure translation in $\mathbb{R}^4$. But, for the homogeneous Lorentz subgroup $SO(1,\,3)$ the same reasoning with an element $(\tilde{\Lambda},\, \mathbf{0}) \in SO(1,\,3)$ results in

$$
(\Lambda,\, \boldsymbol{a})(\Lambda,\, \mathbf{0})(\Lambda,\, \boldsymbol{a})^{-1} = (\Lambda,\, \boldsymbol{a})(\mathbb{1},\, -\boldsymbol{a}) = (\Lambda,\, -\Lambda\boldsymbol{a} + \boldsymbol{a}) \tag{6.3.36}
$$

and the result is a mixture of a rotation and a translation, which takes us out of the homogeneous Lorentz group. Hence, the Lorentz group is not a normal subgroup of the Poincaré group and we have a semidirect product, which is used when one or several subgroups to a group is not normal. Refer to the definition in chapter 2.

6.4 Implications of the Poincaré Algebra

From Noether's theorem, we know that for every continuous symmetry there is a conserved quantity and vice versa. The "quantum version" of this theorem states that:

Theorem 6.4.1. A continuous symmetry transformation with parameter α acts on vectors in Hilbert space through a unitary operator $\hat{U} = \exp(i\alpha\hat{A})$ where $\hat{A}$ is the self-adjoint/Hermitian operator for the corresponding conserved quantity A ([19], p. 70).

If we look at the Lie algebra for the Poincaré group (6.3.33), we have already established that the Hermitian operator $\hat{P}^0$ (the 0:th component of $\hat{\boldsymbol{P}}$) encodes translations in time, induced on vectors in Hilbert space through a unitary operator $\hat{U}$. For time invariance, the conserved quantity is energy. Hence, we deduce that $\hat{P}^0$ is in fact the **Hamilton operator** $\hat{H}$:

$$\hat{P}^0 \stackrel{def}{=} \hat{H}. \tag{6.4.1}$$

The corresponding unitary operator for finite time translations is then

$$\hat{U}\left(\mathbb{1},\, (t,\, 0,\, 0,\, 0)\right) = \exp\left(-\frac{i}{\hbar}\hat{P}^0 t\right) = \exp\left(-\frac{i}{\hbar}\hat{H}t\right) \tag{6.4.2}$$

where t is the time-parameter (c the speed of light set to 1) and $\mathbb{1}$ indicates that we are not rotating the vectors. The $\hbar$ is introduced as a constant to make the theory agree with experimental results. When performing these calculations, one usually sets $\hbar = 1$ and the inconvenience is avoided. The minus sign comes from our definition of the Minkowski metric tensor and hence the operator represents translations forward in time. Consequently we arrive at the same expression as in section 4.2.

We distinguish conserved quantities in quantum mechanics from the fact that their corresponding operators commute with the Hamilton operator. If we investigate the commutation relations between the generators of the Poincaré algebra in expression (6.3.33), we discover that all components $\hat{P}^\mu$ commute with $\hat{P}^0$ (from the last line). Since $\hat{P}^i$ where $i = 1, 2, 3$ will generate spatial translations, they are the **momentum operators** of the x, y and z-directions respectively. From them we can write down a 3-momentum operator $\hat{\boldsymbol{p}}$ according to equation (6.4.3).

$$\hat{\boldsymbol{p}} = \left(\hat{P}^1,\, \hat{P}^2,\, \hat{P}^3\right) = (\hat{p}_x,\, \hat{p}_y,\, \hat{p}_z) \tag{6.4.3}$$

Likewise, from the middle line of (6.3.33), we see that the components $\hat{J}^{23}$, $\hat{J}^{31}$, $\hat{J}^{12}$ of the $\hat{J}$ 4×4 matrix, where the elements on the diagonal are denoted $\hat{J}^{00}$, …, $\hat{J}^{33}$, commute with the Hamilton operator. These operators generate rotations and, since angular momentum is the conserved quantity for a system with rotational symmetry, they are the **angular momentum operators** for rotations with the axis pointing in the x, y and z-directions. They can be thought of as an operator for angular momentum $\hat{\boldsymbol{J}}$ as of equation (6.4.4), which we have encountered numerous times. Although it only has three components, it is written with a capital $\hat{\boldsymbol{J}}$ due to historical convention.

$$\hat{\boldsymbol{J}} = \left(\hat{J}^{23},\, \hat{J}^{31},\, \hat{J}^{12}\right) = \left(\hat{J}_x,\, \hat{J}_y,\, \hat{J}_z\right) \tag{6.4.4}$$

Since the generator $\hat{J}$ (the matrix, not the vector) is antisymmetric in indices, the same goes for the components with switched indices. The remaining components of $\hat{J}$, $\left(\hat{J}^{10},\, \hat{J}^{20},\, \hat{J}^{30}\right)$ which do not commute with the Hamilton operator are referred to as **boosts** as they correspond to accelerations. They can be placed together in an operator $\hat{\boldsymbol{K}}$, see equation (6.4.5), for later use.

$$\hat{\boldsymbol{K}} = \left(\hat{J}^{10},\, \hat{J}^{20},\, \hat{J}^{30}\right) \tag{6.4.5}$$

Now let us find the commutation relations between the operators in equations (6.4.3), (6.4.4) and (6.4.5). Here, the target is to examine 3-vectors so subscript indices $i = 1,\, 2,\, 3$ are used. Firstly, we have established that

$$\left[\hat{p}_i,\, \hat{H}\right] = \left[\hat{J}_i,\, \hat{H}\right] = \left[\hat{H},\, \hat{H}\right] = 0 \tag{6.4.6}$$

whereas for the other components

$$\begin{aligned} \left[\hat{J}_i,\, \hat{J}_j\right] &= i\epsilon_{ijk}\hat{J}_k \qquad \left[\hat{J}_i,\, \hat{p}_j\right] = i\epsilon_{ijk}\hat{p}_k \\ \left[\hat{J}_i,\, \hat{K}_j\right] &= i\epsilon_{ijk}\hat{K}_k \qquad \left[\hat{K}_i,\, \hat{p}_j\right] = i\hat{H}\delta_{ij} \\ \left[\hat{K}_i,\, \hat{K}_j\right] &= -i\epsilon_{ijk}\hat{J}_k \qquad \left[\hat{K}_i,\, \hat{H}\right] = i\hat{p}_i. \end{aligned} \tag{6.4.7}$$

Astoundingly, we once more encounter the fundamental relations for angular momentum on the first line of equation (6.4.7), the same as equation (4.2.10) from section 4.2 albeit with $\hbar = 1$. With their aid we can construct the Lie algebra $\mathfrak{su}(2)$ for the groups $SU(2)$ and $SO(3)$. By counting the generators we can also discern the dimension of the algebra of the defining representation. The generators of the Lorentz group are the three generators of angular momentum as well as the boosts in the x, y and z-directions, all together 6 generators creating a 6-dimensional algebra. For the Poincaré group, four additional generators, the $\hat{P}^\mu$, are required to account for the translations in spacetime and hence we arrive at totally 10 generators and a 10-dimensional algebra.

Moreover, we are finally able to determine a few other familiar operators. First, there is a generalization of the operator in equation (6.4.2), the linear, unitary **spacetime translation operator** which represents finite translations on physical Hilbert space.

$$\hat{U}(\mathbb{1}, \boldsymbol{a}) = \exp\left(-\frac{i}{\hbar}\hat{P}^\mu a_\mu\right) \tag{6.4.8}$$

where

$$\hat{\boldsymbol{P}} = \left(\hat{H},\, \hat{\boldsymbol{p}}\right) \text{ and } \boldsymbol{a} = (t,\, x,\, y,\, z)\,. \tag{6.4.9}$$

In the same manner, we have for an arbitrary rotational matrix, as discussed in section 4.1, $R(\boldsymbol{\theta})$ which describes a rotation $\theta = |\boldsymbol{\theta}|$ around the direction of $\boldsymbol{\theta}$ that the rotation on physical Hilbert space is represented by the **rotational operator** in equation (6.4.10).

$$\hat{U}\left(R(\boldsymbol{\theta}),\, \mathbf{0}\right) = \exp\left(\frac{i}{\hbar}\hat{J}_i\theta^i\right) \tag{6.4.10}$$

With these tools let us proceed to investigate how one may classify elementary particles according to their behavior when in contact with a representation of the Lorentz or the Poincaré group. In addition we will reveal several notions about the different representations of the Lorentz and Poincaré group, especially the crucial role for the little group and the pairing of the Lie algebra of two $\mathfrak{su}(2)$:s with $\mathfrak{so}(1,\, 3)$.

6.5 Representations of the Lorentz and Poincaré Groups

In this section we will find some of the most basic finite and infinite representations for the Lorentz and Poincaré groups. As a start, let us make a change of basis for the Lorentz algebra. We define

$$\begin{aligned}\hat{J}_i^+ &\stackrel{def}{=} \frac{1}{2}(\hat{J}_i + i\hat{K}_i)\\ \hat{J}_i^- &\stackrel{def}{=} \frac{1}{2}(\hat{J}_i - i\hat{K}_i).\end{aligned} \tag{6.5.1}$$

and investigate how these generators commute. Consider

$$\begin{aligned}&[\hat{J}_i^+, \hat{J}_j^+] = [\frac{1}{2}(\hat{J}_i + i\hat{K}_i), \frac{1}{2}(\hat{J}_j + i\hat{K}_j)] = \frac{1}{4}([\hat{J}_i, \hat{J}_i] + i[\hat{J}_i, \hat{K}_j] + i[\hat{K}_i, \hat{J}_j] - [\hat{K}_i, \hat{K}_j]) \stackrel{(6.4.7)}{=}\\ &= \frac{1}{4}(i\epsilon_{ijk}\hat{J}_k - \epsilon_{ijk}\hat{K}_k + \epsilon_{jik}\hat{K}_k + i\epsilon_{ijk}\hat{J}_k) = \frac{1}{2}i\epsilon_{ijk}(\hat{J}_k + i\hat{K}_k) = \frac{1}{2}i\epsilon_{ijk}\hat{J}_k^+.\end{aligned} \tag{6.5.2}$$

The calculation for $[\hat{J}_i^-, \hat{J}_j^-]$ is completely analogous. Next, consider $[\hat{J}_i^+, \hat{J}_j^-]$

$$\begin{aligned}&[\hat{J}_i^+, \hat{J}_j^-] = \frac{1}{4}[\hat{J}_i + i\hat{K}_i, \hat{J}_j - i\hat{K}_j] = \frac{1}{4}([\hat{J}_i, \hat{J}_j] - i[\hat{J}_i, \hat{K}_j] + i[\hat{K}_i, \hat{J}_j] + [\hat{K}_i, \hat{K}_j]) =\\ &= \frac{1}{4}(i\epsilon_{ijk}\hat{J}_k + \epsilon_{ijk}\hat{K}_k - \epsilon_{ijk}\hat{K}_k - i\epsilon_{ijk}\hat{J}_k) = 0.\end{aligned} \tag{6.5.3}$$

So, what we have found is that

$$\begin{aligned}[\hat{J}_i^+, \hat{J}_j^+] &= \frac{1}{2}i\epsilon_{ijk}\hat{J}_k^+\\ [\hat{J}_i^-, \hat{J}_j^-] &= \frac{1}{2}i\epsilon_{ijk}\hat{J}_k^-,\\ [\hat{J}_i^+, \hat{J}_j^-] &= 0\end{aligned} \tag{6.5.4}$$

i.e. the Lorentz algebra can be seen as two $\mathfrak{su}(2)$ algebras,

$$\mathfrak{so}(1,3) = \mathfrak{su}(2) \oplus \mathfrak{su}(2)^* \tag{6.5.5}$$

Representations for $SU(2)$ are usually labeled with the spin quantum number J as (J), or $(2J+1)$, since the module consists of all the states with a particular spin. Due to the fact that $\hat{J}_i = \hat{J}_i^+ + \hat{J}_i^-$, we denote a Lorentz representation with two integers (J_1, J_2) corresponding to spin J. In analogy with *addition of angular momenta*, J takes the values $|J_1 - J_2|, ..., J_1 + J_2$. To see how this works, we take a look at some examples of finite dimensional representations of the Lorentz group inspired by [20].

Example 6.5.1. The trivial representation. In this 1-dimensional representation, every element $\hat{J}^{\mu\nu}$ in the Lie algebra is represented with 0 and every group element Λ with 1. The module which these operators act on is a 1-dimensional vector space spanned by 1-component objects ϕ called Lorentz scalars. These transform under $SO(1,3)$ as

$$\Lambda\phi = 1 \cdot \phi, \tag{6.5.6}$$

where $\Lambda = \exp(-i\omega_{\mu\nu}\hat{J}^{\mu\nu}/2)$ and $\hat{J}^{\mu\nu} = 0$. This representation is denoted $(0,\ 0)$.

Example 6.5.2. The spinorial representation. The double cover of $SO(n,m)$ is $Spin(n,m)$, so for $SO(1,3)$, the double cover is $Spin(1,3)$. Moreover, $Spin(1,3) \cong SL(2,\mathbb{C})$. These relations will be developped in section 6.7.2. Thus, representations for $SL(2,\mathbb{C})$ serve as representations for $SO(1,3)$ as well. In particular, the fundamental representation for $SL(2,\mathbb{C})$ gives us the spinorial representation for $SO(1,3)$. This representation is 2-dimensional with a 2-dimensional vector space, the module, spanned by 2-component objects called **left handed spinors** ψ_α, denoted $(1/2,\ 0)$. If we take the complex conjugate representation of $SL(2,\mathbb{C})$, i.e. we take the complex conjugate of our representation matrices $\mathcal{M}$, we obtain the so called anti-fundamental representation for $SL(2,\mathbb{C})$ with a module spanned by the **right handed spinors** $\bar{\chi}^{\dot{\alpha}}$, denoted $(0,\ 1/2)$. We can think of $(0,\ 1/2)$ as the "complex conjugated representation" of $(1/2,\ 0)$. These representations are *not* equivalent, i.e. $\mathcal{M}^* \neq S\mathcal{M}S^{-1}$ by any matrix S. The left handed spinor ψ_α, $\alpha = 1,2$, transforms as

$$\psi_\alpha \longrightarrow \mathcal{M}_\alpha{}^\beta \psi_\beta, \quad \mathcal{M} \in SL(2,\mathbb{C}) \tag{6.5.7}$$

under the Lorentz group where $\mathcal{M} = \exp(-i\omega_{\mu\nu}S^{\mu\nu}/2))$ and $S^{\mu\nu}$ is a 2×2 matrix representation for $SL(2,\mathbb{C})$. In terms of the Pauli matrices, $S^{ij} = \frac{i}{2}\epsilon^{ijk}\sigma^k$ and $S^{0i} = -\frac{i}{2}\sigma^i$. The complex conjugate representation is defined as

$$\bar{\psi}_{\dot{\alpha}} \longrightarrow (\mathcal{M}^*)_{\dot{\alpha}}{}^{\dot{\beta}} \bar{\psi}_{\dot{\beta}}, \quad \mathcal{M}^* \in SL(2,\mathbb{C})^*. \tag{6.5.8}$$

To see how $\bar{\psi}_{\dot{\alpha}}$ is related to ψ_α we take the complex conjugate of (6.5.7), i.e. $\psi^*_\alpha \longrightarrow (\mathcal{M}^*)_\alpha{}^\beta \psi^*_\beta$, and compare this with (6.5.8) which gives $\psi^*_\alpha = \psi_{\dot{\alpha}}$. Next, we want to see what $\psi_{\dot{\alpha}}$ has to do with the right handed spinor $\bar{\chi}^\alpha$. To do this we have to introduce the $SL(2,\mathbb{C})$ invariant tensor ϵ defined[1] as

$$\epsilon_{\alpha\beta} = \epsilon_{\dot{\alpha}\dot{\beta}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \epsilon^{\alpha\beta} = \epsilon^{\dot{\alpha}\dot{\beta}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}. \tag{6.5.9}$$

Since $\det\mathcal{M} = 1$ it follows that

$$\epsilon'^{\alpha\beta} = \epsilon^{\rho\sigma}\mathcal{M}_\rho{}^\alpha \mathcal{M}_\sigma{}^\beta = \epsilon^{\alpha\beta}\det\mathcal{M} = \epsilon^{\alpha\beta}, \tag{6.5.10}$$

i.e. invariant under $SL(2,\mathbb{C})$. With these tensors we can contract indices as $\psi^\alpha = \epsilon^{\alpha\beta}\psi_\beta$, $\psi_\alpha = \epsilon_{\alpha\beta}\psi^\beta$ and $\bar{\psi}^{\dot{\alpha}} = \epsilon^{\dot{\alpha}\dot{\beta}}\bar{\psi}_{\dot{\beta}}$, $\bar{\psi}_{\dot{\alpha}} = \epsilon_{\dot{\alpha}\dot{\beta}}\bar{\psi}^{\dot{\beta}}$. To see how ψ^α transform, consider

$$\epsilon_{\alpha\sigma}\psi'^\sigma = \mathcal{M}_\alpha{}^\beta \epsilon_{\beta\gamma}\psi^\gamma \Rightarrow \psi'^\sigma = \epsilon^{\sigma\alpha}\mathcal{M}_\alpha{}^\beta \epsilon_{\beta\gamma}\psi^\gamma. \tag{6.5.11}$$

Rearrange (6.5.10) as

$$\begin{aligned} \epsilon^{\sigma\alpha}\mathcal{M}_\alpha{}^\beta \mathcal{M}_\sigma{}^\gamma = \epsilon^{\beta\gamma} \Leftrightarrow \\ \epsilon^{\sigma\alpha}\mathcal{M}_\alpha{}^\beta (\mathcal{M}^T)^\gamma{}_\sigma = \epsilon^{\beta\gamma} \Leftrightarrow \\ \epsilon^{\sigma\alpha}\mathcal{M}_\alpha{}^\beta \epsilon_{\beta\gamma} = ((\mathcal{M}^T)^{-1})^\sigma{}_\gamma. \end{aligned} \tag{6.5.12}$$

Note that we have changed the indices. These two expressions together yield

$$\psi'^\sigma = ((\mathcal{M}^T)^{-1})^\sigma{}_\gamma \psi^\gamma = (\mathcal{M}^{-1})_\gamma{}^\sigma \psi^\gamma \tag{6.5.13}$$

[1]Of course, $\epsilon_{\alpha\beta}$ is a component of ϵ and not the entire matrix, but we present it like this to distinguish between the upper and lower indices.

and similarly, $\bar{\psi}^{\dot{\alpha}}$ transforms as $\bar{\psi}^{\dot{\alpha}} = (\mathcal{M}^{*-1})_{\dot{\beta}}{}^{\dot{\alpha}}\bar{\psi}^{\dot{\beta}}$ which is the right handed spinor[2], i.e. $\bar{\psi}^{\dot{\alpha}} \stackrel{def}{=} \bar{\chi}^{\dot{\alpha}}$. Finally, the right handed spinor $\bar{\chi}^{\dot{\alpha}}$, $\alpha = 1, 2$, transforms as

$$\bar{\chi}^{\dot{\alpha}} \longrightarrow (\mathcal{M}^{*-1})_{\dot{\beta}}{}^{\dot{\alpha}}\bar{\chi}^{\dot{\beta}} \quad \mathcal{M}^* \in SL(2,\mathbb{C})^*, \tag{6.5.14}$$

where $\mathcal{M}^* = \exp(-i\omega_{\mu\nu}S^{\mu\nu}/2))$ and $S^{\mu\nu}$ is a 2×2 matrix representation for $SL(2, \mathbb{C})$. $S^{\mu\nu}$ is the same as for the left handed spionor except that $S^{0i} = \frac{i}{2}\sigma^i$. Spinors were encountered in section 4.3.1 as 1/2-spin states. These spinors are the module for $SU(2) \cong Spin(3)$, i.e. the double cover of $SO(3)$ which is a subgroup of spatial rotations to $SO(1, 3)$. We can for example take the electron as our 1/2-spin particle where the spinors

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{6.5.15}$$

corresponding to a spin up and spin down state of the electron, constitute a basis for the module.

Example 6.5.3. The vector representation. In the vector representation each element $\hat{J}^{\mu\nu}$ is represented by a 4×4 matrix $\mathcal{J}^{\mu\nu}$ acting on the 4-dimensional vector space spanned by 4-vectors x called Lorentz 4-vectors. These 4-vectors transform as

$$x^\rho \longrightarrow (\exp(-i\omega_{\mu\nu}\mathcal{J}^{\mu\nu}/2))^\rho{}_\sigma x^\sigma \tag{6.5.16}$$

under the Lorentz group. The vector representation is the fundamental representation for the Lorentz group since $SO(1, 3)$ is defined as the group of 4×4 matrices acting on a 4-dimensional spacetime. Thus have that $\Lambda^\rho{}_\sigma = (\exp(-i\omega_{\mu\nu}\mathcal{J}^{\mu\nu}/2))^\rho{}_\sigma$ with $\hat{J}^{\mu\nu}$ represented as a 4×4 matrix $\mathcal{J}^{\mu\nu}$. This representation is denoted $(1/2,\ 1/2)$ and constitutes an irreducible representation since time and space mix in a general transformation. An example of an object transforming under the vector representation is the electromagnetic vector potential.

Example 6.5.4. The tensor representation. The tensor representation acts on a vector space spanned by tensors of a particular rank. For every index there must be a transformation matrix Λ. A rank 2 tensor $T^{\rho\sigma}$ transforms as

$$T^{\rho\sigma} \longrightarrow \Lambda^\rho{}_{\rho'}\Lambda^\sigma{}_{\sigma'}T^{\rho'\sigma'}, \tag{6.5.17}$$

where $\Lambda^\rho{}_{\rho'}\Lambda^\sigma{}_{\sigma'}$ are vector representations, i.e. the fundamental representation. Thus, the tensor representation is the direct product of vector representations. If we define

$$\Lambda^\rho{}_{\rho'}\Lambda^\sigma{}_{\sigma'} \stackrel{def}{=} \Lambda^{\rho\sigma}{}_{\rho'\sigma'}, \tag{6.5.18}$$

(6.5.17) becomes

$$T^{\rho\sigma} \longrightarrow \Lambda^{\rho\sigma}{}_{\rho'\sigma'}T^{\rho'\sigma'}, \tag{6.5.19}$$

where $\Lambda^{\rho\sigma}{}_{\rho'\sigma'}$ can be imagined as a 16×16 matrix representation that acts on rank 2 tensors. It is not really a matrix since it is a rank 4 tensor, but for a given pair $\rho\sigma$ there is matrix with components $\rho'\sigma'$. $\Lambda^{\rho\sigma}{}_{\rho'\sigma'}$ together with the set of rank 2 tensors constitute a tensor representation.

[2]It is customary to denote the right handed spinor as $\bar{\chi}^{\dot{\alpha}}$.

Lorentz scalars, spinors, vectors and tensors are examples of finite dimensional representations for the Lorentz group. Since they do not depend on spacetime they are considered as constants.

Now, let us take a look at infinite representations. Consider a field $\Phi(x)$ that depends on the spacetime coordinates. Its components transform under the Lorentz group as

$$\Phi_a \longrightarrow M_a{}^b \Phi_b, \tag{6.5.20}$$

but the Lorentz group also acts on the coordinates in the usual way; $\Lambda^{\mu'}{}_{\mu} x^{\mu}$. There are two different types of symmetries in nature with significant differences; a symmetry related to coordinate transformations is said to be an **external symmetry** and a symmetry related to a transformation of the fields and *not* the coordinates is said to be an **internal symmetry**. We will talk more about internal symmetries in chapter 7. Lorentz invariance is an external symmetry, and because of the coordinate transformation we may say that we get an induced transformation of the field as well. With this taken into account, the transformation of the field $\Phi(x)$ becomes

$$\Phi_a(x) \longrightarrow M_a{}^b \Phi_b(\Lambda^{-1}x), \tag{6.5.21}$$

where M is one of the finite representations from above, i.e. $M = \exp(-i\omega_{\mu\nu}S^{\mu\nu}/2))$. When we make a coordinate transformation, we can think about it in two equivalent ways; either we rotate our coordinate system with Λ or we "rotate the physics" with Λ^{-1}. If we choose the latter we obtain (6.5.21). This is why x appears on the right hand side as well and not x', as it would if we saw this as a rotation of our coordinate system and the physics held still. Thus, x on the right hand side of (6.5.21) are the coordinates in the new system obtained from one of the two approaches. Let us take a closer look at the transformation $\Phi(x) \longrightarrow \Phi(\Lambda^{-1}x)$. Since this is a Lorentz transformation on the coordinates there should be an operator $L^{\mu\nu}$ satisfying the Lie algebra such that

$$\exp(-i\omega_{\mu\nu}L^{\mu\nu}/2)\Phi(x) = \Phi(\Lambda^{-1}x). \tag{6.5.22}$$

To find $L^{\mu\nu}$ we consider the infinitesimal transformation

$$\begin{aligned} &(1 - \frac{i}{2}\omega_{\mu\nu}L^{\mu\nu})\Phi(x) = \Phi((1-\omega)x) \Rightarrow \\ &L^{\mu\nu}\Phi(x) = \frac{-2i\big(\Phi(x) - \Phi\big((1-w)x\big)\big)}{\omega_{\mu\nu}}. \end{aligned} \tag{6.5.23}$$

The only place where L appears in (6.5.22) is in the summation with the antisymmetric $\omega_{\mu\nu}$. Thus, we only need the antisymmetric part of L

$$L^{[\mu\nu]} = \frac{1}{2}(L^{\mu\nu} - L^{\nu\mu}). \tag{6.5.24}$$

This antisymmetric part is commonly denoted $\mathcal{L}$. If we use (6.5.23) in (6.5.24) we obtain

$$\begin{aligned} \mathcal{L}^{\mu\nu}\Phi(x) &= \frac{1}{2}\left(\frac{-2i(\Phi(x) - \Phi((1-\omega)x))}{\omega_{\mu\nu}} + \frac{2i(\Phi(x) - \Phi((1-\omega)x)}{\omega_{\nu\mu}}\right) = \\ &= \frac{-i(\Phi(x) - \Phi((1-\omega)x))}{\omega_{\mu\nu}x^{\nu}}x^{\nu} + \frac{i(\Phi(x) - \Phi((1-\omega)x)}{\omega_{\nu\mu}x^{\mu}}x^{\mu} = \\ &= i(x^{\mu}\partial^{\nu} - x^{\nu}\partial^{\mu})\Phi(x), \end{aligned} \tag{6.5.25}$$

where $\partial^\mu = \frac{\partial}{\partial x^\mu}$. With this result, (6.5.21) can be expressed as

$$\Phi(x) \longrightarrow \exp(-i\omega_{\mu\nu}S^{\mu\nu}/2)\exp(-i\omega_{\mu\nu}\mathcal{L}^{\mu\nu}/2)\Phi_b(x). \tag{6.5.26}$$

One can show that S and $\mathcal{L}$ commute, using the fact that S is a finite constant matrix, which allows us to define[3]

$$\mathcal{J} \stackrel{def}{=} S + \mathcal{L} \tag{6.5.27}$$

such that

$$\Phi(x) \longrightarrow \exp(-i\omega_{\mu\nu}\mathcal{J}^{\mu\nu}/2)\Phi_b(x). \tag{6.5.28}$$

$\mathcal{J}$ is an infinite dimensional representation called the **field representation** of the Lorentz algebra. For the Poincaré algebra, we have to add the momentum generators $\hat{P}^\mu = i\partial^\mu$.

6.6 Classification of Elementary Particles

From the discussion in 6.2 it followed that if we are to consider quantum mechanical theories we can construct unitary representations acting on Hilbert spaces. Since $SO(1,3)$ is non-compact, which we will discover in section 6.7 as we ponder the topological structure of the Lorentz and Poincaré groups, any unitary representation has to be of infinite dimension but we will see how finite representations can arise and be used. The space of all possible states of a particular particle must be an irreducible representation since a reducible representation would imply states that can not be transformed into other states with any Lorentz transformation. So, we are searching after irreducible representations that correspond to one-particle states. A one-particle state does not necessarily mean an elementary particle but in the following pages we will consider one-particle states since the difference is irrelevant.

As we saw in section 6.3, $[\hat{P}^\mu, \hat{P}^\rho]$ commutes for every μ, ρ. This makes it possible to label our one-particle states with the 4-momentum p^μ as $|\psi_{p,\,\sigma}\rangle$, where p^μ is the eigenvalue to the component $\hat{P}^\mu$, i.e.

$$\hat{P}^\mu|\psi_{p,\sigma}\rangle = p^\mu|\psi_{p,\,\sigma}\rangle \tag{6.6.1}$$

where σ represents still unknown degrees of freedom/quantum numbers for the particle. If we make a transformation $\hat{U}(\Lambda,\mathbf{0})$ on $|\psi_{p,\,\sigma}\rangle$ we obtain a transformation of the 4-momentum according to

$$\hat{P}^\mu\hat{U}(\Lambda,\mathbf{0})|\psi_{p,\sigma}\rangle = \hat{U}(\Lambda,\mathbf{0})\left(\hat{U}(\Lambda,\mathbf{0})\right)^{-1}\hat{P}^\mu\hat{U}(\Lambda,\mathbf{0})|\psi_{p,\,\sigma}\rangle. \tag{6.6.2}$$

where we inserted a $\hat{U}\hat{U}^{-1} = \hat{\mathbb{1}}$. Next, we use the result of equation (6.3.18) and distinguish that

$$\begin{aligned}
\hat{U}(\Lambda,\mathbf{0})\hat{P}^\rho\left(\hat{U}(\Lambda,\mathbf{0})\right)^{-1} &= \Lambda_\mu{}^\rho\hat{P}^\mu \Leftrightarrow \\
\hat{P}^\rho = \Lambda_\mu{}^\rho\left(\hat{U}(\Lambda,\mathbf{0})\right)^{-1}\hat{P}^\mu\hat{U}(\Lambda,\mathbf{0}) &= (\Lambda^T)^\rho{}_\mu\left(\hat{U}(\Lambda,\mathbf{0})\right)^{-1}\hat{P}^\mu\hat{U}(\Lambda,\mathbf{0}) \Leftrightarrow \\
\left((\Lambda^T)^\mu{}_\rho\right)^{-1}\hat{P}^\rho &= \left(\hat{U}(\Lambda,\mathbf{0})\right)^{-1}\hat{P}^\mu\hat{U}(\Lambda,\mathbf{0}) \Leftrightarrow \\
\Lambda^\mu{}_\rho\hat{P}^\rho &= \left(\hat{U}(\Lambda,\mathbf{0})\right)^{-1}\hat{P}^\mu\hat{U}(\Lambda,\mathbf{0}).
\end{aligned} \tag{6.6.3}$$

[3]Not the same $\mathcal{J}$ as before.

If we substitute this into (6.6.2) we get

$$\hat{P}^\mu \hat{U}(\Lambda, \mathbf{0})|\psi_{p,\,\sigma}\rangle = \hat{U}(\Lambda, \mathbf{0})\Lambda^\mu{}_\rho \hat{P}^\rho |\psi_{p,\,\sigma}\rangle = \Lambda^\mu{}_\rho p^\rho \hat{U}(\Lambda, \mathbf{0})|\psi_{p,\sigma}\rangle. \tag{6.6.4}$$

This means that when we transform a state $|\psi_{p,\,\sigma}\rangle$ with momentum p^μ we obtain a new state $\hat{U}(\Lambda, \mathbf{0})|\psi_{p,\,\sigma}\rangle$ with momentum/eigenvalue $\Lambda^\mu{}_\rho p^\rho$ that is a new eigenstate to $\hat{P}^\mu$. Now, let us introduce a reference momentum k^μ. Since p^2 is an invariant under any Lorentz transformation we can associate a reference momentum k^μ to every value of p^2. All other momenta that square to p^2 can be obtained through a standard Lorentz transformation (a boost) L defined as

$$L^\mu{}_\nu(p)k^\nu \overset{def}{=} p^\mu. \tag{6.6.5}$$

Moreover, we define our still unknown label σ so that it does not change under $L(p)$, that is

$$N(p)\hat{U}(L(p))|\psi_{k,\,\sigma}\rangle = |\psi_{p,\,\sigma}\rangle, \tag{6.6.6}$$

where $N(p)$ is a normalization factor. Now it is time to shift our focus to the σ-label and figure out what it represents. The physical property connected to σ is defined in the reference frame, i.e. the one where the momentum is k, but for an arbitrary momenta the physical meaning of σ may not be obvious. If we act with an arbitrary Lorentz transformation $\hat{U}(\Lambda, \mathbf{0})$ on $|\psi_{p,\,\sigma}\rangle$ we get

$$\hat{U}(\Lambda)|\psi_{p,\,\sigma}\rangle = N(p)\hat{U}(\Lambda)\hat{U}(L(p))|\psi_{k,\,\sigma}\rangle = N(p)\underbrace{\hat{U}\left(L(\Lambda p)\right)}_{k\to\Lambda p}\underbrace{\hat{U}\left(L^{-1}(\Lambda P)\Lambda L(p)\right)}_{\text{leaves } k \text{ invariant}}|\psi_{k,\,\sigma}\rangle. \tag{6.6.7}$$

Observe that we simply write $\hat{U}(\Lambda, \mathbf{0}) = \hat{U}(\Lambda)$ for convenience. From the definition of L we see that $\hat{U}\left(L^{-1}(\Lambda p)\Lambda L(p)\right)$ acts on the momentum as $k \to p \to \Lambda p \to k$, i.e. k is invariant. The operators $L^{-1}(\Lambda p)\Lambda L(p)$ constitute a subgroup $G_L \subset SO(1,3)$ called **the little group** which is defined to leave k invariant. A more formal mathematical definition can be found in appendix A. If we denote the elements in G_L as W we have that

$$\hat{U}(W)|\psi_{k,\,\sigma}\rangle = \sum_{\sigma'} D_{\sigma\sigma'}|\psi_{k,\,\sigma'}\rangle. \tag{6.6.8}$$

Thus, states with fixed momentum k furnish a representation for the little group G_L. $D_{\sigma\sigma'}$ is the matrix representation for the little group. If we use (6.6.6) and (6.6.8) in (6.6.7) we obtain

$$\begin{aligned}\hat{U}(\Lambda)|\psi_{p,\,\sigma}\rangle &= N(p)\hat{U}(L(\Lambda p))\hat{U}\left(L^{-1}(\Lambda P)\Lambda L(p)\right)|\psi_{k,\,\sigma}\rangle = \\ &= N(p)\hat{U}(L(\Lambda p))\sum_{\sigma'} D_{\sigma\sigma'}|\psi_{k,\,\sigma'}\rangle = \frac{N(p)}{N(\Lambda p)}\sum_{\sigma'} D_{\sigma\sigma'}|\psi_{\Lambda p,\,\sigma'}\rangle.\end{aligned} \tag{6.6.9}$$

This tells us that if we can find representations for the little group G_L, i.e. the matrix D, we can use these as representations for the Lorentz group. Note that we only make the summation over σ' and not the momentum, i.e. the momentum Λp is invariant under the action of D. We may say that the boosted little group leaves the boosted momentum invariant. In fact,

if we consider another little group which leaves a momentum p invariant, then these little groups are equivalent. Before we can determine what σ stands for we have to determine G_L. There are two different values of p^2 of interest, $p^2 < 0$ and $p^2 = 0$, i.e. massive and massless particles,

(i) $p^2 < 0$: This represents the momentum for a massive particle in the rest frame of the particle. We choose the reference momentum to be $k = (mc, 0, 0, 0)$ which squares to $-m^2c^2$. If a massive particle is at rest, its momentum is invariant under spatial rotations. The little group is then $SO(3)$. However, to be able to deal with spinorial representations of the Lorentz group, we lift $SO(3)$ to its double cover $SU(2)$ (more on this in the next section).

(ii) $p^2 = 0$: A massless particle can not be brought to rest by any Poincaré transformation, but with a suitable choice of inertial frame we can take $k = (p, 0, 0, p)$ as a reference momentum which squares to 0. The little group for massless particles is $ISO(2)$, the group of rotations and translations in two dimensions.

So, for a fixed k, $|\psi_{k,\sigma}\rangle$ are states in a representation of spin j for $SO(3)$. This means that σ denotes the spin projection j_z in the particle's rest frame and we may write $|\psi_{k,\sigma}\rangle = |\psi_{k,j_z}\rangle$. For each j we have an irreducible representation of dimension $2j+1$ and the eigenstates under the operator $\hat{J}_z$ are

$$|\psi_{k,-j}\rangle,\ |\psi_{k,-j+1}\rangle,\ \ldots,\ |\psi_{k,j}\rangle. \tag{6.6.10}$$

$D_{\sigma\sigma'}$ is thus a $(2j+1)\times(2j+1)$ matrix. Since $[\hat{J}_z, \hat{P}_j] \neq 0$ we can not label our eigenstates with both the momentum and the $\hat{J}_z$ eigenvalue, but we can take $\hat{P}^2$ and $\hat{J}_z$ since these commute. So, we should think of the label k as the mass of the particle. Both the mass m and the spin quantum number j are invariants. Thus, a massive particle is completely specified with two labels, mass m and spin j, i.e. the irreducible representations for the Poincaré group are specified according to the mass m and spin j. These are in fact quadratic Casimir invariants, see 5.5, and hence Lorentz invariants. The corresponding Casimir operators to m and j are given by (6.6.11),

$$\hat{C}_1 = \hat{P}^\mu \hat{P}_\mu, \quad \hat{C}_2 = W^\mu W_\mu, \tag{6.6.11}$$

where W^μ, the components of $\boldsymbol{W}$, are given by

$$W^\mu = -\frac{1}{2}\epsilon^{\mu\nu\rho\sigma}\hat{J}_{\nu\rho}\hat{P}_\sigma. \tag{6.6.12}$$

where $\epsilon^{\mu\nu\rho\sigma}$ is the Levi-Civita tensor with four indices. $\boldsymbol{W}$ is called the **Pauli-Lubanski pseudo-vector**. The eigenvalue to $\hat{C}_1$ is m^2c^2 and the eigenvalue to $\hat{C}_2$ is $-m^2c^2j(j+1)$.

The problem of finding unitary irreducible representations for the non-compact Poincaré group has been reduced to a problem of finding unitary irreducible representations for the compact $SO(3)$-group or the compact $SU(2)$-group, respectively, which are finite and well known.

The massless particles momenta are invariant under spatial rotation in the xy-plane, i.e. $SO(2)$. The group $SO(2)$ has representations labeled with **helicity** $h = 0, \pm 1/2, \pm 1,....$ Helicity is defined as the projection of the spin onto the direction of the momentum

$$h \overset{def}{=} \boldsymbol{S} \cdot \hat{\boldsymbol{p}} \tag{6.6.13}$$

where $|\hat{\boldsymbol{p}}| = 1$. A massless particle with spin j have $2j + 1$ possible values of helicity and thus $2j + 1$ degrees of freedom. Since the photon's spin projection can be either $+1$ or -1 its helicity is $h = \pm 1$. This corresponds to the two polarization states, **right- or left circularly polarized**. The reference momenta k is not only invariant under rotations, but under a certain combination of rotations, boosts and translations as well. This is why $ISO(2)$ is the little group. An example of this is the following. Consider a boost along the x-axis,

$$k \to \Lambda^{(x)} k = (\gamma p, \gamma\beta p, 0, p) = k', \tag{6.6.14}$$

with $\beta = v/c$, followed with a rotation around the y-axis, see example (3.1.5),

$$k' \to R^{(y)} k' = (\gamma p, 0,\ 0,\ \gamma p) = k'' \tag{6.6.15}$$

where $\sin(\theta) = \beta$, $\cos(\theta) = 1/\gamma$, and finally followed by a boost in the $-z$-direction

$$k'' \to \Lambda^{(-z)} k'' = (p, 0,\ 0,\ p) = k \tag{6.6.16}$$

where $\gamma' = \frac{1}{\beta\gamma + \gamma}$.

A more elaborate discussion about $ISO(2)$ may be found in ([5], pp. 69-74). Another massless particle whose existence has yet to be confirmed is the *graviton* the supposed transmitter for the gravitational force with helicity $h = \pm 2$.
Returning to the j-quantum number, we see that we can classify states according to integer or half-integer values. In the next section we will discover that the composition rule for the unitary operators inducing Poincaré transformations can be written according to

$$\hat{U}(\Lambda,\, \boldsymbol{a})\hat{U}(\tilde{\Lambda},\, \tilde{\boldsymbol{a}}) = e^{2\pi i j}\hat{U}(\Lambda\tilde{\Lambda},\, \Lambda\tilde{\boldsymbol{a}} + \boldsymbol{a}) \tag{6.6.17}$$

where j integer gives a plus-sign and j half-integer a minus-sign. When j is a half integer we obtain *projective representations*.

6.7 Projective Representations

In section 6.2, we discerned that our operators $\hat{U}$ fell short of a phase to being a group representation. Instead the following operator relation appeared:

$$\hat{U}(T_1)\hat{U}(T_2) = e^{i\phi(T_1,\, T_2)}\hat{U}(T_1T_2). \tag{6.7.1}$$

The algebraic nature of this phase as well as its topological interpretation will be presented throughout the following pages. To begin with, in order to achieve associativity between operators according to

$$\hat{U}(T_3)\left(\hat{U}(T_1)\hat{U}(T_2)\right) = \left(\hat{U}(T_3)\hat{U}(T_1)\right)\hat{U}(T_2) \tag{6.7.2}$$

we deduce a condition (6.7.3) on ϕ by using (6.7.1) two times on each side of equation (6.7.2).

$$\phi(T_1,\, T_2) + \phi(T_3,\, T_1T_2) = \phi(T_3,\, T_1) + \phi(T_3T_1,\, T_2) \tag{6.7.3}$$

Here we encounter a tricky situation. On one hand, phases such as

$$\phi(T_1,\, T_2) = \varphi(T_1 T_2) - \varphi(T_1) - \varphi(T_2) \tag{6.7.4}$$

evidently satisfies (6.7.3). Yet they can easily be taken care of we redefine our operator $\hat{U}$ as $\hat{Y}$ such that

$$\hat{Y}(T) \stackrel{def}{=} \hat{U}(T) e^{i\varphi(T)} \tag{6.7.5}$$

which removes the phase from (6.7.1).

$$\hat{Y}(T_1)\hat{Y}(T_2) = \hat{Y}(T_1 T_2) \tag{6.7.6}$$

On the other hand, there might be representations where the phase cannot be eliminated in this manner, representations on the quantum mechanical Hilbert space which are intrinsically projective.[4] Which option is valid when it comes to describing the Poincaré and Lorentz group?

Initially we will examine how the phase appears when describing the representation of the Poincaré algebra.

6.7.1 Central Charges in the Lie Algebra

In order to determine the properties of the additional phase in the Lie algebra picture, we study the group near its identity. If either T_1 or T_2 in equation (6.7.1) is a unit transformation than the phase should obviously disappear. Hence we have

$$\phi(1,\, T_2) = \phi(T_1,\, 1) = 0. \tag{6.7.7}$$

When both T_1 and T_2 are close to the unit element, the phase generated by their difference will be small. Moreover, they can be parametrized by coordinates θ^a and $\tilde{\theta}^a$ respectively. Let $T(0) = 1$, $T_1 = T(\theta)$ and $T_2 = T(\tilde{\theta})$. According to equation (6.7.7), the expansion of $\phi(T(\theta), T(\tilde{\theta}))$ when ϕ is small can be written as

$$\phi(T(\theta), T(\tilde{\theta})) = f_{ab}\theta^a \tilde{\theta}^b + \dots \tag{6.7.8}$$

where f_{ab} are some real constants. Moreover we can perform an expansion of the operators $\hat{U}$ in equation (6.7.1) in a similar manner to what we did in the derivation of the Poincaré algebra in equation (6.3.4), such that

$$\hat{U}(T(\theta)) = \hat{1} + i\theta^a \hat{t}_a + \frac{i}{2}\theta^b \theta^c \hat{s}_{bc} + \dots \tag{6.7.9}$$

with $\hat{t}$ and $\hat{s}$ are Hermitian operators and $\hat{s}_{bc}$ is symmetric in indices (since switching b for c will not alter the θ:s). The composed $\hat{U}(T(\theta)T(\tilde{\theta}))$ can be restated using a function $\Theta(\theta,\, \tilde{\theta})$ which adheres to the group multiplication between two symmetry operators

[4] A set of functions $\phi(T_1,\, T_2)$ which satisfies equation (6.7.3) and where the difference between the different functions can be expressed as functions of the type in equation (6.7.4), is called a **2-cocycle**. These are objects of study in the mathematical field known as *cohomology*, see for instance chapter 6 in [7] and [21]. If the set includes any function $\phi = 0$, the 2-cocycle is said to be *trivial* and we can redefine our operators and remove the phase. Yet, some symmetry groups allow *non-trivial* 2-cocycles and this indicates intrinsically projective representations ([5], p. 82).

$$\hat{U}(T(\theta)T(\tilde{\theta})) = \hat{U}\left(T\left(\Theta(\theta,\,\tilde{\theta})\right)\right) \tag{6.7.10}$$

where the function $\Theta(\theta, \tilde{\theta})$ will obey

$$\Theta^a(\theta,\,0) = \Theta^a(0,\,\tilde{\theta}) = \theta^a \tag{6.7.11}$$

with $\theta^a = 0$ for the identity transformation. Expanding Θ^a in terms of θ^a and $\tilde{\theta}^a$ to the second order will yield

$$\Theta^a(\theta,\,\tilde{\theta}) = \theta^a + \tilde{\theta}^a + f^a{}_{bc}\theta^b.\tilde{\theta}^c \tag{6.7.12}$$

Inserting the result of equation (6.7.12) into the operator $\hat{U}$ in (6.7.10) we have now found all terms of the expression in (6.7.1), now modified according to our parametrization of the symmetry operators

$$\hat{U}(T(\theta))\hat{U}(T(\tilde{\theta})) = \exp\left(i\phi(T(\theta), T(\tilde{\theta}))\right)\hat{U}\left(T\left(\Theta(\theta,\,\tilde{\theta})\right)\right). \tag{6.7.13}$$

By omitting terms of higher order than two, it is possible to place the expanded expression for $\hat{U}$ and ϕ into equation (6.7.13). The left-hand side will then be

$$\begin{aligned}&\hat{U}(T(\theta))\hat{U}(T(\tilde{\theta})) \approx \left(\hat{1} + i\theta^a\hat{t}_a + \frac{i}{2}\theta^b\theta^c\hat{s}_{bc}\right)\left(\hat{1} + i\tilde{\theta}^d\hat{t}_d + \frac{i}{2}\tilde{\theta}^e\tilde{\theta}^f\hat{s}_{ef}\right) \approx \\ &\hat{1} + i\left(\theta^a + \tilde{\theta}^a\right)\hat{t}_a + \frac{i}{2}\left(\theta^b\theta^c + \tilde{\theta}^b\tilde{\theta}^c\right)\hat{s}_{bc} - \theta^a\tilde{\theta}^d\hat{t}_a\hat{t}_d\end{aligned} \tag{6.7.14}$$

where the recurring indices are renamed so that coefficients can be matched. Moving on to the right-hand side, it can be expanded as

$$\begin{aligned}&\exp\left(i\phi(T(\theta), T(\tilde{\theta}))\right)\hat{U}\left(T\left(\Theta(\theta,\,\tilde{\theta})\right)\right) \approx \left(\hat{1} + if_{ab}\theta^a\tilde{\theta}^b\right)\cdot \\ &\cdot\left(\hat{1} + i\left(\theta^a + \tilde{\theta}^a + f^a{}_{bc}\theta^b\tilde{\theta}^c\right)\hat{t}_a + \frac{i}{2}\left(\theta^b + \tilde{\theta}^b\right)\left(\theta^c + \tilde{\theta}^c\right)\hat{s}_{bc}\right) \approx \\ &\approx \hat{1} + i\left(\theta^a + \tilde{\theta}^a + f^a{}_{bc}\theta^b\tilde{\theta}^c\right)\hat{t}_a + \frac{i}{2}\left(\theta^b + \tilde{\theta}^b\right)\left(\theta^c + \tilde{\theta}^c\right)\hat{s}_{bc} + \\ &+ if_{ab}\theta^a\tilde{\theta}^b\hat{1}.\end{aligned} \tag{6.7.15}$$

Setting both sides equal to one another, omitting terms of higher order and those which appear on both sides such as the unit generator we arrive at

$$-\theta^a\tilde{\theta}^d\hat{t}_a\hat{t}_d = if^a{}_{bc}\theta^b\tilde{\theta}^c\hat{t}_a + \frac{i}{2}\left(\theta^b\tilde{\theta}^c + \tilde{\theta}^b\theta^c\right)\hat{s}_{bc} + if_{ab}\theta^a\tilde{\theta}^b\hat{1}. \tag{6.7.16}$$

Next we use that the order of the multiplied parametrizations is irrelevant, $\theta^b\tilde{\theta}^c = \tilde{\theta}^b\theta^c$, and switch indices a and d on the left-hand side to b and c. Hence, the expression changes to

$$-\theta^b\tilde{\theta}^c\hat{t}_b\hat{t}_c = if^a{}_{bc}\theta^b\tilde{\theta}^c\hat{t}_a + i\theta^b\tilde{\theta}^c\hat{s}_{bc} + if_{ab}\theta^a\tilde{\theta}^b\hat{1}. \tag{6.7.17}$$

By removing all terms associated with the parametrization and rearranging, we find that the generator $\hat{s}$ must follow the condition on the first line of (6.7.18) and presents to us the commutation relations between different components of $\hat{t}$.

$$
\begin{aligned}
&i\hat{s}_{bc} = -\hat{t}_b\hat{t}_c - if^a{}_{bc}\hat{t}_a - if_{bc}\hat{1} = i\hat{s}_{cb} = -\hat{t}_c\hat{t}_b - if^a{}_{cb}\hat{t}_a - if_{cb}\hat{1} \\
&\Rightarrow \hat{t}_b\hat{t}_c - \hat{t}_c\hat{t}_b = -if^a{}_{bc}\hat{t}_a + if^a{}_{cb}\hat{t}_a - f_{bc}\hat{1} + if_{cb}\hat{1} \\
&\Leftrightarrow \left[\hat{t}_b,\, \hat{t}_c\right] = -i\left(f^a{}_{bc} - f^a{}_{cb}\right)\hat{t}_a - i\left(f_{bc} - f_{cb}\right)\hat{1}
\end{aligned}
\tag{6.7.18}
$$

Renaming the structure constants as of

$$
C_{bc} \overset{def}{=} f_{bc} - f_{cb} \text{ and } C^a{}_{bc} \overset{def}{=} f^a{}_{bc} - f^a{}_{cb} \tag{6.7.19}
$$

the commutation relations become

$$
\left[\hat{t}_b,\, \hat{t}_c\right] = iC^a{}_{bc}\hat{t}_a + iC_{bc}\hat{1}. \tag{6.7.20}
$$

The coefficients in front of the generator of the unit element are the **central charges** of the Lie algebra. These charges correspond to the phases we encounter when dealing with projective representations of Lie groups. In a moment we shall discover a way to eliminate these charges and also briefly discuss when they constitute an important ingredient for new theories. When the central charges remain in the algebra, they alter the representation theory.

Our C coefficients must adhere to the Jacobi identity in order for us to obtain a Lie algebra according to the definition 2.3.3 in section 2.3. This means that

$$
\begin{aligned}
&\left[[\hat{t}_b,\, \hat{t}_c], \hat{t}_d\right] + \left[[\hat{t}_c,\, \hat{t}_d], \hat{t}_b\right] + \left[[\hat{t}_d,\, \hat{t}_b], \hat{t}_c\right] = 0 \\
&\Rightarrow C^a{}_{bc}C^e{}_{ad} + C^a{}_{cd}C^e{}_{ab} + C^a{}_{db}C^e{}_{ac} = 0 \\
&\text{and } C^a{}_{bc}C_{ad} + C^a{}_{cd}C_{ab} + C^a{}_{db}C_{ac} = 0.
\end{aligned}
\tag{6.7.21}
$$

The latter constraint for the central charges can be followed if

$$
C_{ab} = C^e{}_{ab}\vartheta_e \tag{6.7.22}
$$

where ϑ is an arbitrary set of real constants. With this elaboration, we are able to redefine our generator so that the central charges can be eliminated from the algebra according to equation (6.7.23)

$$
\hat{t}_a \to \hat{t}'_a \overset{def}{=} \hat{t}_a + \vartheta_a \tag{6.7.23}
$$

and the new generators will obey commutation relations without central charges

$$
\left[\hat{t}'_b,\, \hat{t}'_c\right] = iC^a{}_{bc}\hat{t}'_a. \tag{6.7.24}
$$

However, several pertinent issues remain, such as whether this procedure always is a valid option. There is a theorem ([22], pp. 44-45) which states that

Theorem 6.7.1. The generators of semisimple Lie algebras can always be redefined so that all central charges are removed.

As stated in chapter 5, the Lorentz group has a semisimple Lie algebra, which means that its generators can be modified to eliminate central charges. Yet, for the Poincaré group the generators $\hat{P}^\mu$ associated with translations form an invariant Abelian subalgebra and $\mathfrak{iso}(1,\, 3)$ is not semisimple. Still, it is possible to perform similar calculations like those above and shift

the generators $\hat{J}$ and $\hat{\boldsymbol{P}}$ by constant terms in order to obtain the commutation relations of expression (6.3.33) for the defining representation of the Poincaré group. Please refer to appendix F for guidance.

An important case where the central charges cannot easily be disposed of is when dealing with the so-called *super-Poincaré-algebra*, which is an extended form of the Poincaré algebra and is used in *supersymmetry* - a theory which attempts to describe physics beyond the Standard Model. [23] We will not develop its origin and implications here, but the interested reader should be aware of this connection.

Having explored how phases appear in the Lie algebra of a group, it is time to establish some rules regarding when they can be omitted. Fortunately there is a theorem ([5], pp. 83-84) which gives that

Theorem 6.7.2. The phase $e^{i\phi}$ of any representation $\hat{U}(T)$ of a given group can be chosen so that $\phi = 0$ if

(i) The generators of the group can be redefined according to equation (6.7.23) in order to eliminate all central charges from the Lie algebra.

(ii) The group is simply connected, i.e. any loop that starts and ends at the same group element on the group manifold may be shrunken continuously to a point.

Since we have investigated how generators can be redefined properly, it is time to move on to topological arguments.

6.7.2 Phases on the Group Manifold

In order to examine the properties of phases in the entire group picture instead of the Lie algebra, we backtrack a few chapters to section 2.3 and start with the relations between $SU(2)$ and $SO(3)$. Their Lie algebra is isomorphic, $\mathfrak{su}(2) \cong \mathfrak{so}(3)$, yet they yield different groups upon exponentiation. When a Lie algebra gives rise to groups with different manifolds with different topologies, the group with a simply connected manifold is called the *universal covering group*. This definition can be restated with slightly more mathematical rigour as:

Definition 6.7.1. *Let $\mathcal{G}$ and $\tilde{\mathcal{G}}$ be connected matrix Lie groups and thus also topological spaces. Moreover, let $\tilde{\mathcal{G}}$ also be simply connected. $\tilde{\mathcal{G}}$ is together with a Lie group homomorphism $p : \tilde{\mathcal{G}} \mapsto \mathcal{G}$, called the* **projection map***, the* **universal covering group** *of $\mathcal{G}$ with the following properties*

(i) p is surjective, i.e. it maps $\tilde{\mathcal{G}}$ onto $\mathcal{G}$.

(ii) There is neighborhood $U \subset \tilde{\mathcal{G}}$ to the identity element $e \in \tilde{\mathcal{G}}$ which p maps onto a neighborhood $V \subset \mathcal{G}$ of the identity element $e \in \mathcal{G}$.

Definition 6.7.2. *A* **double cover** *is a universal covering group where the projective map p is two-to-one.*

The isomorphism of different Lie algebras is naturally occurring together with a group and its universal covering group, which we have already come across when we first encountered Lie algebras. Their relation is stated below, for the main ideas of the proof please see ([11], p. 99).

PROPOSITION 6.7.1: Suppose $\mathcal{G}$ is a connected matrix Lie group with Lie algebra $\mathfrak{g}$. Let $(\tilde{\mathcal{G}}, p)$ be its universal covering group where $\tilde{\mathcal{G}}$ has a Lie algebra $\tilde{\mathfrak{g}}$. Then the corresponding projective map ϕ between the Lie algebras $\tilde{\mathfrak{g}}$ and $\mathfrak{g}$ is given by

$$\phi : \tilde{\mathfrak{g}} \mapsto \mathfrak{g} \tag{6.7.25}$$

where ϕ is an isomorphism.

We have already established that $SU(2)$ is isomorphic to the embedding of the 3-sphere in $\mathbb{R}^4$ which is the same as the unit ball in $\mathbb{R}^3$. Now, looking at $SO(3)$ we can take all possible rotations angles and map them upon this ball with radius π. The center corresponds to the unit transformation and the Cartesian axes correspond to each spatial rotation axis. Points lying on the surface of the ball correspond to rotations of π- or $-\pi$-radians in some direction. For $SO(3)$, taking a rotational matrix and rotating a vector π radians will be equivalent to rotating it $-\pi$ radians. If we consider $SO(3)$ as a manifold, this means that antipodal points on the 3-sphere S^3 are identified with one another. An antipodal point is on the other side of a sphere/circle for a given point, the diameter of the sphere/circle is the distance between them. This means that a path through the center of the ball connecting two antipodal points is a closed loop. However, this loop cannot be shrunk to a point since the points on the surface must remain antipodal during any deformation. Otherwise the loop would cease to be. In matrix terms the loop in question represents a continuous sequence of rotations about the axis corresponding to the straight line through the center of the sphere starting and ending at the identity matrix, producing a 0- or 2π-rotation.

Still, going through the loop twice will yield a continuously deformable loop. This is a direct consequence of the fact that if we write out the loop between the antipodal points twice, it can be deformed and placed at the surface of the ball. At that position we mirror the second half and place it on the other side of the ball. Then, we have effectively created a circle out of the loop and it can be continuously deformed to a point. This procedure is illustrated in figure 6.1.

Hence we have the following relation between the $SO(3)$ and $SU(2)$-manifolds:

$$SO(3) \cong SU(2)/\mathbb{Z}_2 \tag{6.7.26}$$

where $\mathbb{Z}_2$ is the finite cyclic group of order two which we encountered in chapter 2. It has two elements $\{1, -1\}$ corresponding to the fact that any loop between two antipodal points has to be gone through twice in order for us to obtain the identity element. Thus for every element in $SO(3)$ there exists two corresponding elements in $SU(2)$. One can think of $SO(3)$ as half of $SU(2)$ and topologically this is equivalent to $SO(3)$ being the northern or southern hemisphere of S^3 which is the same topological space as the so-called $\mathbb{R}P^3$, the real projective space of all lines passing through 0 in $\mathbb{R}^4$. With even more advanced mathematics one can state that $\mathbb{Z}_2$ is the first homotopy group, the fundamental group, of $S^3/\mathbb{Z}_2$. However, we will not develop that subject any further, see [7] for details. Basically, homotopy groups are a construct to enable classification of and comparison between different continuous deformations of maps onto one another. To conclude, $SU(2)$ is the double cover of $SO(3)$.

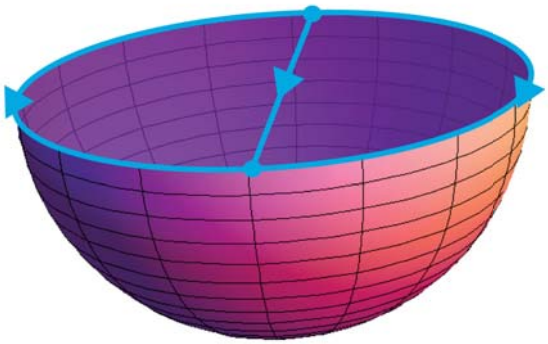


Figure 6.1: The turquoise dots indicate two antipodal points on S^3 and the line between them the loop gone through once. If the loop is gone through twice, we can create a circle on the surface of the ball, the outer loop, which can be continuously deformed to a point.

Moving on to the Lorentz group and the Poincaré group we will find upon scrutiny that they are also doubly connected. This can be achieved if we assign 4-vectors to complex 2×2 Hermitian matrices and note certain important implications.

An arbitrary 4-vector $\boldsymbol{A}$ can be used to construct a 2×2 Hermitian matrix M according to

$$M \overset{def}{=} A^\mu \sigma_\mu = \begin{pmatrix} A^0 + A^3 & A^1 - iA^2 \\ A^1 + iA^2 & A^0 - A^3 \end{pmatrix} \tag{6.7.27}$$

where σ_i are the Pauli matrices and

$$\sigma_\mu = (\sigma_0, \sigma_i) \text{ with } \sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.7.28}$$

The matrices σ_μ form a complete set since any 2×2 matrix can be written as a linear combination of them. Hence we obtain a map from the Minkowski space $\mathbb{R}^{1,3}$ to the Hilbert space $\mathcal{H}$ of all complex 2×2 matrices. Upon investigation of (6.7.27), we discern that the acquired matrix has a determinant

$$\det M = \left(A^0\right)^2 - \left(A^1\right)^2 - \left(A^2\right)^2 - \left(A^3\right)^2 = -A^\mu A_\mu \tag{6.7.29}$$

which is Lorentz-invariant. Moreover, it is possible to preserve the Hermiticity of M through the following transformation

$$M \to CMC^\dagger \tag{6.7.30}$$

where C is an arbitrary complex 2×2 matrix. If C:s determinant $|\det C| = 1$, then the determinant of M is preserved during the transformation (6.7.30) in view of

$$\det(CMC^\dagger) = (\det C)\,(\det M)\left(\det C^\dagger\right) = (\det C)^2\,(\det M)\,. \tag{6.7.31}$$

In this manner, we reach the conclusion that each complex 2×2 matrix C with $|\det C| = 1$ defines a real linear transformation of any 4-vector $\boldsymbol{A}$ that leaves (6.7.29) invariant - a homogeneous Lorentz transformation, $\Lambda(C)$ such that

$$CA^{\mu}\sigma_{\mu}C^{\dagger} = (\Lambda^{\mu}{}_{\nu}(C)A^{\nu})\,\sigma_{\mu}. \tag{6.7.32}$$

However, there is a slight complication due to the fact that it is possible to take two different C-matrices which differ only by a phase and obtain the same result of the transformation in (6.7.30) (since † alter a phase $e^{i\phi}$ to $e^{-i\phi}$). This can be remedied if we restrict the condition on the determinant with

$$\det C = 1. \tag{6.7.33}$$

This implies that $C \in SL(2,\,\mathbb{C})$. With this restriction on the determinant, the group depends on 4-1=3 complex parameters and thus 6 real ones if we decompose each complex number into its real and imaginary parts. Hence, we arrive at the same number of parameters as the Lorentz group has. Moreover if we take two matrices C and $\tilde{C}$, both in $SL(2,\,\mathbb{C})$ and use (6.7.32), then the subsequent relation appears

$$\begin{aligned} &\left(C\tilde{C}\right)A^{\mu}\sigma_{\mu}\left(C\tilde{C}\right)^{\dagger} = C\left(\tilde{C}A^{\mu}\sigma_{\mu}\tilde{C}^{\dagger}\right)C^{\dagger} = \\ &= C\Lambda^{\mu}{}_{\nu}(\tilde{C})A^{\nu}\sigma_{\mu}C^{\dagger} = \Lambda^{\mu}{}_{\kappa}(C)\Lambda^{\kappa}{}_{\nu}(\tilde{C})A^{\nu}\sigma_{\mu} \\ &\Leftrightarrow \Lambda(C\tilde{C}) = \Lambda(C)\Lambda(\tilde{C}) \end{aligned} \tag{6.7.34}$$

and we distinguish a Lie group homomorphism between $SL(2,\,\mathbb{C})$ and $SO(1,\,3)$. To sum up we have that for any matrix $C \in SL(2,\,\mathbb{C})$ there exists an associated Lorentz matrix Λ and there is a composition rule according to the last line of (6.7.34), which is reminiscent of definition 2.1.9 in chapter 2. In principle, it should be possible to go the other way and find a matrix $C \in SL(2,\,\mathbb{C})$ for every $\Lambda \in SO(1,\,3)$ but we do not have an isomorphism. Instead, there is a two-to-one correspondence between elements in $SL(2,\,\mathbb{C})$ and $SO(1,\,3)$. This can easily be distinguished if we consider the matrix $\zeta \in SL(2,\,\mathbb{C})$ defined as

$$\zeta(\theta) = \begin{pmatrix} e^{\frac{i\theta}{2}} & 0 \\ 0 & e^{-\frac{i\theta}{2}} \end{pmatrix} \tag{6.7.35}$$

which yields a Lorentz transformation producing a rotation around the z-axis with an angle θ. With $\theta = 2\pi$ this gives the identity in the Lorentz group but we obtain the element $\zeta(2\pi) = -\mathbb{1} \in SL(2,\,\mathbb{C})$, which is a non-trivial element. Still, taking $\theta = 4\pi$ will result in the identity of both the Lorentz group and $SL(2,\,\mathbb{C})$. We arrive at the conclusion that we have once more discovered a double cover, i.e $SL(2,\,\mathbb{C})$ is the double cover of the Lorentz group $SO(1,\,3)$ with $SL(2,\,\mathbb{C})/\mathbb{Z}_2$ the group of 2×2 complex matrices with determinant one and antipodal points identified with one another

$$SO(1,\,3) \cong SL(2,\,\mathbb{C})/\mathbb{Z}_2. \tag{6.7.36}$$

What is interesting is that, generally, ([7], pp. 117-118), there is a universal covering group $Spin(n)$ for every $SO(n)$, where we for $n = 3$ and $n = 4$ have the relation

$$Spin(3) = SU(2) \text{ and } Spin(4) = SU(2) \times SU(2) \tag{6.7.37}$$

On the Lie algebra level, the group homomorphism between $SO(4)$ and its double cover $Spin(4)$ changes to an isomorphism and we have

$$\mathfrak{so}(4) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2). \tag{6.7.38}$$

Yet, we must take a measure of caution when encountering $Spin(4)$ since we are not in the Euclidean space $\mathbb{R}^4$ but in Minkowski space $\mathbb{R}^{1,3}$ when dealing with the Lorentz and Poincaré groups. This difference was already subtly illuminated in the commutation relations between the generators of angular momentum and the boosts, summarized in (6.4.7). Had we been in $\mathbb{R}^4$ the commutation relations between the boosts should have featured a plus sign instead of minus and been written like

$$\left[\hat{K}_i,\, \hat{K}_j\right] = i\epsilon_{ijk}\hat{J}_k. \tag{6.7.39}$$

This has some consequences which we come across while studying representations of the Lorentz and the Poincaré groups. As we discovered when we dealt with important concepts associated with different representations of the Lorentz and Poincaré groups in the previous section, by combining and complexifying the generators of the boosts and angular momentum one can ascertain that the Lorentz algebra is equivalent to two $\mathfrak{su}(2)$. Since we are in Minkowski space, complex conjugation interchanges the two $\mathfrak{su}(2)$:s. The spin group in Minkowski space is $Spin(1,\, 3)$ and it is the double cover of $SO(1,\, 3)$. Its Lie algebra $\mathfrak{spin}(1,\, 3)$ is isomorphic to the Lorentz group's $\mathfrak{so}(1,\, 3)$ according to:

$$\mathfrak{spin}(1,\, 3) \cong \mathfrak{so}(1,\, 3) \cong \mathfrak{su}(2) \oplus \mathfrak{su}(2)^* \tag{6.7.40}$$

where complex conjugation has interchanged the $\mathfrak{su}(2)$:s. Also, since we have already established that $SL(2,\, \mathbb{C})$ is the double cover of $SO(1,\, 3)$, we deduce the isomorphism

$$Spin(1,\, 3) \cong SL(2,\, \mathbb{C}). \tag{6.7.41}$$

Now, we have come a long way in studying different phases in a topological picture. A few tools are still required to describe the Lorentz and Poincaré groups topologically but then we will finally establish the nature of our phase $e^{i\phi(T_1,\, T_2)}$ from the beginning of this chapter. Foremost there is an important theorem in linear algebra and complex analysis called the polar decomposition theorem ([24], p. 1126). It states that any non-singular matrix M may be written as

$$M = Ue^H \tag{6.7.42}$$

with U is a unitary matrix and H an Hermitian matrix. For $\det U$ we discerned in section 2.2 that its determinant is nothing but a phase factor since $|\det U| = 1$. Regarding the determinant of the exponential, it must be that $\det(\exp H) = \exp(\mathrm{Tr}\ H)$ with some elaborations in elementary linear algebra and $\exp(\mathrm{Tr}\ H)$ is real and positive due to the Hermiticity of H. Our condition for the determinant for a matrix $C \in SL(2,\, \mathbb{C})$ demands that

$$\det U = 1 \text{ and } \mathrm{Tr}\ H = 0 \tag{6.7.43}$$

where U encodes the rotations and H the translations. Moreover, this decomposition is unique which implies that $SL(2,\, \mathbb{C})$ topologically is the direct product of the space of all matrices H and all U. For the translational matrices H we have that any Hermitian traceless 2×2 matrix can be written as

$$\begin{pmatrix} x_3 & x_1 - ix_2 \\ x_1 + ix_2 & -x_3 \end{pmatrix}. \tag{6.7.44}$$

The parameters x_i are real but are not constrained with any additional condition. Thus they will result in a non-compact manifold with three variables, i.e. the ordinary three-dimensional space $\mathbb{R}^3$. Provided with clues such as unitary, 2×2 and a determinant which is equal to one, we easily conclude that the U:s are a representation of our familiar friend $SU(2)$. Its corresponding manifold is of course the surface of the unit ball in $\mathbb{R}^4$. Still, we had that the elements $C \in SL(2, \mathbb{C})$ were subjected to two-to-one homomorphism when mapped onto the Lorentz group. Let the elements C and $-C$ be elements in $SL(2, \mathbb{C})$ mapped onto the same element in the Lorentz group. If they are decomposed according to (6.7.42), we see that since e^H is always positive U and $-U$ are mapped onto the same element. In this manner we once more arrive at the topological structure $S^3/\mathbb{Z}$.

The topological structure of the Lorentz group is thus given by

$$\mathbb{R}^3 \times S^3/\mathbb{Z}_2 \tag{6.7.45}$$

where $\mathbb{R}^3$ which is non-compact represents the boosts and our familiar double-connected structure $S^3/\mathbb{Z}_2$ the rotations. For the Poincaré group an additional non-compact factor $\mathbb{R}^4$ is introduced in order to account for the translations in $\mathbb{R}^4$. Hence, its topological structure is given by

$$\mathbb{R}^4 \times \mathbb{R}^3 \times S^3/\mathbb{Z}_2 \tag{6.7.46}$$

since its decomposition is unique. Nevertheless, on the group level we obtain a semi-direct product of the Lorentz group and $\mathbb{R}^4$ for the Poincaré group due to the reasons described at the end of the vast calculation of section 6.3. In addition, it is easy to deduce their dimension by examining the topological structure. For the Lorentz group $S^3/\mathbb{Z}_2$ is 3-dimensional and with $\mathbb{R}^3$ another three dimensions are added, which gives totally six dimensions. For the Poincaré group the four dimensions of $\mathbb{R}^4$ are included and this yields ten dimensions.

We have found that the Lorentz and Poincaré groups are not simply connected and then by theorem 6.7.2 they have associated intrinsic projective representations. Yet, since they have a compact, double-connected part, any loop can be contracted to a point if it is circled twice. This fact bestow upon us the following condition for the phase seen in beginning of this curious section.

$$\left(\hat{U}(T_1)\hat{U}(T_2)\right)^2 = \left(e^{i\phi(T_1, T_2)}\hat{U}(T_1T_2)\right)^2 \Leftrightarrow \left(\hat{U}(T_1)\hat{U}(T_2)\hat{U}(T_1T_2)^{-1}\right)^2 = 1 \tag{6.7.47}$$

and the phase $e^{i\phi(T_1, T_2)}$ is simply a sign $\pm$. Hence the operator composition relation reduces to

$$\hat{U}(\Lambda, \boldsymbol{a})\hat{U}(\tilde{\Lambda}, \tilde{\boldsymbol{a}}) = \pm\hat{U}(\Lambda\tilde{\Lambda}, \Lambda\tilde{\boldsymbol{a}} + \boldsymbol{a}). \tag{6.7.48}$$

This sign $\pm$ corresponds to the different states we can have as irreducible representations of the Lorentz and Poincaré groups, those with integer spin, the **bosons**, and those with half-integer spin, the **fermions**. We use spinorial representations when dealing with fermions

and if we act on a state with spin 1/2 an angle of 4π we will return to the identity. We will see in the derivation of the little group that the constraint of finite mass provides only these two possible options for massive particles. However, topologically we can also impose a constraint on the helicity of the photon. If we rotate a photon's momentum with an angle 4π the rotation can be continuously deformed and shrunken to a point, i.e. equivalent to no rotation at all. Moreover, the phase $\exp(i4\pi j)$ must be unity and then the only valid choices for j is integer or half-integer.

In order to eliminate topologically intrinsic projective representations we usually lift all calculations to the universal covering group of the group we would like to study. The universal covering group has no projective representations, simply ordinary ones. This is what we do when we favor $SU(2)$ over $SO(3)$ when describing intrinsic angular momentum as well as $SL(2, \mathbb{C})$ for the Lorentz group. [21]

In the next two chapters we will go beyond the irreducible representations of the Poincaré group, characterized by mass or the lack of it and spin, outlining the external or spacetime-dependent symmetries and instead look at internal symmetries such as color and flavor. To be able to describe these quantities we require a new set of rules, the remarkable formalism known as *gauge theory*.

Chapter 7

Internal Symmetries

So far we have mainly been concerned about symmetries under different coordinate transformations of spacetime. As we saw in chapter 6 such symmetries are often called *external*. In this chapter and for the rest of the survey we will instead be looking at symmetries under transformations which do not act directly on spacetime itself but on the fields defined on spacetime. Hence they are called *internal* symmetries. These are often studied through theories of gauge which will be the approach of this chapter.

As physicists during the 20th century became increasingly aware of the importance of symmetries the studies were extended to include symmetries in the Hilbert spaces and thus to internal properties of the theories. The first great success was the reformulation of Maxwell's electrodynamics as a theory emerging from a $U(1)$ symmetry of spinor fields. It was done by Weyl, Fock and London putting the long-known degree of freedom in the potentials in a fundamental context and thus firmly established the theory of gauge. In 1954 Chan Ning Yang and Robert Mills widened the study of internal symmetries to non-Abelian groups through the extension to $SU(2)$ as an attempt to explain the presumed symmetry of protons and neutrons. Although not very successful in the beginning the studies of $SU(n)$ as symmetry groups, known as **Yang-Mills theories**, later have resulted in many major breakthroughs in theoretical physics and are the true cornerstones of the Standard Model. They are responsible for the unification of three of the four fundamental forces and are the foundation of the classification of elementary particles through the symmetry group of the Standard Model, $U(1) \times SU(2) \times SU(3)$.

In our exposure we will start with the historically first encounter with these symmetries in electromagnetism before going on to the more general studies. These in turn will be generalized to an insight of the gauge theories behind the Standard Model and the last chapter about particle physics. The chapter ends with an additional section about fiber bundles to give the keen reader a more geometric picture of the mathematics behind the theories. It is however complementary and is in no way necessary in order to read any other part of this survey.

7.1 Gauge Theory

A **gauge theory** is a theory which is invariant under a local transformation acting on the fields of the theory but not on the coordinates of spacetime. Thus the theory is said to have an internal symmetry referring to that property. The word *local* is essential in these theories

and amounts to the fact that the transformations depend on the spacetime coordinates so that they may be different in each point. The meaning of this will be more obvious when formulated mathematically in succeeding examples.

The first and simplest example of gauge invariance discovered in physics was in electrodynamics. Actually, the adding of a constant to the electric potential V is in fact a gauge transformation but the standard example is what follows.

Example 7.1.1. The Maxwell equations for the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$ are

$$\nabla \cdot \boldsymbol{E} = \frac{\rho}{\varepsilon_0} \tag{7.1.1a}$$

$$\nabla \times \boldsymbol{E} + \frac{\partial \boldsymbol{B}}{\partial t} = 0 \tag{7.1.1b}$$

$$\nabla \cdot \boldsymbol{B} = 0 \tag{7.1.1c}$$

$$\nabla \times \boldsymbol{B} - \frac{1}{c^2}\frac{\partial E}{\partial t} = \mu_0 \boldsymbol{j}. \tag{7.1.1d}$$

The homogeneous equations (7.1.1b) and (7.1.1c) allow us to express $\boldsymbol{E}$ and $\boldsymbol{B}$ in terms of potentials

$$\boldsymbol{E} = -\nabla V - \frac{\partial \boldsymbol{A}}{\partial t}, \qquad \boldsymbol{B} = \nabla \times \boldsymbol{A}. \tag{7.1.2}$$

In the study of these equations the two physical fields $\boldsymbol{E}$ and $\boldsymbol{B}$ were found to be invariant under the simultaneous transformations

$$V \to V - \frac{\partial f}{\partial t} \tag{7.1.3a}$$

$$\boldsymbol{A} \to \boldsymbol{A} + \nabla f \tag{7.1.3b}$$

since all the added terms cancel in the Maxwell equations, f being any differentiable function. These transformations are local since the arbitrary function f may depend on the coordinates. So the physics is invariant under a local transformation acting on the fields and hence electromagnetism is found to be a gauge theory.

This internal freedom of the theory can be utilized and to specify a particular function f is called to choose or to *fix a gauge*. This choice may have great importance for the simplicity in calculations. As a specific example of such a gauge we may take the Lorenz[1] gauge. It is Lorentz invariant and simplifies the wave equations for the potentials. To write these down we use equations (7.1.2), (7.1.1a) and (7.1.1d). By direct substitution we find

$$\nabla \cdot \left(-\nabla V - \frac{\partial \boldsymbol{A}}{\partial t}\right) = \frac{\rho}{\varepsilon_0} \tag{7.1.4}$$

$$\nabla \times (\nabla \times \boldsymbol{A}) - \frac{1}{c^2}\frac{\partial}{\partial t}\left(-\nabla V - \frac{\partial \boldsymbol{A}}{\partial t}\right) = \mu_0 \boldsymbol{j} \tag{7.1.5}$$

and by using the relation $\nabla \times (\nabla \times \boldsymbol{X}) = -\nabla(\nabla \cdot \boldsymbol{X}) - \Delta \boldsymbol{X}$ in equation (7.1.5) we get

$$\begin{aligned} -\Delta \boldsymbol{A} + \nabla(\nabla \cdot \boldsymbol{A}) + \frac{1}{c^2}\frac{\partial \nabla V}{\partial t} + \frac{1}{c^2}\frac{\partial^2 \boldsymbol{A}}{\partial t^2} = \\ = -\Box \boldsymbol{A} + \nabla\left(\nabla \cdot \boldsymbol{A} + \frac{1}{c^2}\frac{\partial V}{\partial t}\right) = \mu_0 \boldsymbol{j} \end{aligned}$$

[1] It is named after the Danish physicist Ludvig Lorenz.

where $\Box = \Delta - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}$ is the D'Alembert operator or the wave operator. Now, we can make use of the gauge invariance and choose an f to make

$$\nabla \cdot \boldsymbol{A} + \tfrac{1}{c^2}\tfrac{\partial V}{\partial t} = 0. \tag{7.1.6}$$

This gives us a neat equation and as a welcomed consequence this choice also affects equation (7.1.4) since we can substitute the $\nabla \cdot \frac{\partial \boldsymbol{A}}{\partial t}$ and arrive at

$$\Box V = -\frac{\rho}{\varepsilon_0} \tag{7.1.7}$$

$$\Box \boldsymbol{A} = -\mu_0 \boldsymbol{j}. \tag{7.1.8}$$

Electromagnetism turned out to be a gauge theory already from the beginning. However, since the study of external symmetries proved to be extraordinary successful it gave rise to the question if we would be equally rewarded by studying the internal counterparts. Thus when physicist started to explore this field they took on the task to construct new gauge theories from other possible symmetries, and we follow in their footsteps.

7.1.1 Constructing a Gauge Theory

After this appetizer we take on a more general approach to gauge theories and how to construct them. As mentioned in the beginning of this chapter the starting point is a theory invariant under some sort of global transformation of the fields and the construction procedure is to make these transformations local, still keeping the invariance of the theory. The group of transformations is called the **gauge group** and its properties will have a major impact on the theory. We will now study the construction procedure by doing it explicitly for the simplest case.

Suppose we have a theory invariant under global transformations $\mathrm{e}^{i\alpha} \in U(1)$ on the fields of the theory

$$\psi(x) \to \psi'(x) = \mathrm{e}^{i\alpha}\psi(x), \quad \alpha \in \mathbb{R}, \tag{7.1.9}$$

where x is a coordinate in spacetime and α is constant. The invariance means that all equations governing the physics remain unchanged after such a transformation. To upgrade it into a local transform we let $\alpha = \alpha(x)$ depend on the coordinates, making $U(1)$ to our gauge group of interest. The fields still transform as

$$\psi(x) \to \psi'(x) = \mathrm{e}^{i\alpha(x)}\psi(x) \tag{7.1.10}$$

but any expression involving a derivative will no longer transform accordingly since

$$\partial_\mu\psi(x) \to \partial_\mu\psi'(x) = \partial_\mu \mathrm{e}^{i\alpha(x)}\psi(x) = \mathrm{e}^{i\alpha(x)}\partial_\mu\psi(x) + \psi(x)\mathrm{e}^{i\alpha(x)}\partial_\mu i\alpha(x) \tag{7.1.11}$$

which introduce a non-vanishing additional term. Equations involving derivatives are thus no longer invariant and the invariance of our theory is broken. Both the reason and remedy to this issue lie in the definition of the derivative

$$n^\mu\partial_\mu\psi(x) \overset{def}{=} \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}[\psi(x+\varepsilon n) - \psi(x)], \tag{7.1.12}$$

where n is any unit vector. Looking at the right hand side it is now obvious why the derivative does not transform as wished since the two fields have slightly different coordinates and thus

transform under different transformations. Clearly, if we are serious in constructing a gauge theory we need to design a new derivative that transform as the fields, and which measures the change of only the fields and not the change of the transformation group. To fabricate such a derivative we define a way to relate two fields depending on points apart from each other, $\psi(y)$ and $\psi(x)$, through a **comparator** $U(y,x)$ depending on the coordinates y and x. In addition, we also demand it to transform as

$$U(y,x) \to U'(y,x) = \mathrm{e}^{i\alpha(y)} U(y,x) \mathrm{e}^{-i\alpha(x)}. \tag{7.1.13}$$

This rule ensures that $\psi(y)$ and $U(y,x)\psi(x)$ transform in the same way since

$$\psi(y) \to \psi'(y) = \mathrm{e}^{i\alpha(y)}\psi(y), \tag{7.1.14}$$

$$U(y,x)\psi(x) \to U'(y,x)\psi'(x) = \mathrm{e}^{i\alpha(y)} U(y,x) \mathrm{e}^{-i\alpha(x)} \mathrm{e}^{i\alpha(x)} \psi(x) = \mathrm{e}^{i\alpha(y)} U(y,x)\psi(x). \tag{7.1.15}$$

For $U(y,x)$ to make sense we set $U(x,x) = 1$ and we make the assumption that it is a smooth function of the coordinates. As $U(y,x)$ relates two transformations it takes one group element to another and hence it has to be an element of the gauge group. Thus $U(x,x)$ is the identity. We can now use $U(y,x)$ to compare the two fields in the derivative. We define this new derivation operator as

$$n^\mu D_\mu \psi(x) \stackrel{def}{=} \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} [\psi(x + \varepsilon n) - U(x + \varepsilon n, x)\psi(x)] \tag{7.1.16}$$

and notice that it satisfies the desired transformation law

$$D_\mu \psi(x) \to (D_\mu \psi(x))' = \mathrm{e}^{i\alpha(x)} D_\mu \psi(x). \tag{7.1.17}$$

Since $D_\mu \psi(x)$ transforms covariantly we call this new operator **covariant derivative**. However, we still do not have a explicit expression for $U(y,x)$. Using the fact that $U(x + \varepsilon n, x)$ is infinitesimally close to 1 and smooth we may expand it with respect to the first argument (i.e. the derivative only acts on the first argument)

$$\begin{aligned} n^\mu D_\mu \psi(x) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\psi(x + \varepsilon n) - \left(U(x,x) + \varepsilon n^\mu \partial_\mu U(y,x)\Big|_{y=x} + \mathcal{O}(\varepsilon^2) \right) \psi(x) \right] \\ &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\psi(x + \varepsilon n) - \psi(x) - \varepsilon n^\mu \partial_\mu U(y,x)\Big|_{y=x} \psi(x) \right] \\ &= n^\mu \left[\partial_\mu - \partial_\mu U(y,x)\Big|_{y=x} \right] \psi(x) \end{aligned} \tag{7.1.18}$$

and with the definition

$$-\partial_\mu U(y,x)\Big|_{y=x} \stackrel{def}{=} i\lambda A_\mu(x), \tag{7.1.19}$$

where A_μ is a new introduced field and λ is a coupling constant for later convenience, the covariant derivative becomes

$$D_\mu = \partial_\mu + i\lambda A_\mu. \tag{7.1.20}$$

A_μ is a field depending on the spacetime coordinates and is called a **gauge field** or sometimes interaction field. It is the derivative of a transformation in the identity and relates two transformations infinitesimally apart. Thus it takes values in the Lie algebra to the gauge

group and because of its relating role A_μ is also called a **connection**. This term will be given a more geometrical interpretation in the additional section 7.2 about fiber bundles.

A_μ originates from $U(x+\varepsilon n, x)$ and hence its transformation law may be deduced from equation (7.1.13)

$$U(x+\varepsilon n, x) \to U'(x+\varepsilon n, x) = \mathrm{e}^{i\alpha(x+\varepsilon n)} U(x+\varepsilon n, x) \mathrm{e}^{-i\alpha(x)} = (1-\varepsilon n^\mu i\lambda A'_\mu + \mathcal{O}(\varepsilon^2)). \quad (7.1.21)$$

We have

$$\begin{aligned} \mathrm{e}^{i\alpha(x+\varepsilon n)} U(x+\varepsilon n, x)\, \mathrm{e}^{-i\alpha(x)} = \\ &= \left(\mathrm{e}^{i\alpha(x)} + \varepsilon n^\mu \partial_\mu \mathrm{e}^{i\alpha(\tilde{x})}\Big|_{\tilde{x}=x} + \mathcal{O}(\varepsilon^2)\right) \Big(1 - \varepsilon n^\mu i\lambda A_\mu + \mathcal{O}(\varepsilon^2)\Big) \mathrm{e}^{-i\alpha(x)} \\ &= 1 - \mathrm{e}^{i\alpha(x)} \varepsilon n^\mu i\lambda A_\mu \mathrm{e}^{-i\alpha(x)} + \varepsilon n^\mu \partial_\mu (\mathrm{e}^{i\alpha(\tilde{x})})\Big|_{\tilde{x}=x} \mathrm{e}^{-i\alpha(x)} + \mathcal{O}(\varepsilon^2) \quad (7.1.22) \end{aligned}$$

so by equation (7.1.21) discarding terms of order 2 or higher

$$\varepsilon n^\mu i\lambda A'_\mu = \mathrm{e}^{i\alpha(x)} \varepsilon n^\mu i\lambda A_\mu \mathrm{e}^{-i\alpha(x)} - \varepsilon n^\mu \partial_\mu (\mathrm{e}^{i\alpha(\tilde{x})})\Big|_{\tilde{x}=x} \mathrm{e}^{-i\alpha(x)}. \quad (7.1.23)$$

Since all $\mathrm{e}^{i\alpha(x)}$ are just numbers everything commutes which eventually gives us the following transformation rule for the gauge field

$$A_\mu \to A'_\mu = A_\mu - \frac{1}{\lambda}\partial_\mu \alpha(x). \quad (7.1.24)$$

Introducing a gauge field A_μ changes in general what physics the theory describes. This new field interacts with other parts of the theory and may e.g. carry energy within itself. However, since we are demanding the physics to be invariant under the local transformations only those quantities in the theory that ensures this property can be physical. Hence we must find all such quantities that can be constructed from the gauge field. It turns out that there is only one originating from A_μ and it can be found through the following construction[6].

Let us make a loop of comparators comparing in a chain fields of four infinitesimal close points starting and ending at the same point x. Let the points be x, $x+\varepsilon\hat{1}$, $x+\varepsilon\hat{1}+\varepsilon\hat{2}$ and $x+\varepsilon\hat{2}$, where $\hat{1}$ and $\hat{2}$ are two different vectors. Thus the loop is

$$\mathbf{U}(x) = U(x, x+\varepsilon\hat{2})\, U(x+\varepsilon\hat{2}, x+\varepsilon\hat{1}+\varepsilon\hat{2})\, U(x+\varepsilon\hat{1}+\varepsilon\hat{2}, x+\varepsilon\hat{1})\, U(x+\varepsilon\hat{1}, x) \quad (7.1.25)$$

which is illustrated in figure 7.1. This loop transforms covariantly since by (7.1.13)

$$\mathbf{U}(x)\psi(x) \to \mathbf{U}'(x)\psi'(x) = \mathrm{e}^{i\alpha(x)} \mathbf{U}(x)\psi(x) \quad (7.1.26)$$

and may hence contain a physical quantity. To find this we will expand $U(y,x)$ in two steps and make two assumptions. First we require $U(y,x) = (U(x,y))^\dagger$ which is sensible if we have unitary representations of the comparator since for two equal fields, apart from the transformations, we would have

$$\begin{aligned} U(y,x)\psi(x) = \psi(y) \quad \Leftrightarrow \quad \Big(U(y,x)\Big)^\dagger U(y,x)\psi(x) = \Big(U(y,x)\Big)^\dagger \psi(y) \\ \Leftrightarrow \quad \psi(x) = U(x,y)\psi(y). \quad (7.1.27) \end{aligned}$$

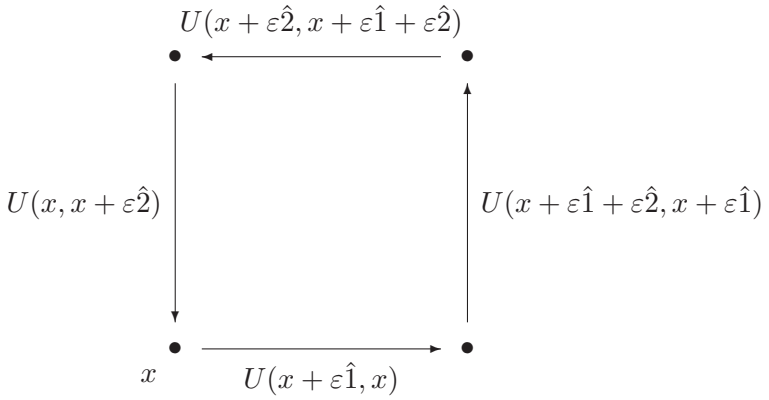


Figure 7.1: A loop of comparators. See text for details.

Second we assume that $U(y,x)$ is a pure phase. By some technical trickery based on these assumptions, which can be found in appendix G, we may expand the comparator between two infinitesimally close points as

$$U(x+\varepsilon n, x) = \exp\left[-i\lambda\varepsilon A_\mu(x+\tfrac{\varepsilon}{2}n) + \mathcal{O}(\varepsilon^3)\right]. \tag{7.1.28}$$

Moreover, by our assumptions we have that

$$U(y,x) = \exp[i\phi(y,x)] = \exp[-i\phi(x,y)] = (U(x,y))^\dagger, \tag{7.1.29}$$

where ϕ is a real valued scalar function. This requires

$$\phi(y,x) = -\phi(x,y) + 2\pi i m, \quad m \in \mathbb{Z} \tag{7.1.30}$$

implying

$$\partial_\mu \phi(y,x) = -\partial_\mu \phi(x,y). \tag{7.1.31}$$

Expanding $\mathbf{U}(x)$ in (7.1.25) in the way of equation (7.1.28) yields

$$\begin{aligned}\mathbf{U}(x) = \exp\Big[&- i\lambda\varepsilon\big[- A_2(x+\tfrac{\varepsilon}{2}\hat{2}) - A_1(x+\tfrac{\varepsilon}{2}\hat{1}+\varepsilon\hat{2}) \\ &+ A_2(x+\varepsilon\hat{1}+\tfrac{\varepsilon}{2}\hat{2}) + A_1(x+\tfrac{\varepsilon}{2}\hat{1})\big] + \mathcal{O}(\varepsilon^3)\Big]\end{aligned} \tag{7.1.32}$$

where equation (7.1.31) has been used for the first two terms. The indices 1 and 2 denote the componentes in direction $\hat{1}$ and $\hat{2}$. Note that we silently have used the Baker-Campbell-Hausdorff formula in equation (2.3.21) for the commuting A's. Now we can expand and get rid of the half-ε terms in the argument through

$$\begin{aligned}\mathbf{U}(x) = \exp\Big[&- i\lambda\varepsilon\big[- A_2(x) - \tfrac{\varepsilon}{2}\partial_2 A_2(x) + A_2(x+\varepsilon\hat{1}) + \tfrac{\varepsilon}{2}\partial_2 \overbrace{A_2(x+\varepsilon\hat{1})}^{A_2(x)+\varepsilon\partial_1 A_2(x)+\ldots} \\ &- A_1(x+\varepsilon\hat{2}) - \tfrac{\varepsilon}{2}\partial_1 \underbrace{A_1(x+\varepsilon\hat{2})}_{A_1(x)+\varepsilon\partial_2 A_1(x)+\ldots} + A_1(x) + \tfrac{\varepsilon}{2}\partial_1 A_1(x)\big] + \mathcal{O}(\varepsilon^3)\Big] = \\ = \exp\Big[&- i\lambda\varepsilon\big[- A_2(x) + A_2(x+\varepsilon\hat{1}) - A_1(x+\varepsilon\hat{2}) + A_1(x)\big] + \mathcal{O}(\varepsilon^3)\Big]\end{aligned} \tag{7.1.33}$$

and then in turn expanding the exponential in ε we find

$$\mathbf{U}(x) = 1 - i\lambda\varepsilon^2 \left[\partial_1 A_2(x) - \partial_2 A_1(x)\right] + \mathcal{O}(\varepsilon^3). \tag{7.1.34}$$

This expression reveals the invariant quantity: the field strength

$$F_{\mu\nu} \stackrel{def}{=} \partial_\mu A_\nu - \partial_\nu A_\mu, \tag{7.1.35}$$

also called the **curvature** for reasons explained in the fiber bundle section 7.2.

There is yet another way to find this tensor where the gauge invariance is as apparent as above. The covariant derivative of anything that transform covariantly also transform covariantly. Hence the second covariant derivative also follows the same transformation law (7.1.17) and we can form the commutator transforming as

$$[D_\mu, D_\nu]\psi(x) \to \mathrm{e}^{i\alpha(x)}[D_\mu, D_\nu]\psi(x). \tag{7.1.36}$$

However, writing out the commutator we find that it is in fact no derivative at all since

$$\begin{aligned}[D_\mu, D_\nu]\psi = [\partial_\mu, \partial_\nu]\psi + i\lambda\left([\partial_\mu, A_\nu] - [\partial_\nu, A_\mu]\right)\psi - \lambda^2[A_\mu, A_\nu]\psi \\ = i\lambda\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right)\psi.\end{aligned} \tag{7.1.37}$$

The left hand side transforms according to equation (7.1.36) and transforming the right hand side using the transformation law for A_μ in (7.1.24) everything cancel but the transformation of ψ. Hence the commutator of the covariant derivative is invariant under the local transformations and we may equally well define the field strength as

$$i\lambda F_{\mu\nu} = [D_\mu, D_\nu]. \tag{7.1.38}$$

From this follow that any function of $F_{\mu\nu}$ and its derivatives is gauge invariant. These two ways of deriving $F_{\mu\nu}$ is in fact equivalent since the commutator can be viewed geometrically as the comparator loop.

So far in our theory we have defined a new derivative operator which introduced a gauge field, which in turn brought new physics into the theory. But still, we have not yet found any equation to govern the dynamics of the theory. This will be the topic of the subsequent section.

7.1.2 Construction Through the Lagrangian Formalism

A fruitful manner to construct gauge theories is through the principle of least action[2]. We denote the action as

$$I = \int \mathcal{L}(x)\,\mathrm{d}^4x \tag{7.1.39}$$

where $\mathcal{L}(x)$ is the Lagrangian, or rather the Lagrangian density which is also integrated over space. We work here with natural units in which $c = \hbar = 1$ making the Lagrangian a dimensionless quantity.

In classical mechanics the Lagrangian is the difference between kinetic and potential energy, but to construct new theories we need new Lagrangians. Being a scalar the Lagrangian

[2]The reader is assumed to be familiar with the action principle and the calculus of variation, otherwise see appendix H for a brief introduction.

is manifestly Lorentz invariant and by only including terms which keep the action gauge invariant all equation derived from it will be invariant as well. The big issue here is thus to find these terms. The criteria for the terms are that they have to be scalars and keep the action gauge invariant. Moreover, for the theory to be renormalizable[3] only terms with dimension down to length^{-4} is allowed since otherwise the associated coupling constants need to have dimension of a positive power of length in order to preserve the dimensionlessness of the Lagrangian. This is because an additional length^4 comes from the measure in the integral and thus the 4 is due to the dimension of spacetime, other numbers are relevant when working with other spaces. A positive power of length of the coupling constant leads to a dependence on the cut-off and thus to a non-renormalizable expression. Further restrictions on the terms follow if one requires the theory to be invariant under parity and time reversal transformations.

When faced with the task of finding the Lagrangian one usually have some experience of the formalism from other theories and one also knows that the equations it produces must be consistent with already verified theories. This is great guidance in a quest like ours and even if the reader might be unfamiliar with other results, such as the Klein-Gordon equation for scalar fields, we will still let ourself be inspired by them, after a short introduction. The Klein-Gordon equation is the first relativistic version of the Schrödinger equation for spinless particles and reads in natural units

$$\partial_\mu \partial^\mu \phi = m^2 \phi \tag{7.1.40}$$

following from the Lagrangian

$$\mathcal{L}_{\text{K-G}}(x) = \partial_\mu \phi^* \partial^\mu \phi - m^2 \phi^* \phi. \tag{7.1.41}$$

The particles are here described by complex scalar fields ϕ. The Klein-Gordon equation exhibit a global $U(1)$ symmetry but to remain invariant under local transformations all derivatives has as learnt to be replaced by covariant derivatives. The first term in the Klein-Gordon Lagrangian is called the kinetic term and involves, as is very reasonable for a kinetic energy term, first order derivatives. We also know that we need such a term with first derivatives to yield a second order differential equation for the dynamics, so we ought to look for something mimicing this one. The second term is the mass term in analogy to the classical potential energy and one can naively think of it as the total mass energy times the probability distribution of the particle's position.

Now, our Lagrangian is to consist of the fields, A_μ and their derivatives. Trying to find the terms involving A_μ a good starting point is to find something similar to the kinetic term in the Klein-Gordon Lagrangian. We thus need a scalar valued, gauge invariant combination of first order derivatives of A_μ. We know of such a quantity and the first natural try would be something proportional to $F_{\mu\nu}F^{\mu\nu}$. This satisfies all our requirements so far and as the dimension is of length^{-4} it is renormalizable with a dimensionless coupling constant. Hence we have

$$\mathcal{L}_{\text{kinetic}} = kF_{\mu\nu}F^{\mu\nu}, \quad k \in \mathbb{R}. \tag{7.1.42}$$

[3]In perturbation calculations in these theories integrals commonly diverge. Renormalization is a way around this through introducing a cut-off in the integrals, continue the calculations and let the cut-off tend to infinity once the final result is obtained. If such a result is independent of the cut-off the theory is said to be renormalizable.

Next we search for an analogue of the mass term with the obvious first choice $m^2 A^*_\mu A^\mu$. However, this term violates the gauge invariance which easily can be seen through the transformation law for A_μ ((7.1.24)). In fact, the requirement of a gauge invariant Lagrangian rule out all combinations of A_μ in our $U(1)$-example but $F_{\mu\nu}$. This forces us the conclusion that our gauge field is a massless quantity. Even though less relevant for the rest of this section the most general gauge invariant and renormalizable Lagrangian for fermions under the gauge group $U(1)$ is [6]

$$\mathcal{L} = m^2\psi^*\psi + \psi^* i\gamma^\mu D_\mu \psi - \frac{1}{4}\left(F_{\mu\nu}\right)^2 - c\epsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F^{\mu\nu} \tag{7.1.43}$$

where γ^μ is the Dirac matrices [4], ψ are Dirac spinors and $\epsilon^{\alpha\beta\mu\nu}$ is the totally antisymmetric tensor equal to -1 for all odd permutations of the indices. However, as mentioned it is not crucial that the Lagrangian is gauge invariant as long as it holds for the action. We are now about to pursue such an additional term which in itself is not gauge invariant but keeps that property of the action.

Recalling Noether's theorem from chapter 6 we ought to have a conserved quantity following from this continuous symmetry we are fabricating. This is indeed true and we may find it by studying how the action transforms. As we already have presumed we have a theory invariant under a global transformation of the fields which infinitesimally can be written

$$\psi_l \to \psi_l' = \psi_l + i\varepsilon T_l(x) \tag{7.1.44}$$

where we have gathered all the details of the transformation in the function $T_l(x)$ depending on both the coordinates, the transformation and the transformed field. The dependence on the coordinates follows from that the *same* transformation results in different added terms at different points as the field varies over spacetime. The transformation is thus still global. The index l runs over the relevant fields of the theory. The variation of the action I under this transformation is then[5]

$$\delta I = \int \mathrm{d}^4x \left(\frac{\delta\mathcal{L}}{\delta\psi_l} i\varepsilon T_l(x) + \frac{\delta\mathcal{L}}{\delta\partial_\mu\psi_l} i\varepsilon\partial_\mu T_l(x)\right) = 0. \tag{7.1.45}$$

The equality to zero is due to the requirement that the action is invariant under the specific global transformation and holds even if the fields would not satisfy the Euler-Lagrange equations. Now, making the transformation local we let $\varepsilon = \varepsilon(x)$ depend on the coordinates and once again vary the action. In addition to the terms in the equation (7.1.45), which still vanish through the same requirement, we also get

$$\frac{\delta I}{\delta\psi_l} = \int \mathrm{d}^4x \frac{\delta\mathcal{L}}{\delta\partial_\mu\psi_l} T_l(x)\partial_\mu i\varepsilon(x). \tag{7.1.46}$$

Integrating by parts and letting $\varepsilon(x)$ vanish at infinity yields

$$\frac{\delta I}{\delta\psi_l} = -i\int \mathrm{d}^4x \partial_\mu \left(\frac{\delta\mathcal{L}}{\delta\partial_\mu\psi_l} T_l(x)\right)\varepsilon(x) \tag{7.1.47}$$

[4] This should not bother the unacquainted reader.

[5] Here δ denotes functional derivative but readers unfamiliar with functionals can just think of it as a partial derivative.

and if we now demand the fields to satisfy Euler-Lagrange equations the action has to be invariant under any infinitesimal transformation, including local ones. Hence we have found a conserved quantity, that is

$$\partial_\mu J^\mu \overset{def}{=} \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta \partial_\mu \psi_l} T_l(x) \right) = 0. \tag{7.1.48}$$

This quantity, often referred to as the **conserved current**, will come in handy in our search for the Lagrangian term. Suppose this term $\mathcal{L}_0$ depends on the gauge field A_μ but not on any of its derivatives. The variation of the action under the transformation in equation (7.1.24) then becomes

$$\delta I = \int \mathrm{d}^4 x \left(\frac{\delta \mathcal{L}_0}{\delta A_\mu} (-\frac{1}{\lambda} \partial_\mu \alpha(x)) + \frac{\delta \mathcal{L}_0}{\delta \partial_\nu A_\mu} (-\frac{1}{\lambda} \partial_\nu \partial_\mu \alpha(x)) \right) = -\frac{1}{\lambda} \int \mathrm{d}^4 x \frac{\delta \mathcal{L}_0}{\delta A_\mu} \partial_\mu \alpha(x) \tag{7.1.49}$$

and through a partial integration we get

$$\delta I = \frac{1}{\lambda} \int \mathrm{d}^4 x \partial_\mu \left(\frac{\delta \mathcal{L}_0}{\delta A_\mu} \right) \alpha(x). \tag{7.1.50}$$

For this to be zero, i.e. the action to be invariant, we must have

$$\partial_\mu \left(\frac{\delta \mathcal{L}_0}{\delta A_\mu} \right) = 0. \tag{7.1.51}$$

We know already of such a quantity and setting $\delta \mathcal{L}_0 / \delta A_\mu$ proportional to J^μ the action will be invariant. Since the dimension of the term is length^{-4} the proportionality constant can be absorbed in λ and we simply set

$$\frac{\delta \mathcal{L}_0}{\delta A_\mu} = J^\mu \tag{7.1.52}$$

which gives us the term $\mathcal{L}_0 = J^\mu A_\mu$.

We have now found two terms satisfying the requirements and by reasons which will become apparent we choose the constant in the kinetic term to $k = \frac{1}{4}$. We have the simple Lagrangian

$$\mathcal{L} = -\frac{1}{4} F^{\mu\nu} F_{\mu\nu} + J^\mu A_\mu. \tag{7.1.53}$$

We will now illustrate the Lagrangian way of constructing gauge theories by continuing the $U(1)$-example above and find the governing equations. We will use our Lagrangian from equation (7.1.53) where the first term is gauge invariant while the second, being the $\mathcal{L}_0$ from above, is not. To once more verify that the action is invariant we have from the transformation law (7.1.24) that the non-trivial part of the action transforms as

$$\begin{aligned} S = \int \mathrm{d}^4 x J^\mu A_\mu \to \int \mathrm{d}^4 x (J^\mu A_\mu - J^\mu \frac{1}{\lambda} \partial_\mu \alpha(x)) \\ = /\text{p.i}/ = \int \mathrm{d}^4 x J^\mu A_\mu - \int \mathrm{d}^4 x \partial_\mu \left(J^\mu \frac{1}{\lambda} \alpha(x) \right) + \int \mathrm{d}^4 x \frac{1}{\lambda} \alpha(x) \partial_\mu J^\mu. \end{aligned} \tag{7.1.54}$$

If J^μ is required to vanish at infinity the middle term is zero on the boundaries, the third term is zero by equation (7.1.48) and then we are left with only the first term. Hence the action is invariant under the gauge transformation. Proceeding in the spirit of the principle of least action we calculate the variation

$$\delta S = \int \mathrm{d}^4x \left[\frac{\delta \mathcal{L}}{\delta A_\nu} \delta A_\nu + \frac{\delta \mathcal{L}}{\delta(\partial_\mu A_\nu)} \delta(\partial_\mu A_\nu) \right] = 0 \tag{7.1.55}$$

and as usual we integrate partially

$$\delta S = \int \mathrm{d}^4x \partial_\mu \left[\frac{\delta \mathcal{L}}{\delta(\partial_\mu A_\nu)} \delta A_\nu \right] + \int \mathrm{d}^4x \delta A_\nu \left[\frac{\delta \mathcal{L}}{\delta A_\nu} - \partial_\mu \frac{\delta \mathcal{L}}{\delta(\partial_\mu A_\nu)} \right] = 0. \tag{7.1.56}$$

Demanding δA_ν to be zero at infinity leaves us with the second term and for this to be zero we must have

$$\frac{\delta \mathcal{L}}{\delta A_\nu} = \partial_\mu \frac{\delta \mathcal{L}}{\delta(\partial_\mu A_\nu)}. \tag{7.1.57}$$

Looking at each side at a time we find from equation (7.1.53) that

$$\frac{\delta \mathcal{L}}{\delta A_\nu} = J^\nu, \qquad \frac{\delta \mathcal{L}}{\delta(\partial_\mu A_\nu)} = -\frac{1}{4} \frac{\delta \left(F_{\alpha\beta} \eta^{\alpha\rho} \eta^{\beta\sigma} F_{\rho\sigma} \right)}{\delta \partial_\mu A_\nu} = \tag{7.1.57a, b}$$

$$= -\frac{1}{4} \eta^{\alpha\rho} \eta^{\beta\sigma} \left[F_{\rho\sigma} \left(\delta^\alpha_\mu \delta^\beta_\nu - \delta^\beta_\mu \delta^\alpha_\nu \right) + F_{\alpha\beta} \left(\delta^\rho_\mu \delta^\sigma_\nu - \delta^\sigma_\mu \delta^\rho_\nu \right) \right] =$$

$$= -\frac{1}{4} \left[F^{\mu\nu} - F^{\nu\mu} + F^{\mu\nu} - F^{\nu\mu} \right] = /\text{antisymmetry}/ =$$

$$= -F^{\mu\nu}$$

which plugged into (7.1.57) finally yields

$$\partial_\mu F^{\mu\nu} = -J^\nu \qquad \Longleftrightarrow \qquad \partial_\mu F^{\nu\mu} = J^\nu. \tag{7.1.58}$$

As told in the introduction and as the notation already may have hinted $U(1)$ is precisely the gauge group in electromagnetic theory. The gauge field $A = (V, \boldsymbol{A})$ is the 4-potential with the electric potential V and the magnetic vector potential $\boldsymbol{A}$ as components. The field strength $F_{\mu\nu}$ is the electromagnetic field tensor

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \tag{7.1.59}$$

which can easily be verified through equations (7.1.2) and (7.1.35). Note that in electromagnetism as a gauge theory the potentials are more fundamental quantities than the electric and magnetic fields. Initially this arose objections but a clever experiment proposed in 1959 by Aharanov and Bohm proved the effect of the potential on electrons even in the complete absence of electric and magnetic fields. The experiment involved a long, impenetrable coil containing an isolated magnetic field onto which electron where scattered. The electrons did not experience any magnetic field at all but still the vector potential surrounding the coil

resulted in a shift of the scattering pattern. Thus the gauge field has been proven to be the truly fundamental field.

The coupling constant λ is in this case of electromagnetism the elementary charge e and the continuous gauge symmetry is through Noether's theorem the reason to the conservation of electric charge. Because electromagnetism was the first gauge theory the coupling constant in general sometimes is called charge. J^μ is the 4-current $j^\mu = (\rho c, \boldsymbol{j})$ where ρ is the charge density and $\boldsymbol{j}$ the current density. The name interaction field now gets its explanation since A_μ describes the massless photon, the electromagnetic interaction particle.

In this light, we see that the transformation law (7.1.24) for A_μ is in fact precisely the gauge freedom (7.1.3) in our introductory example, just shifting A_μ with the derivative of a scalar function $A'_\mu = A_\mu + \partial_\mu \alpha(x)$, and choosing a specific transformation $\mathrm{e}^{i\alpha(x)}$ is the same as fixing the gauge.

Our derived equations (7.1.58) are the inhomogeneous Maxwell equations in tensor notation while the homogeneous ones follow directly from the definition $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. Hence the construction of a gauge invariant action through the Lagrangian density has given us the complete gauge theory electromagnetism. Since electromagnetism was gauge invariant from the very beginning we have just derived it from another principle but still this demonstrates how powerful the formalism may be for constructing new theories.

7.1.3 Non-Abelian Gauge Groups

So far everything has been explained in the light of the Abelian $U(1)$ as gauge group but the generalization to groups of higher rank involves few surprises. In our exposure we follow the outline of the previous sections pointing out the differences. Accordingly, we start with some Lie group $\mathcal{G}$ as gauge group and let the fields transform as

$$\psi \to \psi' = g\psi, \qquad g \in \mathcal{G} \tag{7.1.60}$$

and require our theory of interest to be invariant under these global transformations. Note that the fields may be vectors and that group elements act on them through representations of the same dimension, making it possible for non-commuting transformations. As in the section above we promote this to a local transformation by letting $g = g(x)$ depend on the coordinates. We can express this in terms of the generators t_a of the Lie algebra $\mathfrak{g}$, $a = 1, 2, \ldots, \dim \mathfrak{g}$, to get the local transformation law

$$\psi \to \psi' = \mathrm{e}^{i\alpha^a(x)t_a}\psi \tag{7.1.61}$$

where $\alpha^a(x)$ are real valued functions of the coordinates and the summation over a is understood. Once more we need a covariant derivative and define a comparator as $U(y, x)$ with the transformation law

$$U(y, x) \to U'(y, x) = g(y)U(y, x)g^{-1}(x). \tag{7.1.62}$$

We set $U(x, x) = \hat{\mathbb{1}}$ and may expand every small displacements in the first argument in terms of the generators

$$U(x + \varepsilon n, x) = \hat{\mathbb{1}} + i\lambda\varepsilon n^\mu A^a_\mu(x)t_a + \mathcal{O}(\varepsilon^2) \tag{7.1.63}$$

which now introduce a new gauge field A^a_μ for every generator in the Lie algebra. Note that the gauge fields are as in the case of $U(1)$ scalars and it is the linear combination of $A^a_\mu t_a$

which is Lie algebra valued (recall that the generator of $U(1)$ is i). With this expansion in a definition of the covariant derivative adopted from (7.1.16) we get the covariant derivative

$$D_\mu = \partial_\mu - i\lambda A^a_\mu t_a \tag{7.1.64}$$

in analogy to (7.1.18). To find the transformation law for the gauge fields we make use of the work done in equations (7.1.22) and (7.1.23) and note that the commuting property of $U(1)$ actually was not used until later. Hence we can simply restate the result in equation (7.1.22) as

$$A^a_\mu t_a \to A'^a_\mu t_a = g(x)A^a_\mu t_a g^{-1}(x) - \frac{i}{\lambda}\partial_\mu(g(\tilde{x}))\Big|_{\tilde{x}=x} g^{-1}(x). \tag{7.1.65}$$

and through $\partial_\mu(g)g^{-1} = \partial_\mu(gg^{-1}) - g\partial_\mu g^{-1} = -g\partial_\mu g^{-1}$ we get the more appealing look

$$A^a_\mu t_a \to A'^a_\mu t_a = g(x)\left(A^a_\mu t_a + \frac{i}{\lambda}\partial_\mu\right) g^{-1}(x). \tag{7.1.66}$$

If the transformations are infinitesimal we can expand them to the first order of α^a which yields the infinitesimal version

$$\begin{aligned} A^a_\mu t_a \to A'^a_\mu t_a &= (1 + i\alpha^a t_a)A^a_\mu t_a(1 - i\alpha^a t_a) + (1 + i\alpha^a t_a)\frac{i}{\lambda}\partial_\mu(1 - i\alpha^a t_a) = \\ &= A^a_\mu t_a + \left[i\alpha^a t_a, A^b_\mu t_b\right] + \frac{1}{\lambda}\partial_\mu\alpha^a t_a = \\ &= A^a_\mu t_a - \alpha^a A^b_\mu f_{ab}{}^c t_c + \frac{1}{\lambda}\partial_\mu\alpha^a t_a \end{aligned} \tag{7.1.67}$$

where $f_{ab}{}^c$ are the structure constants of the algebra. Note that this in the case of an Abelian gauge group yields our former rule in equation (7.1.24) (the sign is changed because of the different definitions of the gauge fields).

This transformation rule ensures that the covariant derivative of a field transforms under infinitesimal transformations as

$$D_\mu\psi \to (D_\mu\psi)' = (1 + i\alpha^a t_a)D_\mu\psi \tag{7.1.68}$$

which can be verified through direct computation.

Next step is to derive the field strength tensor, or curvature. We now adopt the second approach from the Abelian procedure and construct the commutator for the covariant derivative. Still it transforms covariantly as in

$$[D_\mu, D_\nu]\psi(x) \to [D'_\mu, D'_\nu]\psi'(x) = g(x)[D_\mu, D_\nu]\psi(x) \tag{7.1.69}$$

and we define the field strength tensor as

$$-i\lambda F^a_{\mu\nu}t_a \overset{def}{=} [D_\mu, D_\nu] \tag{7.1.70}$$

in analogy to (7.1.35). Here as well we find that it is no derivative but a constant, in this case a matrix:

$$\begin{aligned} F^a_{\mu\nu}t_a &= \partial_\mu A^a_\nu t_a - \partial_\nu A^a_\mu t_a - i\lambda[A^a_\mu t_a, A^b_\nu t_b] \\ &= \partial_\mu A^a_\nu t_a - \partial_\nu A^a_\mu t_a + \lambda A^a_\mu A^b_\nu f_{ab}{}^c t_c \end{aligned} \tag{7.1.71}$$

that is

$$F^c_{\mu\nu} = \partial_\mu A^c_\nu - \partial_\nu A^c_\mu + \lambda A^a_\mu A^b_\nu f_{ab}{}^c. \tag{7.1.72}$$

This expression differs from the Abelian case in equation (7.1.35) through the commutator and recalling that we used the Baker-Campbell-Hausdorff in the comparator loop we should not be that surprised. The transformation law follows from (7.1.69) and reads

$$F^a_{\mu\nu} t_a \to F'^a{}_{\mu\nu} t_a = g(x) F^a_{\mu\nu} t_a g^{-1}(x) \tag{7.1.73}$$

or infinitesimally

$$F^a_{\mu\nu} \to F'^a{}_{\mu\nu} = F^a_{\mu\nu} - \alpha^b F^c_{\mu\nu} f_{bc}{}^a, \tag{7.1.74}$$

which means that the invariance in the Abelian case no longer holds. There are however still ways to construct invariant terms in the Lagrangian from F. The kinetic term for A^a_μ is ([6] p. 489)

$$\mathcal{L}_{\text{kinetic}} = -\frac{1}{4} F^a_{\mu\nu} F_a^{\mu\nu} = -\frac{1}{4}\text{Tr}(F^2). \tag{7.1.75}$$

This term gives rise to the equation of motion

$$\partial^\mu F^a_{\mu\nu} + \lambda f^a{}_{bc} A^b_\mu F^c_{\mu\nu} = D^\mu F^a_{\mu\nu} = 0 \tag{7.1.76}$$

which differ from the Maxwell equations in (7.1.58) ($j^\mu = 0$) by terms cubic and quartic in the gauge field which involve the gauge groups structure constants. These terms describe how the gauge fields interact with themselves, a new feature distinguishing the new fields from the non-interacting photons. In this way the symmetry group determines the physics to a large extent and the time development of a system depend directly on its structure. However, there is still no allowed mass term for the gauge fields so the interaction particles of our theory is massless. This was a major drawback of the early gauge theories and because of this Wolfgang Pauli who inspired Yang and Mills even refrained to publish any of his results on the subject. But the solutions to the mass problem came some decade later through symmetry breaking. This however falls out of our scope here.

When the possibility of massive gauge fields were discovered the interest of Yang-Mills theories was substantially renewed. During the 60's and 70's they became the base of the standard model and of the classification of elementary particles. These classifications will be the topic of the following chapter 8 about particle physics and as a conclusion of this one we list the three relevant gauge theories for the standard model.

Example 7.1.2. Electrodynamics, $\mathcal{G} = U(1)$: The gauge group of electromagnetism with one gauge field A_μ describing the massless photon. The gauge field has no self-interaction and the field strength tensor is simply $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ since the group is Abelian, i.e. $f^{abc} = 0$. Fields of the theory transform as $\psi \to \mathrm{e}^{i\alpha(x)}\psi$ and the gauge field as $A_\mu \to A_\mu - \frac{1}{e}\partial_\mu \alpha(x)$.

Example 7.1.3. Isotopic gauge invariance, $\mathcal{G} = SU(2)$: This theory contains three gauge fields as there are three generators for $SU(2)$. The fields describe the three bosons of weak interactions and the generators are the Pauli matrices giving the transformations $\psi \to \exp[i\alpha^j(x)\sigma_j/2]\psi$ and $A^j_\mu \to A^j_\mu + \frac{1}{\lambda}\partial_\mu \alpha^j + \epsilon^{jkl} A^k_\mu \alpha^l$.

Example 7.1.4. Quantum chromodynamics, $\mathcal{G} = SU(3)$: The eight generators give rise to eight gauge fields A^a_μ describing the gluons, bosons of the strong interaction. The generators are the Gell-Mann matrices λ_a and the transformations are $\psi \to \exp[i\alpha^a(x)\lambda_a/2]\psi$ and $A^a_\mu \to A^a_\mu + \frac{1}{\lambda}\partial_\mu \alpha^a + f^a{}_{bc} A^b_\mu \alpha^c$. [6]

[6] λ without any index is still the coupling constant.

7.2 *Fiber Bundles

This section is a non-rigorous introduction to fiber bundles and the material presented here is not necessary for understanding the following chapter. The idea is just to offer a more geometrical picture of what we have done in this chapter and the reader may jump with no loss of the track to the next chapter. This section will however also function as an example on how groups and group theory are natural parts of other mathematical branches, as the projective representations illumined some of their topological aspects. In the presentation several technicalities are omitted and the aim is merely to present the ideas enough to get a grasp of how to formulate the mathematical structure behind gauge theory. However, to really get the geometrical pictures a number of apperently unrelated definitions and purely mathematical examples are required before the connection to previous sections can be made. A reader already familiar with the concepts and notions in fiber bundles may thus jump to section 7.2.2 for the actual discussion of gauge theories.

Before we state the definition of a fiber bundle we prepare for the concept with an example and some notation. A fiber bundle consists mainly of three manifolds[7] and mappings between them so let us create a construction of such things which will lead us to the definition.

Example 7.2.1. Consider a manifold M, e.g. a sphere. To each point $x \in M$ there is a tangent space which we denote T_xM, i.e. the space of all tangent vectors to paths passing through x. The union of all these tangent spaces we denote TM. The manifold M over which TM is defined is called a base space. We require M to have a set of open coverings, $\{U_i\}$, which means that to every $x \in M$ there exists an open set U_i which includes x. One can think of it as a patchwork covering the whole manifold M. To each U_i there is a subset $T_{U_i}M \subset TM$ and we require that the relation between $T_{U_i}M$ and U_i could be expressed as the Cartesian product $U_i \times T_{U_i}M$. This means that

$$T_{U_i}M \stackrel{def}{=} \bigcup_{x \in U_i} T_xM \tag{7.2.1}$$

and that every element p of $T_{U_i}M$ can be described by the ordered pair (x, v) where v is a vector in T_xM. Thus every $p \in T_{U_i}M$ is associated with a point $x \in M$ and we can define a projection $\pi : T_{U_i}M \to M$ such that $\pi(p) = x$, projecting (x, v) to the first factor (thus loosing all information but x). The inverse projection naturally associates x with the tangent space T_xM, i.e. $\pi^{-1}(x) = T_xM$. This is an example of what will be called a fiber over x. The geometrical situation is illustrated in figure 7.2.

We can now handle the relations between our elements within a set U_i but for our construction to be a fiber bundle we need ways to move between the different patches. These will be provided by something called transition maps defined in the definition of a fiber bundle, which we now are ready for.

Definition 7.2.1. *A (differentiable) fiber bundle $(E, \pi, M, F, \mathcal{G})$ consist of the following elements*

(i) A differentiable manifold E, called the **fiber bundle** *(or total space).*

[7] Manifolds are discussed in appendix A.

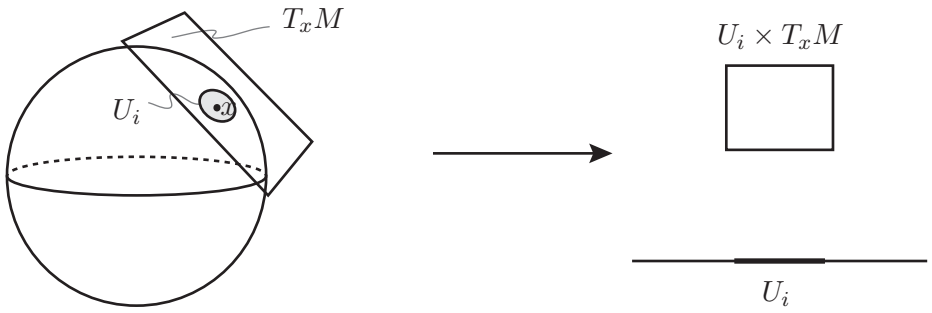


Figure 7.2: A sphere and its tangent space T_xM at x which locally can be viewed as a Cartesian product $U_i \times T_xM$ between some neighbourhood U_i to x and the tangent space.

(ii) *A differentiable manifold M, called the* **base space**.

(iii) *A differentiable manifold F, called the* **fiber**.

(iv) *A surjection (onto map) $\pi : E \to M$, called the* **projection**, *with the inverse $\pi^{-1}(x \in M) \stackrel{def}{=} F_x \cong F$. F_x is called the fiber at x.*

(v) *A group*[8] *$\mathcal{G}$ acting on F from the left. $\mathcal{G}$ is called the* **structure group**.

(vi) *A set of open coverings $\{U_i\}$ and a smooth map (diffeomorphism) $\phi_i : U_i \times F \to \pi^{-1}(U_i)$ such that $\pi \circ \phi_i(x, f) = x$, with $x \in U_i, f \in F$.*

(vii) *Denote the restriction of ϕ_i to the point $x \in U_i$ by $\phi_{i,x}$. Then $\phi_{i,x} : F \to F_x \subset E$. Now, to every pair of neighborhoods $U_i \cap U_j \neq \emptyset$ there is a map $t_{ij} = \phi_{i,x}^{-1}\phi_{j,x} : F \to F$ required to be an element of $\mathcal{G}$. Such a map is called a* **transition map** *or transition function. With $x \in U_i \cap U_j$ and $f \in F$ this can also be expressed as $\phi_{i,\,x}^{-1}\phi_{j,\,x}(f) = t_{ij}(x)f$. The transitions map satisfies*

a) $t_{ii}(x) = 1$

b) $t_{ij}(x) = t_{ji}^{-1}(x)$

c) $t_{ik}(x) = t_{ij}(x)t_{jk}(x)$ *for* $U_i \cap U_j \cap U_k \neq \emptyset$.

The fiber bundle or the total space is thus a manifold which to every point x in the base space (through the inverse projection π^{-1}) assigns a fiber F_x. Each such fiber is a manifold isomorphic to F. To get a geometrical picture one often visualizes the fiber bundle to reside above the base space and in this schematic picture one can see the fiber manifolds as one-dimensional fibers rising up from each point in the base space. This intuitive picture is illustrated in figure 7.3. It is often a helpful image even though the fiber manifolds in general have higher dimensions. There will follow similar illustrations through out this section but as for this one, they all should be considered as mere pictures of the concepts.

The smooth map ϕ_i is called to be a local trivialization since it allows us to look at the subspace $\pi^{-1}(U_i) \subset E$ as a simple Cartesian direct product $U_i \times F$. This map is such that

[8] We are almost exclusively dealing with Lie groups.

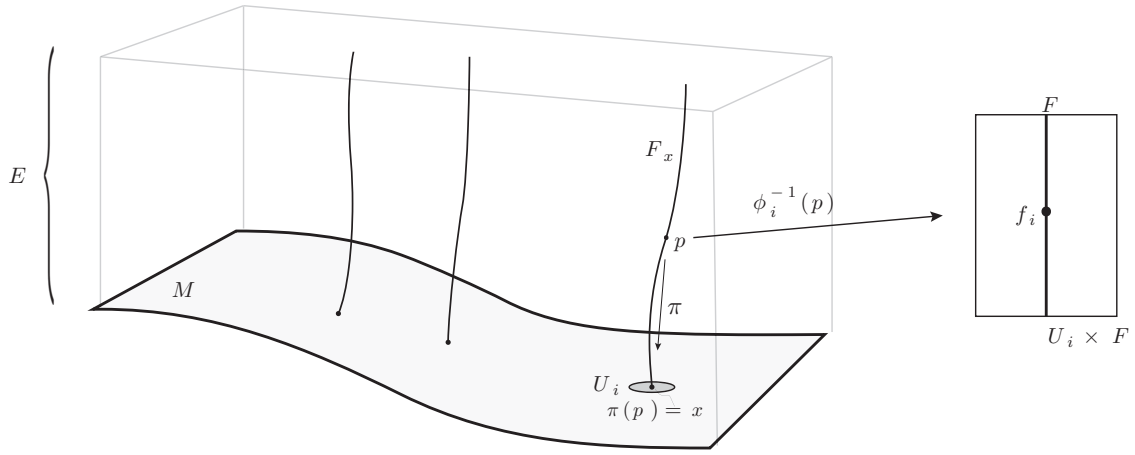
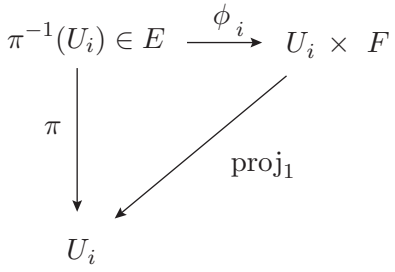


Figure 7.3: A schematic picture of a fiber bundle. See text for details.

the projection π corresponds to the projection on the first factor in the Cartesian product $\text{proj}_1(U_i \times F) = U_i$ on the other side of ϕ. That is $\pi(p \in \pi^{-1}(U_i)) = \text{proj}_1(\phi^{-1}(p))$ which can be illustrated as



where the scheme commutes. The ϕ_i could equally well be defined to map in the opposite direction and it might help in this first acquaintance to point out that ϕ^{-1} is as important.

If the relation between the fiber bundle and the base space is such that it can be expressed as a Cartesian product globally, $E = M \times F$, then the fiber bundle is called a **trivial bundle**. In a non-trivial case one can look at the fiber bundle E as a patchwork of local trivial spaces $U_i \times F$ glued together in a certain way. This way is precisely what the transition maps describes.

Let x be a point in $U_i \cap U_j \subset M$ and a point $p = \pi^{-1}(x) \in E$ be associated by the local trivialization ϕ_i with the ordered pair (x, f_i) and at the same time by ϕ_j^{-1} to (x, f_j), with $f_i, f_j \in F$. Then f_i and f_j are related to one other by the transition map as

$$f_i = t_{ij}(p) f_j, \tag{7.2.2}$$

which can easily be seen from the definition $t_{ij} = \phi_{i,x}^{-1} \phi_{j,x}$. Thus the transition maps describes how one moves from one local trivial space to another. These relations are illustrated in figure 7.4.

As a first example on a non-trivial bundle we may take the Möbius strip above the base space S^1, in contrast to the trivial cylinder.

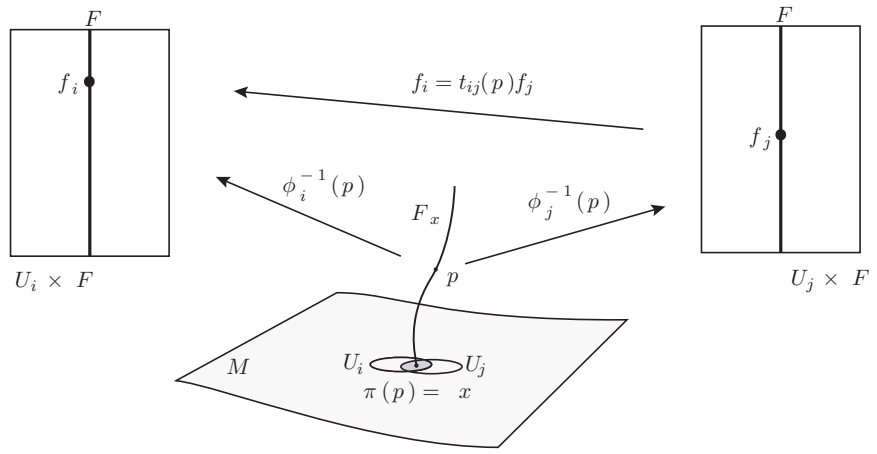
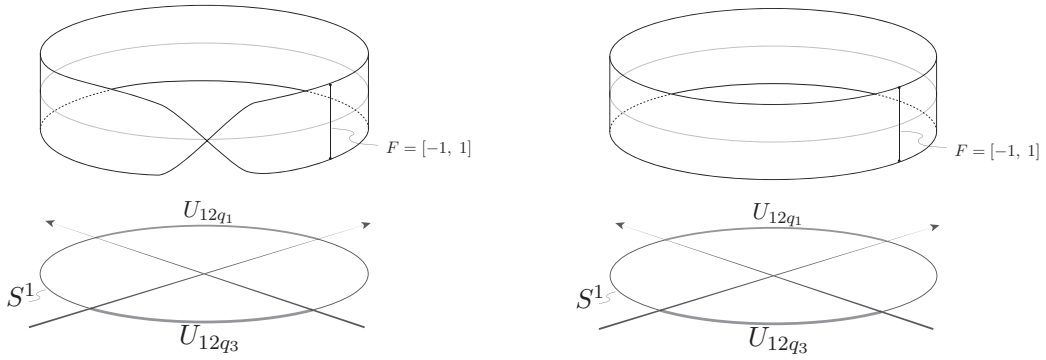


Figure 7.4: An illustration of how the transition map t_{ij} relates the two local trivializations over U_i and U_j. See text for details.



(a) The Möbius strip as a non-trivial fiber bundle over S^1.

(b) The trivial fiber bundle over S^1 is a cylinder.

Figure 7.5: A simple example of the difference between trivial and non-trivial fiber bundles. See text for details.

Example 7.2.2. Take the manifold S^1 as a base space and the Möbius strip as a fiber bundle with fibers consisting of the interval $[-1, 1]$. Then each open set $\pi^{-1}(U_i)$ (the inverse projection of an arc of the circle) maps by ϕ_i to the Cartesian product $U_i \times [-1, 1]$. The construction is illustrated in the figure 7.5a. If we instead of the Möbius strip would have a cylinder as the fiber bundle this would be a global relation and the cylinder is thereby the trivial bundle over S^1, as in figure 7.5b.

What makes the Möbius strip globally different than the cylinder is the twist and this non-triviality appears in the transition maps. Let us as an example of this define two open overlapping sets on S^1 as $U_1 = (0, \frac{3}{2}\pi)$ and $U_2 = (-\pi, \frac{1}{2}\pi)$. This gives the two intersections $U_{12q_1} = (0, \frac{1}{2}\pi)$ and $U_{12q_3} = (\pi, \frac{3}{2}\pi)$ in which the transition maps t_{12q_1} and t_{12q_3} can be defined as

$$t_{12q_1} : [-1, 1] \to [-1, 1], \qquad t_{12q_3} : [-1, 1] \to [-1, 1] \tag{7.2.3}$$

$$t \mapsto t \qquad t \mapsto -t. \tag{7.2.4}$$

Here $t \in [-1, 1]$ is used to parameterize the fibers. It is the minus sign in t_{12q_3} which makes the non-trivial twist in the Möbius strip. Replacing it with a plus, i.e. making t_{12q_3} to the identity map, would yield the trivial cylinder bundle. The structure group of our example is the mirror group $\mathbb{Z}_2$ and clearly t_{12q_1} and t_{12q_3} are the two elements of this group with $t_{12q_3}^2 = t_{12q_1} = e$.

We can now look back at our introductory example 7.2.1 and identify TM as the total space E, M as the mentioned base space and T_xM as the fiber F_x at x. The identification between $T_{U_i}M$ and $U_i \times T_xM$ is handled by ϕ_i. This bundle is called the **tangent bundle** and since the fibers in this case are vector spaces it is a **vector bundle**.

We move on with another important definition.

Definition 7.2.2. *A* **section** *is a smooth map $\sigma : M \to E$ such that $\pi \circ \sigma(x) = x$.*

Intuitively the section assigns to every $x \in M$ a point $f \in F_x$ in the fiber over x. A path $\gamma(t) \subset M$ thus maps by σ to a specific path $\tilde{\gamma}(t) \subset E$, illustrated in figure 7.6. In this way a section is quite analogous to the graph of a function. Take the ordered pair $\sigma(x) = (x, f(x)) \in E$ for some $f(x) \in F$, then the projection onto the first factor $\pi(\sigma(x)) = x$ is like it is for a standard function graph $(x, g(x)) \in X \times Y$. In the case of our tangent bundle a section of E assigns to every point $x \in M$ a specific tangent vector in T_xM. This is just what we recognize as a vector field and thus a vector field can be described as a section in a tangent bundle.

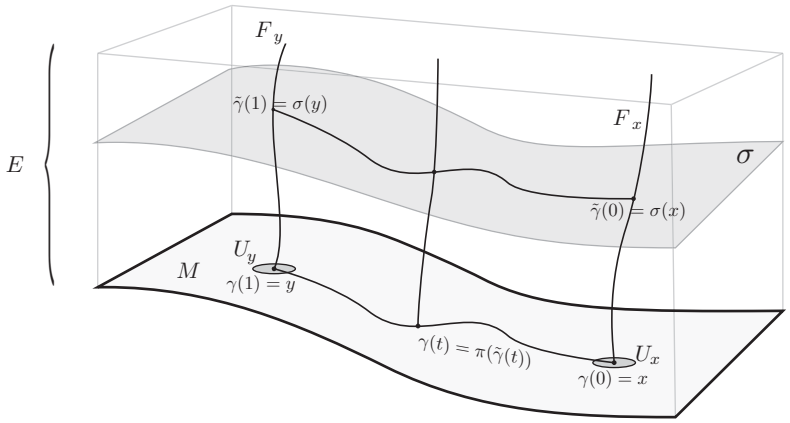


Figure 7.6: A section σ of a fiber bundle is an assignment of a point $p \in E$ to every $x \in M$.

As two points on a fiber always is related by an element of the structure group this expands to an entire section. This means that given a section σ_1 any new section σ_2 can be defined through the first by

$$\sigma_2 = g(x)\sigma_1 \qquad g(x) \in \mathcal{G} \tag{7.2.5}$$

where the group element g depend on coordinates of the base space.

In general it is not necessarily possible to define a section globally on a base space and thus it is useful to define local sections on open sets $U_i \subset M$.

In the context of gauge theory spacetime (or what ever space the theory describes) is the base space and the coordinate dependent transformations are elements of the fibers. The

fibers are in fact themselves copies of the gauge group and hence there are a copy of the group assigned to every point in spacetime. To choose a section of the bundle is the same thing as fixing the gauge.

In physics, and gauge theory in particular, one is often concerned with this special case of fiber bundles, called principal bundles.

7.2.1 Principal Bundles

Definition 7.2.3. *A* **principal bundle** P *is a fiber bundle in which the structure group* $\mathcal{G}$ *is a topological group to which the fibers are identical. In the smooth category* $\mathcal{G}$ *is required to be a Lie group. In addition there is a continuous right action denoted*

$$\begin{aligned}&\Phi : \mathcal{G} \times P \to P,\\ &\Phi(g,p) = \Phi_g(p) = pg, \quad g \in \mathcal{G}, p \in P\end{aligned}$$

which preserves the fibers of P *and acts freely and transitively*[9] *on them. The right action is defined through the local trivialization as*

$$pg = \Phi_g(p) = \phi_i(x, f_i g)$$

where $\phi_i(x, f_i) = p \in P$.

Once again we may make use of the simplifying visualization in figure 7.3 and think of Φ as pushing a point $p \in P$ *along* the fiber thus preserving the projection $\pi(\Phi_g(p)) = \pi(p)$. Although the right action is defined in terms of the local trivializations it is in fact independent of the actual map since

$$pg = \phi_i(x, f_i g) = \phi_i(x, t_{ij} f_j g) = \phi_j(x, f_j g) = pg. \tag{7.2.6}$$

For a principal bundle it is often convenient to identify the fibers F directly with the group $\mathcal{G}$.

We now arrive at the point where the name of the gauge field as a "connection" will get its explanation. We first define the term in the context of fiber bundles and then relate it to the known gauge field. There are in fact two ways of defining the connection and we will present them both. The first is from a geometrical point of view and the latter is through differential forms where we will try to keep the formalism to a minimum.

Geometrical Approach
We precede the definition with some notation. We denote the tangent space to a point p in the principal bundle as T_pP. The subspace of T_pP which is projected onto $x = \pi(p)$ we call the **vertical space**, V_pP. This is parallel to the fiber F_x and can be written as $V_pP = \ker(\pi)$. Thus a vector in V_pP is a tangent vector to the fiber, i.e. the structure group $\mathcal{G}$ as P is a principal bundle, and we have isomorphism between V_pP and the Lie algebra $\mathfrak{g}$ of $\mathcal{G}$. It is important not to confuse T_pP with the tangent space to M.

[9] An action is free if for any $f \in F$ $fg = fh$ implies $g = h$, with $g, h \in \mathcal{G}$. It is transistive if for any $f_1, f_2 \in F$ there exists a $g \in \mathcal{G}$ such that $f_1 g = f_2$.

Definition 7.2.4. *A* **connection** *on a principal bundle P is a smoothly-varying assignment to each point $p \in P$ of a subspace H_pP of the tangent space T_pP at p such that*

$$\begin{aligned} &(i) \qquad T_pP = H_pP \oplus V_pP \quad \forall p \in P \\ &(ii) \qquad \Phi_{g*}(H_pP) = H_{pg}P \quad \forall p \in P, \forall g \in \mathcal{G}. \end{aligned}$$

In words (i) says that H_pP is the orthogonal complement to $V_pP = \ker(\pi)$, i.e. transverse to the fiber. (ii) says that pushing the subspace H_pP along the fiber with Φ_{g*}[10] yields the same space as to first push the point p to $\Phi_g(p)$ and there assign a subspace $H_{\Phi_g(p)}P$.

(i) can also be reformulated as

$$(i)' \qquad \pi_*(H_pP) = T_{\pi(p)}M \qquad \forall p \in P. \tag{7.2.7}$$

The effect of the connection on the tangent space T_pP is schematically depicted in figure 7.7.

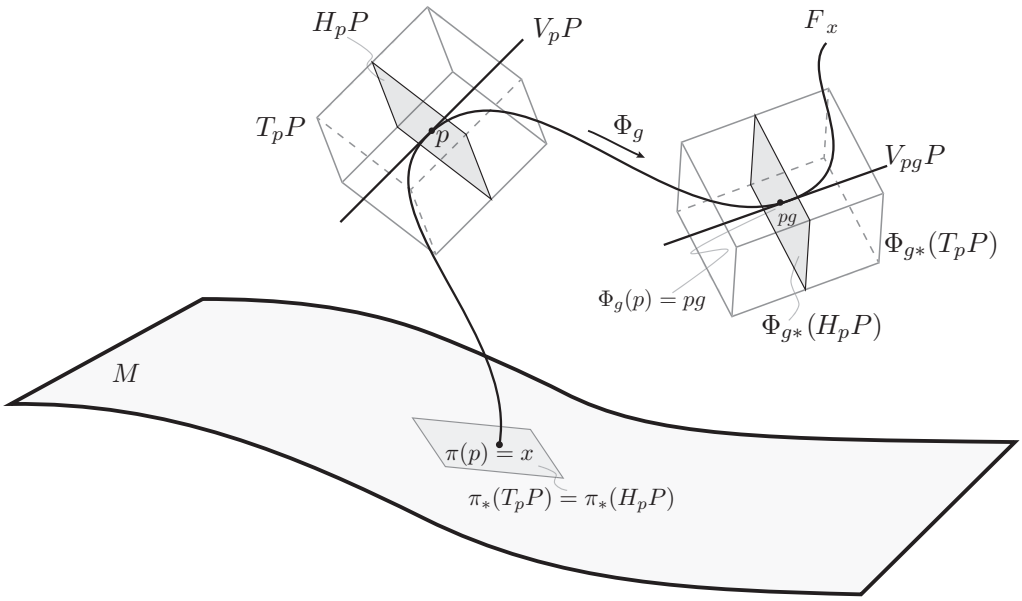


Figure 7.7: A connection decomposes the tangent space T_pP at $p \in P$ into a vertical space V_pP and horizontal space H_pP and commutes with the right action Φ.

Differential Approach

The differential approach to connections requires some knowledge of differential geometry. We will define the connection as a **differential one-form** without going into any details about forms. If the term is unfamiliar the reader can think of it as a linear functional from a vector space to its field of scalars, or equivalently as an element of the dual vector space

[10] Whenever a function acts on an entire set we add a $*$ as a subscript.

(see appendix D). The one-form of the connection is however combined with the Lie algebra through a tensor product which makes it Lie algebra valued. It will function as a projection from the tangent space T_pP onto the vertical space V_pP and to be able to specify this we introduce a vector field[11] X over P. It is at every point p a tangent to the fiber and is defined by the right action as

$$Xf(p) = \frac{\mathrm{d}}{\mathrm{d}t} f(p \exp[t\xi])\Big|_{t=0}, \qquad \xi \in \mathfrak{g} \tag{7.2.8}$$

where $f : P \to \mathbb{R}$ is an arbitrary function and t is a parameter. This field allows for the definition of the differential connection.

Definition 7.2.5. *A* **connection 1-form** $\omega_p \in \mathfrak{g} \otimes T_p^*P$ *on* P *is a Lie algebra-valued 1-form* $w_p : T_pP \to \mathfrak{g}$*, satisfying for each* $p \in P$

$$\begin{aligned} &(i) \qquad \omega_p(X(p)) = \xi && \forall \xi \in \mathfrak{g} \\ &(ii) \qquad \omega_{pg}(\Phi_{g*}v) = \mathrm{Ad}_g(\omega_p(v)) && \forall g \in \mathcal{G}, \forall v \in T_pP. \end{aligned}$$

As mentioned the connection is a projection onto V_pP through the isomorphism between the vertical space and the Lie algebra. This is the first requirement. The second requirement (ii) says that doing a projection of the pushed vector $\Phi_{g*}v$ yields the same Lie algebra element as if going back with g^{-1} apply ω_p to v and then move forward along the fiber with g. Thus (ii) can be rewritten

$$(ii) \qquad \omega_{pg}(\Phi_{g*}v) = g^{-1}(\omega_p(v))g \quad \forall g \in \mathcal{G}, \forall v \in T_pP.$$

If we now define the horizontal space as

$$H_pP \stackrel{def}{=} \ker \omega_p \tag{7.2.9}$$

the connection one-form divides the tangent space into a vertical space and a horizontal space with the same properties as in definition 7.2.4. Thus the two view points give the same result. The connection one-form is illustrated in figure 7.8 which displays the kinship of the two approaches.

Still we have not yet explained the relation to the gauge field. This is defined on spacetime which plays the role of the base space while the connection ω projects in the bundle. The missing step before we can see the gauge field as a connection is to define a Lie algebra-valued one-form $\mathcal{A}_i$ on a tangent space to a subset U_i of the base space, i.e.

$$\mathcal{A}_i : TU_i \to \mathfrak{g}. \tag{7.2.10}$$

This can be made through our connection ω in the bundle and the concept of **pullback**, which first is to be explained. Let σ_i be a local section on U_i. We then denote the pullback on any function f by this section as $\sigma_i^* f$ which means informally that f takes its arguments from the range of σ_i and that the pullback of f has the same domain as σ_i. As an example the pullback of a local trivialization takes points in the base space to an ordered pair with a point in the fiber according to $\sigma_i^*\phi_i : U_i \to \sigma_{i*}(U_i) \to U_i \times F$. The simplest case of a pullback

[11] This field is called the *fundamental vector field* in the literature.

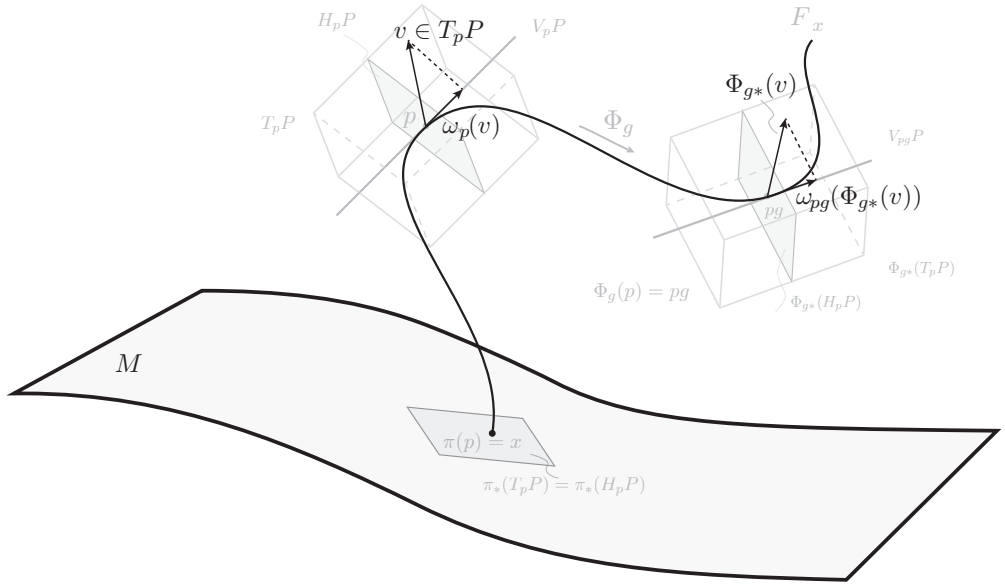


Figure 7.8: A connection one-form ω_p projects the tangent space T_pP onto the vertical space V_pP. The horizontal space H_pP is defined as its kernel which yields the same result as in figure 7.7.

is the composition of two functions $f(g(x))$ in elementary calculus. Now, we can define $\mathcal{A}_i$ through a pullback of the connection by σ_i as

$$\mathcal{A}_i \overset{def}{=} \sigma_i^* \omega \in \mathfrak{g} \otimes T^*U_i. \tag{7.2.11}$$

Conversely we can define a connection given a Lie algebra-valued one-form $\mathcal{A}_i$ and a local section over U_i. That this is true and that we indeed can define a global connection based on local $\mathcal{A}_i$'s can be shown quite straight forward ([7] pp. 333-335). However, for ω to be unique this implies $\mathcal{A}_i$ to undergo a specific transformation law when moving between different sections. In terms of the transition functions this law reads

$$\mathcal{A}_j = t_{ij}^{-1} \mathcal{A}_i t_{ij} + t_{ij} \, \mathrm{d}_P t_{ij}^{-1} \tag{7.2.12}$$

involving the *exterior derivative*[12] d_P on P. This seems familiar and if we let two sections σ_i and σ_j be related as $\sigma_j(x) = \sigma_i(x)g(x)$, recalling equation (7.2.5), then the transformation law in components becomes

$$\mathcal{A}_{j\mu} = g^{-1}(x)\mathcal{A}_{i\mu}g(x) + g^{-1}(x)\partial_\mu g(x). \tag{7.2.13}$$

This is precisely the transformation law for the gauge field we saw in equation (7.1.66) and in fact $\mathcal{A}_i$ *is* the gauge field through the relation

$$\mathcal{A}_i(x) = i\lambda A_\mu^a(x) t_a \, \mathrm{d}x^\mu. \tag{7.2.14}$$

[12]We will make no further use of the exterior derivative why we omit any discussion of it. We leave it with the remark that it turns a scalar function into a one-form as the gradient of the function, giving the unacquainted reader a slight hunch.

We have noted that changing section is the same thing as choose a new gauge, and accordingly the transformation laws are the same.

We have in the previous section seen how the covariant derivative algebraically removed the change of the transformations and thus only measured the rate of change of the fields. We can now see this geometrically. The connection divides T_pP into H_pP and V_pP and thus every tangent vector v into horizontal and vertical components $v_H + v_V$. If v is the directional derivative of the transformed field then the vertical component describes the change of the transformation. Removing the projection onto the vertical component thus only measure the change of the fields and this is the geometrical picture behind the equation $D_\mu = \partial_\mu - i\lambda A^a_\mu t_a$. This relates very much to the notions horizontal lift and parallel translation which are the last introduced notions in this section.

7.2.2 Horizontal Lifts and Parallel Translation

We move right onto the definitions.

A **lift** of a path $\gamma(t) \subset M$ in the base space is a path $\tilde{\gamma}(t) \subset P$ in the bundle such that $\pi(\tilde{\gamma}(t)) = \gamma(t)$, i.e. a path which is projected down on $\gamma(t)$.

Definition 7.2.6. *A **horizontal lift** $\bar{\gamma}(t) : [0,1] \to P$ with respect to a connection ω is a lift such that its tangent vector at p always lies in H_pP for all $p \in P$ and $t \in [0,1]$. Equivalently, it is a lift such that*

$$\omega\left(\frac{\mathrm{d}}{\mathrm{d}t}\bar{\gamma}(t)\right) = 0 \qquad \forall t \in [0,1]. \tag{7.2.15}$$

An image of a horizontal lift is found in figure 7.9. The definition allows us to construct a horizontal lift $\bar{\gamma}(t)$ of a path in M in infinitesimal steps, respecting the equation (7.2.15) at each step. In accordance with the relations between two different sections in equation (7.2.5) we can express the horizontal path as

$$\bar{\gamma}(t) = \sigma_*(\gamma(t))g(t) \tag{7.2.16}$$

for a section σ and some varying $g(t) = g(\gamma(t)) \in \mathcal{G}$. We can choose these such that $\bar{\gamma}(0) = \sigma_*(\gamma(0))$, that is $g(\gamma(0)) = e$. Since the $\bar{\gamma}(t)$ corresponds to another section than σ a similar relation holds between tangent vectors to $\bar{\gamma}(t)$ and $\sigma_*(\gamma(t))$ as between $\mathcal{A}_j$ and $\mathcal{A}_i$ in equation (7.2.12). We will not state that relation here but by derivating it with respect to t (denoted with a dot) and applying ω on both sides yields the differential equation

$$\frac{\mathrm{d}g(t)}{\mathrm{d}t} = -\omega\left(\sigma_*(\dot{\gamma}(t)\right) = -\mathcal{A}\left(\dot{\gamma}(t)\right). \tag{7.2.17}$$

The details on this are given in [7] (pp. 336-337). This equation is solvable and together with the condition $g(0) = e$ we get

$$g(t) = \mathrm{P}\exp\left[-\int_0^t \mathcal{A}_\mu \frac{\mathrm{d}x^\mu}{\mathrm{d}t}\,\mathrm{d}t\right] = \mathrm{P}\exp\left[-\int_{\gamma(x)}^{\gamma(t)} \mathcal{A}_\mu(\gamma(t))\,\mathrm{d}x^\mu\right]. \tag{7.2.18}$$

The P means that the exponential is path ordered, i.e. its expansion is done in a way which respects the path[13]. This is important since Lie algebra elements in $\mathcal{A}$ may not commute.

[13] An expression for path ordered expansion is given in appendix G, in fact of this specific integral.

The horizontal lift is now given by the expression

$$\bar{\gamma}(t) = \sigma_*(\gamma(t))\mathrm{P}\exp\left[-\int_{\gamma(x)}^{\gamma(t)} \mathcal{A}_\mu(\gamma(t))\ \mathrm{d}x^\mu\right]. \tag{7.2.19}$$

The path ordered exponential can be viewed as a measurement of how much $\sigma_*\gamma(t)$ deviates from being horizontal. In the same way, when integrating from a point x to another point y in the base space, the exponential provides a measure of how far we have traveled in the fiber when we arrive at y. Such a shift is seen in figure 7.9. Thus it relates the different transformations associated to the two points and that sounds very much like a comparator. This is more than a resemblance as equation (7.2.14) yields

$$U_\gamma(y,x) = \mathrm{P}\exp\left[-\int_x^y \mathcal{A}_\mu(\gamma(t))\ \mathrm{d}x^\mu\right] = \mathrm{P}\exp\left[-i\lambda\int_x^y A_\mu(x')\,\mathrm{d}x'^\mu\right], \tag{7.2.20}$$

integrating along some path γ. One can easily verify that this expression indeed satisfies what we imposed in previous sections, and it is the same as what appeared in the expansion of U in equation (7.1.28) (see appendix G). The right hand side of equation (7.2.20) is called a Wilson line and as mentioned it depends on the path. This has the important implication that it is non-trivial for closed loops. If we close the path $\gamma(0) = \gamma(1)$ and thus create a loop in M the integral, now a Wilson loop, will not vanish. Geometrical this corresponds to that the horizontal lift of γ does not close but starts and ends at different points but on the same fiber. The net transformation along the fiber is called the **holonomy** of the connection around γ and is illustrated in figure 7.10.

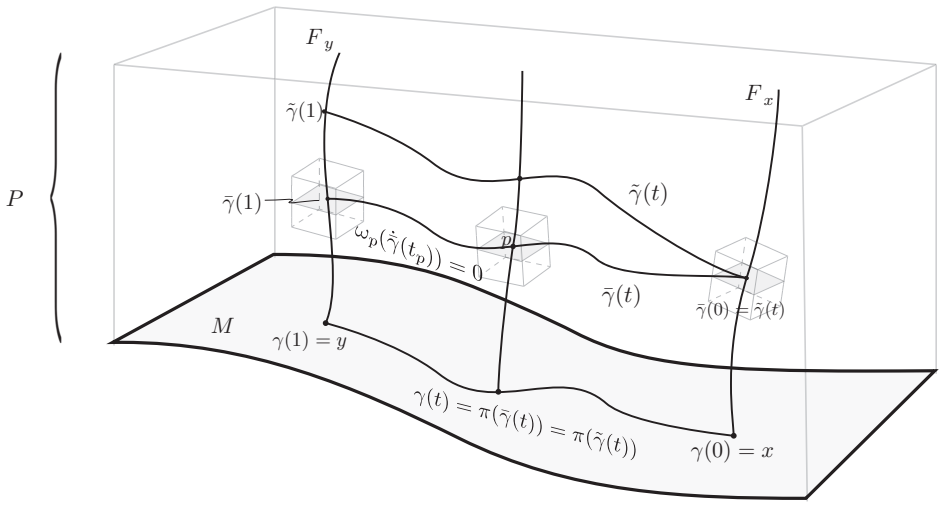


Figure 7.9: A horizontal lift $\bar{\gamma}(t)$ is defined such that its tangent vector $\dot{\tilde{\gamma}}(t_p)$ at p always lies in H_pP, for all $p \in \bar{\gamma}(t)$. Note how the non-horizontal path $\tilde{\gamma}(t)$ differs from $\bar{\gamma}(t)$ which can be expressed as in (7.2.16).

Evaluating some quantity along a horizontal path is called a **parallel transport** or parallel translation. This is affected by the holonomy and an accessible geometric picture of this is

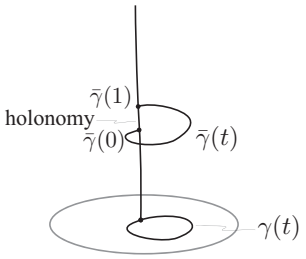
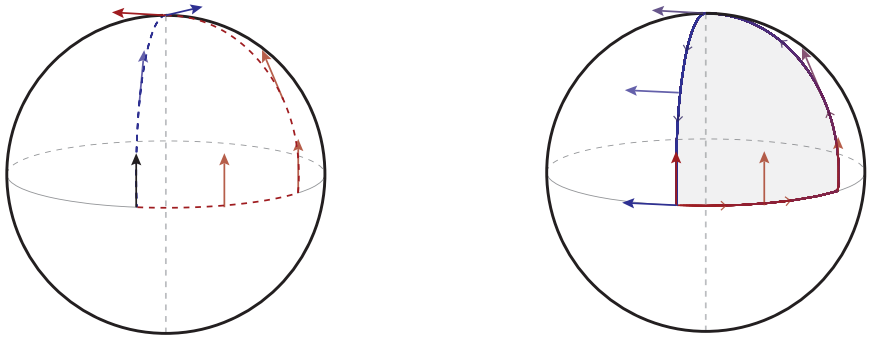


Figure 7.10: A horizontal lift of a closed path does not necessarily start and end at the same point on the fiber. The net transformation is called holonomy.



(a) If a vector is parallel transported along different paths on a sphere its direction at the end point depends on the path taken.

(b) A vector is parallel transported along a closed path on a sphere and returns to the start point in an angle to the original position. This angle is related to the enclosed area shaded in the figure.

Figure 7.11: The effect of curvature on parallel translation. See text for details.

given by the example of a tangent bundle to a sphere. Opposed to a flat surface there is no direct way of translating a vector on a sphere. However, a parallel transport along a path is a translation under which the covariant derivative of the vector is zero, which corresponds to a horizontal lift. This means that the vector does not change when one moves from one tangent space to another infinitesimally apart. However, parallel transports along different paths do not necessarily end in the same vector. In the figure 7.11a translations of a vector along two different paths from the equator to the north pole is illustrated. One is going straight north while the other moves along the equator for some distance before turning north. Arriving at the pole the vectors are pointed in different directions, although both are parallel transported. This is a very physical manifestation of the path dependence of the Wilson line. Connecting the paths yields a loop and the holonomy of the connection around this is depicted in figure 7.11b. The effect of holonomy is due to the curvature of the sphere and the magnitude of it is related to the area enclosed by the loop.

We will end this introduction to fiber bundles by illustrating this area dependence in an example of a Wilson loop, referring back to the preceding sections. This will shed some light

on the last unmentioned quantity from the review of gauge theories. Let A_μ be the gauge field emerging from the local symmetry of $U(1)$, as in electromagnetism. If we calculate the Wilson loop for a closed path γ in spacetime we may apply the Stoke's theorem,

$$\mathrm{P}\exp\left[-i\lambda\oint_x^y A_\mu(x')\,\mathrm{d}x'^\mu\right] = \mathrm{P}\exp\left[-i\frac{\lambda}{2}\int_S F_{\mu\nu}\,\mathrm{d}s^{\mu\nu}\right], \tag{7.2.21}$$

where S is the area enclosed by γ, $\mathrm{d}s^{\mu\nu}$ an infinitesimal area element and $F_{\mu\nu}$ the field strength. This directly exposes the relation between the holonomy and the enclosed area and introduces the field strength in the context of fiber bundles. For an infinitesimal loop this is precisely what we did in section 7.1.1 when we constructed the loop of comparators, which should be apparent from (7.2.20). The field strength $F_{\mu\nu}(x)$ is thus shown to be a measure of the curvature locally around x and hence the name given in section 7.1.1 has found its reason.

We have now given a geometrical interpretation of all the quantities in the preceding sections of this chapter. For a much more thorough treatment of fiber bundles and differential geometry in physics [7] is recommended and on which much of this exposure is based.

Leaving the underlaying mathematical structure behind gauge theories we now turn towards some of its very important applications. The following chapter will discuss the results of the application of gauge theory in particle physics.

Chapter 8

Particle Physics

We have now arrived at the last chapter of this survey. During our journey we have learned the fundamentals of group theory and representation theory, explored both external and internal symmetries and seen numerous examples of the natural merging of the aforementioned mathematical disciplines with physical theories such as quantum physics and special relativity.

It is a continuous endeavor of science to connect seemingly random pieces of knowledge into a single clear image of reality. The most complete picture of nature that we have today is called the *standard model of particle physics* and it incorporates everything that we have discussed so far. Since no introductory text on group theory and symmetries in nature would be complete without it, let us therefore explore this theory.

We will not attempt to fully explain the standard model, but rather try to describe how such a theory can be constructed and outline its connection to what we have learned so far. In order to do so, we will introduce a concept called *isospin* which allows us to perceive particles belonging to a set of particles called a *charge multiplet* as different states of a single particle corresponding to distinct values of the *isospin projection number*. In essence, this means that a particle can be "rotated" into another particle through the action of elements of the underlying set of a representation of the group $SU(2)$, in much the same way as an ordinary spin vector is rotated!

Moreover, by exploring the internal structures of certain composite particles, such as neutrons and protons, and by introducing additional quantum numbers, we will be able to group these particles into *supermultiplets*. We will then find the connection between these supermultiplets and the weight and root diagrams of chapter 5. However, we must first acquaint ourselves with elementary particles and the standard model!

8.1 Elementary Particles

An **elementary particle** is by definition a point-like object without any internal structure or excited states. Such a particle is characterized by, among other things, its spin, rest mass and electric charge. Since it has the same nature in every inertial frame one may also define an elementary particle as an object which is invariant under Poincaré transformations. The latter observation led Wigner to the conclusion that the quantum mechanical states for a given elementary particle must be elements of a Hilbert space which is the module of a representation of the Poincaré group (this was discussed in section 6.2).

Previously we have dealt with external symmetries and internal symmetries separately but

to completely describe elementary particles both of them are required. In 1967 the physicists Sidney Coleman and Jeffrey Mandula proved a theorem on the impossibility of combing the Poincaré symmetry group, describing external symmetries, and any symmetry group encoding internal symmetries described with gauge theory in any but a trivial way. This implies that there is no group which have representations that describe both of these phenomena. Hence, it follows that all elementary particles are irreducible representations of the symmetry group

$$ISO(1,\,3) \times \mathcal{G}_{\text{int.sym}} \tag{8.1.1}$$

where $\mathcal{G}_{\text{int.sym}}$ is some internal symmetry group in direct product with the Poincaré group[1].[25] Since we have a direct product, this means that there is no mixing between the generators of the different groups. This theorem is known as the **Coleman-Mandula theorem** and it is of vital importance in contemporary research in supersymmetry. The interested reader is referred to [26] for a deeper discussion.

One may wonder what the internal symmetry group in equation (8.1.1) is? According to the standard model the answer is:

$$\mathcal{G}_{\text{int.sym}} = U(1) \times SU(2) \times SU(3). \tag{8.1.2}$$

Let us now take a closer look at this theory!

8.2 The Standard Model

Since ancient times mankind has wondered what the world is composed of and what holds it together. The modern answer to these fundamental questions has been named the **standard model** of particle physics. It is a theory which tries to explain all known phenomena in particle physics, except for those caused by gravity, in terms of the properties and interactions of elementary particles. Gravity is not included in the standard model because general relativity has yet to be successfully quantized.

Particles with half-integer spin are called **fermions** while particles with integer spin are called **bosons**. The standard model divides all known elementary particles into two distinct **families** of spin-1/2 fermions called **quarks** and **leptons** and one spin-1 family of bosons (i.e. vector bosons) known as **gauge bosons** or "force carriers". In addition to these particles at least one spin-0 elementary particle called the **Higgs boson** is postulated to exist. This particle is necessary in the standard model in order to explain the origin of other particle's rest masses and it is widely believed that the discovery of the Higgs boson will be announced by CERN in the near future!

Quarks and leptons may also be divided into three **generations** of particles (se figure 8.1) and the atoms of the universe are composed entirely by particles from the first generation of the standard model. It is a mystery why nature is equipped with three families when only one seems to be needed.

The electron is a lepton and because of its electric charge it is a subject to the **electromagnetic interaction** which is one of the four known forces of nature. Neutrinos are also leptons and they are for instance produced in β decays. Such decays are caused by a second fundamental force known as the **weak interaction**.

[1]Their paper investigates the nature of the *S-matrix*, which is an array of transition amplitudes between different states corresponding to different particles (it expresses scattering processes). By studying its symmetries it is possible to draw conclusions regarding the explicit behavior of spacetime and internal symmetries.

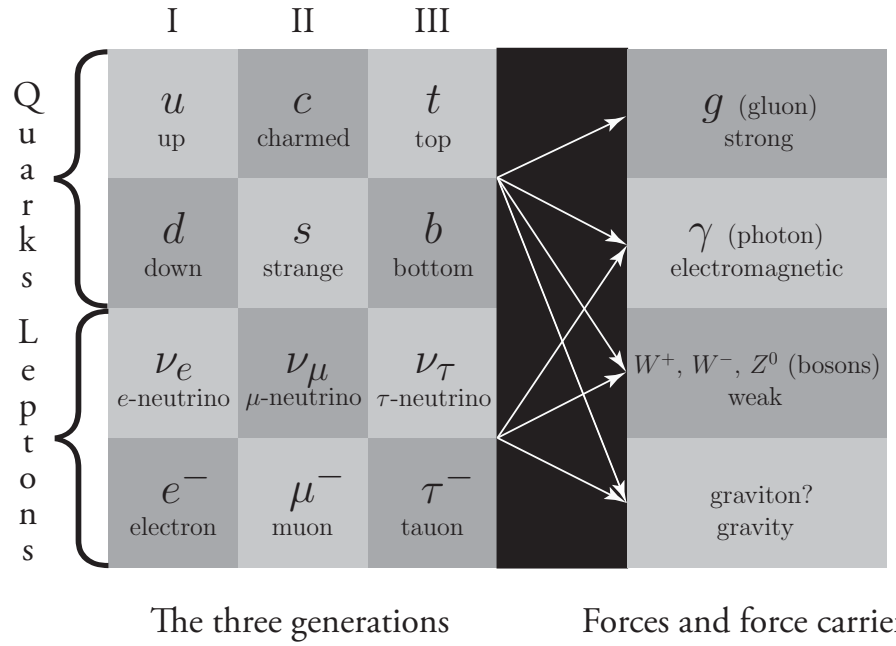


Figure 8.1: The three generations of particles in the standard model of particle physics. The table also shows the interactions which the particles participate in (including gravity) and the corresponding force carriers.

Quarks are always found in bound states called **hadrons**. The proton and the neutron are both hadrons and there are literally *hundreds* of more (excluding atomic nuclei) and most of them are unstable. Hadrons are held together by a third force known as the **strong interaction**. The fourth and final force is the **gravitational interaction**. A complete description of nature would have to include this force as well, but it is so weak compared to the others that it may, for all practical purposes, be omitted in particle physics. In this field one therefore usually speaks of the *three*, rather than four, forces of nature. We will follow this convention throughout this chapter.

In the standard model forces between elementary particles are caused by the exchange of force carriers. The electromagnetic interaction is due to the exchange of *virtual* photons. We refer to them as **virtual** particles since they are only involved in the interactions and do not appear as products in the final state. The same can be said about other force carriers. The weak force is mediated by the exchange of W^+, W^- and Z^0 bosons (where the superscripts denote the particle's electric charges) while the force carriers for the strong force are called **gluons**. There are eight different gluons, all of which have zero rest mass and no electric charge. It has been suggested that gravity might be due to an exchange of **gravitons**, but the existence of such particles has yet to be experimentally verified.[2]

Every quark can have one out of three different *colors* and to every elementary particle in the standard model there is an associated **antiparticle** with the same rest mass but

[2]However, the existence of *gravitational waves* has been indirectly confirmed experimentally. In 1974 Hulse and Taylor discovered the first *binary pulsar*. It is a double star system consisting of two neutron stars of equal mass, which are rotating about the system's center of mass. Measurements show that the orbital period of these stars is decreasing by about 75 μs per year, which suggests that the system is loosing energy through emission of gravitational waves. The data deviates less than 0.5% from the predictions of general relativity[27].

opposite electric charge. A given particle and its antiparticle also differ in terms of other quantum numbers. An antiquark for instance has the *anti-color* of its corresponding quark. The antiparticle of an arbitrary quark q is denoted by $\bar{q}$ and the antiparticle of a hadron is obtained by replacing each quark in the hadron with its corresponding antiquark. In general, the antiparticle of a given particle P is denoted by $\bar{P}$, but for some common elementary particles the bar is omitted. The antiparticle of the electron is for instance called the **positron** and is denoted by e^+.

The reason why force carriers are called *gauge* bosons is because the three forces in particle physics possess gauge invariance.

8.3 Feynman Diagrams

Feynman diagrams are powerful tools in particle physics which allow one to calculate the quantum mechanical probabilities for events to occur. We will use them in a more modest fashion, mainly to depict the exchange of force carriers between two elementary particles.

The Feynman diagrams below illustrate the exchange of a single virtual photon γ between two electrons e^- and two positrons e^+, respectively.

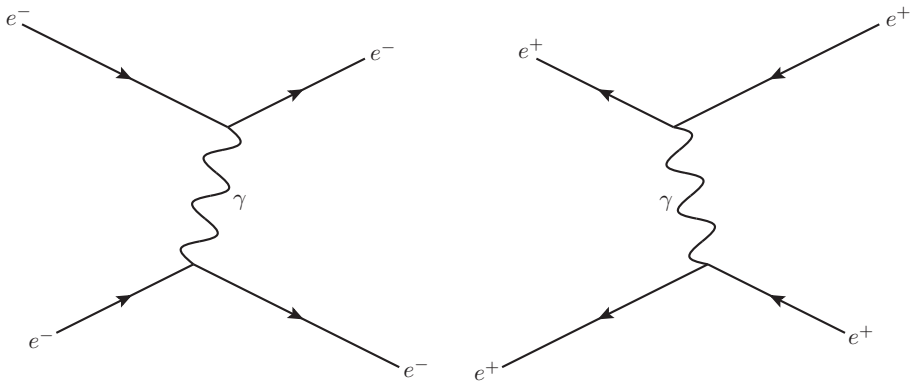


Figure 8.2: Electromagnetic interaction via photon exchange for two electrons and two positrons, respectively.

Note that time increases from the left to right and that the relative distance between the interacting particles is shown on the vertical axis in these diagrams (some authors let the horizontal axis be the spatial axis and the vertical axis be the temporal axis). Photons and W^+, W^- and Z^0 bosons are drawn as wiggly lines while leptons and quarks are drawn as straight lines and gluons as coiled lines. In order to distinguish a particle from its antiparticle one adds to the particle's line an arrowhead pointing to the right and to the antiparticle's line an arrowhead pointing to the left. It is important to remember that these arrowhead *do not* indicate the direction of the particle's motion.

The weak interaction and the strong interaction are illustrated in the two diagrams below, where ν_e, q and g denote an *electron neutrino*, a quark and a gluon, respectively.

Feynman diagrams can also be used to depict processes involving hadrons. The left diagram below shows the decay of a neutron n into a proton p and a W^- hadron, followed by the decay of the W^- particle into an electron e^- and an antielectron neutrino $\bar{\nu}_e$. The right diagram illustrates the exchange of a pion π^- between a proton and a neutron, where

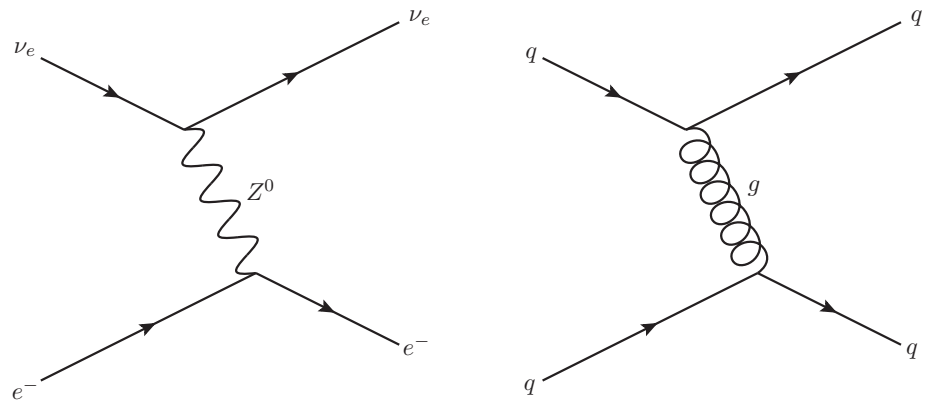


Figure 8.3: Weak interaction between an electron and an electron neutrino and strong interaction between two quarks.

the pion is drawn as a dashed line.

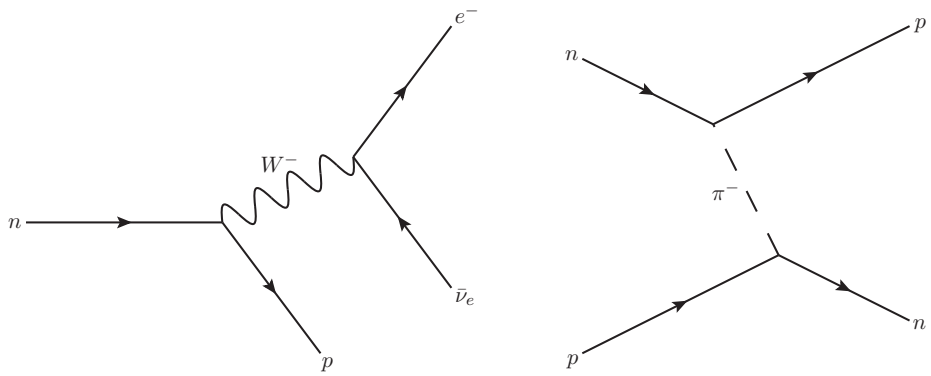


Figure 8.4: The decay of a neutron via a W^- particle and a pion exchange between a neutron and a proton.

We will study the latter process in great detail in the following sections since it is a major contribution to the **strong nuclear force** which holds atomic nuclei together. Note that this force is not the same as the fundamental strong interaction between quarks, but rather a consequence of it.

Moreover, we should mention in this context that the most basic hadrons are called **baryons** and **mesons**. While baryons are composed of three quarks mesons are composed of one quark and its corresponding antiquark. **Pions** (*pion* is an abbreviation of **pi meson**) are mesons and the most common unstable hadrons. They can have tree different electric charges and are consequently denoted by π^+, π^0 and π^-.

Finally, note that there are other conventions for drawing Feynman diagrams than the ones which we have described here. For instance, many authors draw exchange particles as vertical (or horizontal) lines, even though this erroneously implies that the particles are moving with infinite speed. It is also common to denote both attractive and repulsive interactions in the same way, even though the former type of interactions tends to decrease the distance between the interacting particles while the latter type tends to increases it.

Table 8.1: Properties of leptons. All leptons have spin 1/2. The antiparticles have the same rest masses m_0 as their corresponding leptons but the opposite electric charges Q and lepton quantum numbers L_l, where $l = e, \mu, \tau$. The data in this table comes from p. 72 in [28].

Leptons

Name	Symbol	Q [e]	L_e	L_μ	L_τ	m_0 [MeV/c^2]
Electron	e^-	-1	1	0	0	0.511
Electron neutrino	ν_e	0	1	0	0	$< 2(10^{-6})$
Muon (mu)	μ^-	-1	0	1	0	105.7
Muon neutrino	ν_μ	0	0	1	0	< 0.19
Tauon (tau)	τ^-	-1	0	0	1	1770.0
Tauon neutrino	ν_τ	0	0	0	1	< 18.2

Table 8.2: Properties of gauge bosons. All gauge bosons have spin 1 and their corresponding antiparticles have the same rest masses m_0 but the opposite electric charges Q. The data in this table comes from [29].

Gauge bosons

Force	Name	Symbol	Q [e]	m_0 [GeV/c^2]
Electromagnetism	Photon	γ	0	0
Strong	Gluon	g	0	0 (theretical)
Weak	–	W^+, W^-, Z^0	$+, -, 0$	$80.4, 91.2$

8.4 Electrons and Photons

Let us now begin to explore the properties of the elementary particles in the standard model in more detail, starting with the electron and the photon.

The electron is denoted by e^- and carries an electric charge of $-e$, where e is the elementary (or unit) charge. It has a rest mass of about $0.511\,\mathrm{MeV/c^2}$ and spin 1/2. The properties of the electron as well as other leptons are displayed in table 8.1.

As we mentioned above the photon is the mediator of the electromagnetic interaction, which only affects charged particles. It is denoted by γ, has no electric charge, zero rest mass and spin 1. The properties of the photon as well as other gauge bosons are shown in table 8.2. Also note that the photon moves with the speed of light through free space in every inertial frame.

8.5 Isospin

Atomic nuclei are composed of neutrons and protons. Since protons repel one another through electromagnetic interaction there must be an even stronger force holding a nucleus with more than one proton together. This *strong nuclear force* must affect neutrons as well as protons since the former particles are also bound to the nucleus. Moreover, it is a well known fact

that protons are repelled by a nucleus at distances greater than a few fm, which means that the strong nuclear force must decrease with distance much faster that the Coulomb force.

After the discovery of the neutron in 1932 physicists were struggling to understand how the neutron and the proton can have such similar rest masses but opposite electric charges. Heisenberg suggested that the two particles might in fact be two different states of a single particle, which is now called a **nucleon**. The quantum number which distinguishes the neutron from the proton was eventually named the **isospin**[3] **projection number** (in analogy with the ordinary spin projection number). Emperical data revealed that the strong nuclear force affects neutrons and protons equally much, which suggests that this force is independent of the isospin projection number. Moreover, in a paper published in 1936 by Cassen and Condon the authors suggest that the Hilbert space of a nucleon is acted on by the elements of the underlying set of a representation of the group $SU(2)$ in such a way that the probability of detecting a nucleon as for instance a neutron is altered.

We now know that isospin is not an exact symmetry of nature, but the *unification* of the neutron and the proton into a single nucleon was a predecessor of the modern theories of unification. Let us therefore explain how these particles were united.

The quantum mechanical state vectors of a physical system are elements of a complex Hilbert space $\mathcal{H}$ with either a finite or infinite number of dimensions (see appendix C for more details).

The Hilbert space $\mathcal{H}$ of a system composed of n distinct particles with the respective Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n$ is isomorphic to the tensor product of these spaces:

$$\mathcal{H} \cong \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \cdots \otimes \mathcal{H}_n. \tag{8.5.1}$$

If the system consists of a single particle with states in several distinct Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_n$ (for instance a particle which is confined to several distinct intervals on the x-axis) then the total Hilbert space $\mathcal{H}$ of the particle is isomorphic to the direct sum of these spaces:

$$\mathcal{H} \cong \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \cdots \oplus \mathcal{H}_n. \tag{8.5.2}$$

(See section 3.2 for the definitions of the tensor product and direct sum of vector spaces.)

This is how we construct Hilbert spaces in general for any system consisting of several parts and/or with states in distinct Hilbert spaces.

Since a given nucleon can be detected as either a neutron or a proton its state vectors must lie in both the Hilbert space $\mathcal{H}_n$ of a neutron and the Hilbert space $\mathcal{H}_p$ of a proton. In other words: the Hilbert space $\mathcal{H}$ of a nucleon satisfies

$$\mathcal{H} \cong \mathcal{H}_n \oplus \mathcal{H}_p. \tag{8.5.3}$$

If the neutron and the proton have no degrees of freedom then their respective Hilbert spaces contain only a single ray, which means that $\mathcal{H}_n = \mathbb{C} = \mathcal{H}_p$ and $\mathcal{H} = \mathbb{C}^2 \cong \mathbb{C} \oplus \mathbb{C}$. As normalized base vectors for $\mathcal{H}$ we may therefore select $\psi_n = (1, 0)$ and $\psi_p = (0, 1)$, which we define as state vectors for the "neutron state" and the "proton state", respectively.

The normalized state vectors of the nucleon compose the "unit circle" $C = \{c\psi_n + d\psi_p;\ |c|^2 + |d|^2 = 1,\ c, d \in \mathbb{C}\}$ in $\mathbb{C}^2$. If the state vector $c\psi_n + d\psi_p$ of the nucleon is

[3]The original name was *isotopic spin*, but isotopes differ in the number of nucleons while a change of a nucleus' total isospin has no effect on the number of nucleons. It was therefore suggested that the name should be changed to *isobaric spin*, but the modern term is the abbreviation *isospin*.

normalized then the probability of detecting it as a neutron or as a proton is given by $|c|^2$ and $|d|^2$, respectively.

In order to change the probability of detecting the nucleon as for instance a neutron one needs a unitary operator $\hat{U} : C \to C$. In other words, the Hilbert space $\mathbb{C}^2$ of the nucleon is the module of a representation of the group $SU(2)$. Note the analogy with the rotation operators acting on the Hilbert space of a spin 1/2 system (see section 4.3.1). In spinor notation these operators together with the Hilbert space $\mathbb{C}^2$ also form a representation (the defining representation) of $SU(2)$. Moreover, the quantum projection number m_s for such a system can only take on two values, namely $-1/2$ and $1/2$. Since a nucleon can have precisely two different *charges* (± 1) it is therefore natural to define the neutron as the **isospin down** state and the proton as the **isospin up** state.

In general, a set of N hadrons with similar masses but different electric charges is assigned an **isospin** vector $\boldsymbol{I}$. Just as angular momentum is a vector in $\mathbb{R}^3$, the isospin vector $\boldsymbol{I}$ may be thought of as a vector in an *abstract* real 3-dimensional vector space V. The projection of $\boldsymbol{I}$ on an imaginary "third" axis in V is given by the isospin projection number $I_3 = -I, -I+1, \ldots, I$ and the value of I is determined by

$$N = 2I + 1 \Leftrightarrow I = \frac{N-1}{2}. \tag{8.5.4}$$

Moreover, the values of I_3 for the N states of the particle are assigned so that the electric charge becomes a strictly increasing function of I_3. Also note that the isospins $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ of two different "unified" particles may be added, $\boldsymbol{I} = \boldsymbol{I}_1 + \boldsymbol{I}_2$, and that the total isospin quantum number I satisfies $|I_1 - I_2| \leq I \leq I_1 + I_2$.

For a nucleon we have that $N = 2$ and $I = 1/2$. Since the neutron has a lower charge than the proton it therefore corresponds to $I_3 = -1/2$ (isospin down), while the proton corresponds to $I_3 = +1/2$ (isospin up).

The strong nuclear force depends on the value of I but not on the value of I_3 (i.e. not on the orientation of $\boldsymbol{I}$). This means that the strong nuclear force is invariant under the action of $SU(2)$ since the elements of a representation of this group simply rotates the isospin vector $\boldsymbol{I}$. To be a bit more precise, let $\Pi(SU(2))$ be a representation of $SU(2)$ with $\mathbb{C}^2$ as the module. Moreover, let $\hat{F}$ be an operator of the strong nuclear force acting on $\mathcal{H}_\mathrm{f} \otimes \mathbb{C}^2$, where $\mathcal{H}_\mathrm{f}$ is the Hilbert space of the exchange particle (force carrier) for the strong nuclear force. Then the invariance of the strong nuclear force under the action of $\Pi(SU(2))$ means that

$$\hat{F}\rho(g)\psi = \rho(g)\hat{F}\psi, \, \forall \psi \in \mathcal{H}_\mathrm{f} \otimes \mathbb{C}^2 \text{ and } \forall \rho(g) \in \Pi(SU(2)), \tag{8.5.5}$$

where g is any element in $SU(2)$.

In other words, the operator $\hat{F}$ commutes with every element $\rho(g) \in \Pi(SU(2))$. Generally, if $\mathcal{G}$ is the underlying set of a group $(\mathcal{G}, \star)$ and the elements of $\mathcal{G}$ act as unitary operators on two finite dimensional Hilbert spaces $\mathcal{V}$ and $\mathcal{W}$ then a linear operator $\hat{F} : \mathcal{V} \to \mathcal{W}$ is called an **intertwining operator** if

$$\hat{F}g\psi = g\hat{F}\psi, \, \forall g \in \mathcal{G} \text{ and } \forall \psi \in \mathcal{V}. \tag{8.5.6}$$

Moreover, let $-i\tau$ be the operator for the *total* isospin of the nucleon and the exchange particle. It can be shown that τ is an element of the underlying set of a representation of $\mathfrak{su}(2)$ ([30], p. 9), which means that $\exp(\tau) \in \Pi(SU(2))$. Equation (8.5.5) therefore gives

that for any $\psi \in \mathcal{H}_{\mathrm{f}} \otimes \mathbb{C}^2$

$$\begin{aligned} \hat{F}\exp(\tau)\psi = \exp(\tau)\hat{F}\psi \Leftrightarrow \\ \hat{F}\psi + \hat{F}\tau\psi + \ldots = \hat{F}\psi + \tau\hat{F}\psi + \ldots, \end{aligned} \tag{8.5.7}$$

where we have used the linearity of $\hat{F}$.

Equation (8.5.7) gives that $\hat{F}(-i\tau)\psi = (-i\tau)\hat{F}\psi$ for all $\psi \in \mathcal{H}_{\mathrm{f}} \otimes \mathbb{C}^2$. Specifically, this means that if ψ is an eigenvector of $-i\tau$ corresponding to the eigenvalue I_3 then the same is true for $\hat{F}\psi$ since

$$-i\tau\hat{F}\psi = \hat{F}(-i\tau)\psi = I_3\hat{F}\psi. \tag{8.5.8}$$

In other words, the operators for the strong nuclear force preserve the total isospin of any physical system.

One may wonder what kind of physical process corresponds to the action of $\Pi(SU(2))$ on the Hilbert space of a nucleon? The answer was given by Yukawa in 1934 when he suggested that the strong nuclear force is due to an exchange of particles of a previously unknown type (later named *mesons*). Yukawa predicted that these new particles should have spin 0, a rest mass of about 200 times that of the electron's and electric charges of $\pm e$[31]. In his own words:

> "In order to obtain exchange forces, we must assume that these mesons have the electric charge $+e$ or $-e$, and that a positive (negative) meson is emitted (absorbed) when the nucleon jumps from the proton state to the neutron state, whereas a negative (positive) meson is emitted when the nucleon jumps from the neutron to the proton. Thus a neutron and a proton can interact with each other by exchanging mesons just as two charged particles interact by exchanging photons." (From Yukawa's Nobel lecture given in 1949.)

There were no particle accelerators in 1934 capable of generating the required energy, but physicist believed that mesons could be observed in cosmic rays. In the early 1930s Anderson and Neddermeyer observed tracks in their cloud chamber which they initially believed to be the tracks of ultra high energy electrons subjected to new laws of nature. By 1936 they had proven that the tracks had in fact been made by a previously unknown particle with a rest mass somewhere between that of the electron's and the proton's. They therefore named it the **mesotron** ("middle particle") and for several years the mesotron was mistaken for Yukawa's meson ([32], p. 71). Their new particle eventually turned out to be a lepton which is now called a **muon** or *mu* (when it had been confirmed that the muon was indeed a new particle Rabi supposedly cried out "Who ordered that?").

Yukawa's *pions* π^+ and π^- were discovered at last in 1947 by Cecil Powell and Guiseppe Occhialini. Their masses turned out to be approximately $139.6\,\mathrm{MeV/c^2}$. In addition to these two pions a neutral pion π^0 with a mass of about $135.0\,\mathrm{MeV}$ was discovered in 1950 by Bjorklund et el. in the gamma emissions following the bombardment of various targets with high energy protons ([33], pp. 370, 371). The transformations of a nucleon in connection with pion absorption/emission are illustrated in figure 8.5.

Since the total isospin of the systems are conserved quantities we can deduce from these interactions that the isospin projection number I_3 of the pions π^-, π^0 and π^+ are $-1, 0$ and $+1$, respectively, which is in full agreement with equation (8.5.4) ($(N = 3) \Rightarrow (I = 1) \Rightarrow (I_3 = -1, 0, 1)$). Note that the value of I_3 is the same for a pion as its electric charge Q expressed

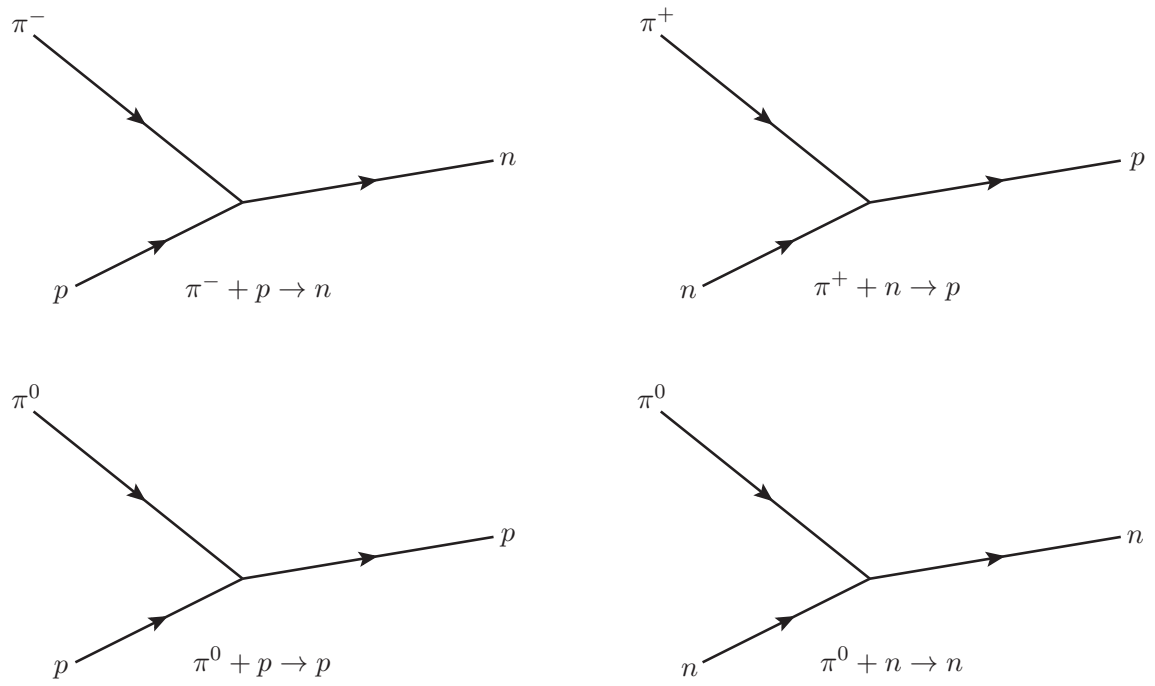


Figure 8.5: Nucleon absorption of pions.

in e, while for a neutron and a proton $I_3 = Q - 1/2$. In general, the relation between I_3 and Q is given by the **Gell-Mann–Nishijima** formula

$$Q = I_3 + Y/2, \tag{8.5.9}$$

where Q is the *hypercharge* of a unified particle. The hypercharge Y depends on the representation of $SU(2)$. For a pion $Y = 0$ and for a nucleon $Y = 1$.

Let us now see how the physical action of a pion on a nucleon corresponds to the action of $\Pi(SU(2))$ on the Hilbert space $\mathcal{H} = \mathbb{C}^2$ of the nucleon. We know that $\mathbb{C}^2$ is a complex 2-dimensional irreducible representation of $SU(2)$ (where we have chosen to identify the representation with its module). In the same way, the Hilbert space of the pions is a complex 3-dimensional irreducible representation of $SU(2)$.

With respect to isomorphism there is only one complex $(n+1)$-dimensional irreducible representation of $SU(2)$, which is $Sym^n\mathbb{C}^2$ (symmetric tensors of rank n defined on $(\mathbb{C}^2)^{\otimes n}$). Physicists call it the spin$(n/2)$ representation of $SU(2)$, but in the present context we may call it the **isospin**$(n/2)$ representation. Note that this is *not* the group $Spin(m,n)$ which is the double cover of $SO(m,n)$ (see section 6.5).

So, the Hilbert space of a nucleon and a pion are realizations of the isospin$(1/2)$ and isospin(1) representation of $SU(2)$, respectively. Moreover, $\mathrm{Sym}^2\mathbb{C}^2$ is the complexification of $\mathfrak{su}(2)$, i.e.

$$\mathrm{Sym}^2\mathbb{C}^2 \cong \mathfrak{su}(2) \otimes \mathbb{C} \cong \mathfrak{sl}(2,\mathbb{C}). \tag{8.5.10}$$

The pions thus span the realization of a representation of $\mathfrak{sl}(2,\mathbb{C})$. Since the pions are the force carriers for the strong nuclear force, we can conclude that the Hilbert space $\mathcal{H}_\mathrm{f}$ mentioned above is in fact a realization of $\mathfrak{sl}(2,\mathbb{C})$.

Table 8.3: Properties of quarks. All quarks have spin 1/2 and their corresponding antiparticles have the same rest masses m_0 but the opposite electric charges Q. The data in this table comes from p. 90 in [28].

Quarks

Name	Symbol	Q [e]	m_0 [GeV/c^2]
down	d	$-1/3$	≈ 0.3
up	u	$2/3$	≈ 0.3
strange	s	$-1/3$	≈ 0.5
charmed	c	$2/3$	≈ 1.5
bottom	b	$-1/3$	≈ 4.5
top	t	$2/3$	≈ 171

8.6 Quarks

We have seen how a neutron and a proton can be unified into a single nucleon and how the strong force between two nucleons is mediated by an exchange of pions. This is a good approximation, but it is not entirely true since it is based on the idea that neutrons and protons are fundamental particles.

We know today that they are in fact composed of quarks. There are six distinct types, or **flavors**, of quarks (see table 8.3) and an equal number of antiquarks. The neutron is made of one up quark and two down quarks, which we write as $n = udd$. Similarly, the proton is made of two up quarks and one down quark, i.e. $p = uud$. Moreover, the pions have the compositions $\pi^+ = u\bar{d}$, $\pi^0 = u\bar{u}, d\bar{d}$, $\pi^- = d\bar{u}$ (π^0 is a linear combination of $u\bar{u}$ and $d\bar{d}$, i.e. its state vector is a linear combination of the $u\bar{u}$ and $d\bar{d}$ states).

The strong and electromagnetic interactions preserve the six **quark numbers** N_f given by

$$N_f \stackrel{def}{=} N(f) - N(\bar{f}),\ f = d, u, s, c, b, t, \tag{8.6.1}$$

where $N(f)$ is the number of quarks of flavor f and $N(\bar{f})$ the number of antiquarks of flavor $\bar{f}$ within an isolated system.

However, the weak interaction only preserves the **total quark number** N_q defined by

$$N_q \stackrel{def}{=} N(q) - N(\bar{q}), \tag{8.6.2}$$

where $N(q)$ and $N(\bar{q})$ are the total number of quarks and antiquarks, respectively, of arbitrary flavors.

It is therefore convenient to define a **baryon number** B as

$$B \stackrel{def}{=} N_q/3 = [N(q) - N(\bar{q})]/3 \tag{8.6.3}$$

which is preserved by all of forces in the standard model.

There are also quantum numbers called **strangeness** (S), **charm** (C), **bottom** ($\tilde{B}$) and **top** (T), which are defined as

$$S \stackrel{def}{=} -N_s,\ C \stackrel{def}{=} -N_c,\ \tilde{B} \stackrel{def}{=} -N_s,\ T \stackrel{def}{=} -N_t. \tag{8.6.4}$$

Quarks participate in the strong interaction which is *flavor independent*. This means that at a given distance between two quarks the strong force between them does not depend on their flavors. The flavor independence of the strong force is the reason why the strong nuclear force affects neutrons and protons equally much. It is also the reason why we can group hadrons into sets of particles with approximately the same masses and consider the particles within each set as different states of a single particle corresponding to different isospin projection numbers. Such sets are called **charge multiplets**. The particles within a charge multiplet have the same baryon number, strangeness, charm and bottom but different electric charges. Examples of charge multiplets are of course (p, n) and (π^+, π^0, π^-).

Moreover, the **hypercharge** Y of a hadron is defined by

$$Y \stackrel{def}{=} B + S + C + \bar{B} + T. \tag{8.6.5}$$

A set of hadrons with the same spin, parity and baryon number is called a **supermultiplet**. Meson supermultiplets have nine constituent particles and are therefore called *nonets*, while baryon supermultiplets have one, eight or ten constituent particles and are called *singlets*, *octets* and *decuplets*. If one considers mason states with spin-parity 0^- (quarks and anti-quarks have opposite parities) composed of u, d and s quarks then one finds the multiplets $(K^0, K^+), (\bar{K}^0, K^-)$ and (π^+, π^0, π^-) and two neutral particles which are denoted by η and η'. Together these particles form a nonet which can be shown on a plot of Y as a function of I_3 (see figure 8.6). A baryon octet is shown in figure 8.7.

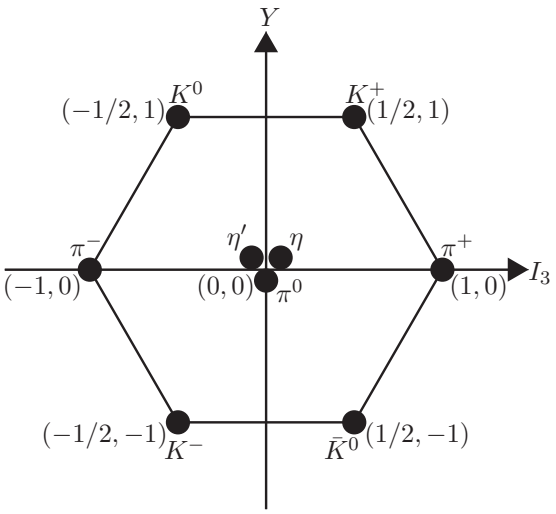


Figure 8.6: A meson supermultiplet.

The quark model has been very successful, but at first glance it does raise some interesting questions. For instance, why are all the observed *basic* hadrons either baryons $(qqq, \bar{q}\bar{q}\bar{q})$ or mesons $(q\bar{q})$? Also, the quarks in for instance the baryon $\Omega^- = sss$ are in the same spatial and spin state, which seems to violate the Pauli principle. So how is this possible?

In order to answer these questions one has to assume that quarks have another degree of freedom called **color**. Any quark can be in one out of three different color states called **red** (r), **green** (g) and **blue** (b), and an antiquark can be in one out of the three anti-color states $\bar{r}, \bar{g}$ and $\bar{b}$. These states are characterized by two quantum numbers called **color charges**,

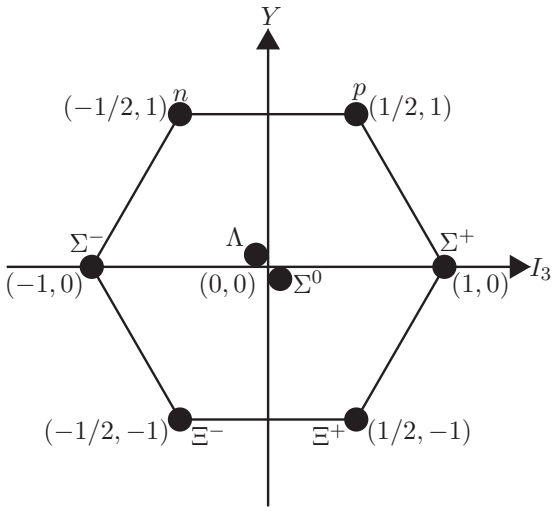


Figure 8.7: A baryon supermultiplet.

Table 8.4: Color charges for different color states for quarks and antiquarks.

Quarks	I_3^{C}	Y^{C}	Antiquarks	I_3^{C}	Y^{C}
r	$1/2$	$1/3$	$\bar{r}$	$-1/2$	$-1/3$
g	$-1/2$	$1/3$	$\bar{g}$	$1/2$	$-1/3$
b	0	$-2/3$	$\bar{b}$	0	$2/3$

namely **color isospin** I_3^{C} and **color hypercharge** Y^{C}. These quantum numbers depend only on the color states (r, g, b) and not on the flavors $(d, u, s \,\text{etc.})$. Also note that I_3^{C} and Y^{C} are both preserved under the strong interaction. The values of the color charges for different color states are shown in table 8.4. These values are the result of a symmetry for the strong force called $SU(3)$ *color symmetry*.

For multiparticle states the total color charges are the sums of the color charges of the individual states. The **color confinement** hypothesis states that sets of quarks can only be observed as free particles if their total color charges are zero. The corresponding multiparticle states are called *color singlets* (we will see in the next section that this is not the actual definition of a color singlet). Color confinement implies that baryons have the color combination rgb and that mesons have the color combinations $r\bar{r}, g\bar{g}$ and $b\bar{b}$. Since a combination of red, green and blue gives the color white one say that hadrons have to be "white" or "colorless" (one can imagine that an anti-color cancels out a color).

8.7 The Eightfold Way

With our new knowledge of particle physics we are ready to take the next step by approaching it from a group theoretical point of view. The aim is to find a physical interpretation of the root and weight diagrams from chapter 5 which we already have established correspond to irreducible representations. These irreducible representation can in turn be used to classify

particles depending on how they transform. For external symmetries we saw that particles fall into massive or massless irreducible representation of the Poincaré group. Figure 8.7 has already hinted what the root diagram from chapter 5 can be interpreted as.

The **Eightfold Way** is a theory to organize our mesons and baryons into irreducible representations of the symmetry group of which our mesons and baryons transform under. Since mesons and baryons are bound states of quarks, we begin by finding the irreducible representations for the three lightest quarks u, d and s. Gell-Mann postulated that the three quarks form a basis for $\mathbb{C}^3$ where each quark is given by

$$u = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad d = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad s = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{8.7.1}$$

The antiquarks form a basis for the dual space $(\mathbb{C}^3)^*$ where each antiquark is given by

$$\bar{u} = \begin{pmatrix} 1,0,0 \end{pmatrix}, \quad \bar{d} = \begin{pmatrix} 0,1,0 \end{pmatrix}, \quad \bar{s} = \begin{pmatrix} 0,0,1 \end{pmatrix}. \tag{8.7.2}$$

It turns out that (8.7.1) are the basis for the module of the fundamental representation for $\mathfrak{sl}(3,\mathbb{C})$. This can be verified if we consider the action of the Cartan elements $\pi(x_3) = \pi(h_1)$ and $\pi(z_8) = \pi(h_2)$, given by (5.3.5), on the three quark states u, d and s. For example,

$$\begin{aligned} \pi(h_1)u &= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \\ \pi(h_2)u &= \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \end{aligned} \tag{8.7.3}$$

$$\begin{aligned} \pi(h_1)d &= \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = -\frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\ \pi(h_2)d &= \frac{1}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}. \end{aligned}$$

That is, u and d are two eigenstates to the Cartan elements $\pi(h_1)$ and $\pi(h_2)$ with eigenvalues $(1/2, 1/3)$ and $(-1/2, 1/3)$, respectively. In a similar way we will find that s is the third eigenstate with eignevalues $(0, -2/3)$. But this is just the fundamental representation for $\mathfrak{sl}(3,\mathbb{C})$ presented in section 5.3 by the weight diagram 5.5. We now have the physical interpretation of the irreducible representation denoted **3** talked about in section 5.3. Thus, the three quark states u, d and s form the fundamental representation for $\mathfrak{sl}(3,\mathbb{C})$. u is the highest weight state corresponding to Λ_1. With this in mind it is not an entirely unfounded guess that the antiquarks constitute the antifundamental representation $\bar{\mathbf{3}}$.

The antifundamental representation is the complex conjugate of the fundamental representation. If we take the complex conjugate of the commutation relations for $\mathfrak{sl}(3,\mathbb{C})$ we obtain

$$[\pi(t_i)^*, \pi(t_j)^*] = -if_{ijk}\pi(t_k)^* \quad \Rightarrow [-\pi(t_i)^*, -\pi(t_j)^*] = if_{ijk}(-\pi(t_k)^*) \tag{8.7.4}$$

where the $\pi(t_{i,j,k})$ is a representation of the Lie algebra elements and f_{ijk} are the real structure constants. From (8.7.4) we see that $-\pi(h_1)^*$ and $-\pi(h_2)^*$ are elements from the antifundamental representation. It is now easy to see that the three antiquarks $\bar{u}$, $\bar{d}$ and $\bar{s}$ are eigenstates to $-\pi(h_1)^*$ and $-\pi(h_2)^*$. For example

$$\begin{aligned} -\pi(h_1)^*\bar{s} &= \frac{1}{2}\begin{pmatrix}0,0,1\end{pmatrix}\begin{pmatrix}-1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{pmatrix} = 0 \\ -\pi(h_2)^*\bar{s} &= \frac{1}{3}\begin{pmatrix}0,0,1\end{pmatrix}\begin{pmatrix}-1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 2\end{pmatrix} = \frac{2}{3}\begin{pmatrix}0,0,1\end{pmatrix}, \end{aligned} \tag{8.7.5}$$

i.e. $\bar{s}$ is an eigenstate with eigenvalues $(0, -2/3)$ known as the highest weight state for the antifundamental representation corresponding to Λ_2. The other antiquarks states $\bar{d}$ and $\bar{u}$ have eigenvalues $(1/2, -1/3)$ and $(-1/2, -1/3)$, respectively. Thus, the three aniquark states $\bar{u}$, $\bar{d}$ and $\bar{s}$ form the antifundamental representation for $\mathfrak{sl}(3,\mathbb{C})$.
What we have found so far is that the quarks and antiquarks constitute the 3-dimensional irreducible representations obtained in section 5.3. These quarks are simultaneous eigenstates to the two Cartan operators $\pi(h_1)$ and $\pi(h_2)$ which corresponds to the two physical observables; isospin projection and hypercharge mentioned in the previous section. Our quark states can thus be labeled with isospin projection I_3 and hypercharge Y. This means that the so far unknown labels x and y in figure 5.4 and 5.5 can be changed to I_3 and Y, respectively. This is illustrated in figure 8.8 and 8.9. Let us now consider the mesons. They consist of one quark and one antiquark. This compound system is thus a tensor product of a quark and an antiquark state,

$$V = \mathbb{C}^3 \otimes (\mathbb{C}^3)^*, \tag{8.7.6}$$

thus giving us a 9-dimensional representation with module V. This representation is *reducible* into two irreducible representations; one 1-dimensional and one 8-dimensional. The meson states can be obtained since we have the quark and antiquark states. For example, the π^+ meson consist of one u quark and one $\bar{d}$ antiquark,

$$u \otimes \bar{d} = \begin{pmatrix}1\\0\\0\end{pmatrix} \otimes \begin{pmatrix}0,1,0\end{pmatrix} = \begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}. \tag{8.7.7}$$

This is the $\pi(e_1)$ element from the adjoint representation. The π^- is the antiparticle to π^+, i.e. consists of one d quark and one $\bar{u}$ antiquark

$$d \otimes \bar{u} = \begin{pmatrix}0\\1\\0\end{pmatrix} \otimes \begin{pmatrix}1,0,0\end{pmatrix} = \begin{pmatrix}0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix} \tag{8.7.8}$$

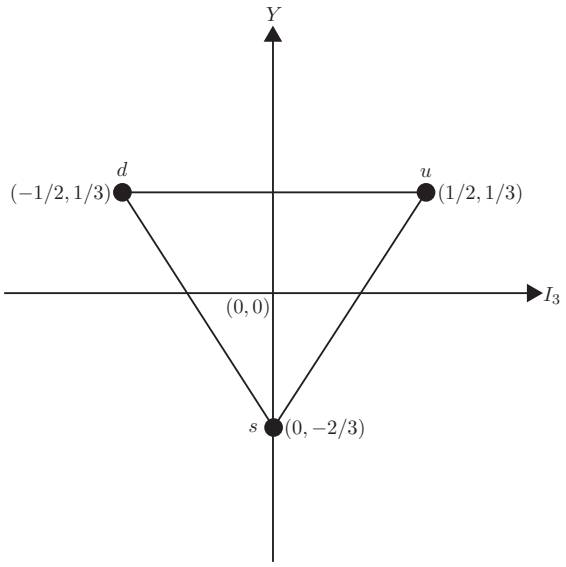


Figure 8.8: The three lightest quarks u, d and s in the fundamental representation for $\mathfrak{sl}(3,\mathbb{C})$.

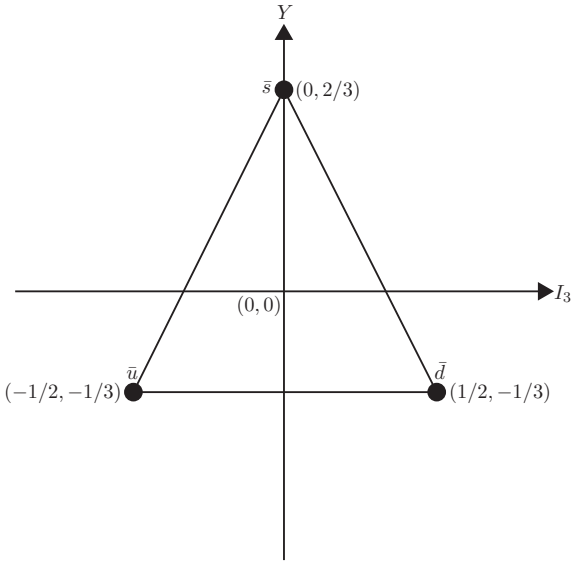


Figure 8.9: The three lightest antiquarks $\bar{u}$, $\bar{d}$ and $\bar{s}$ in the antifundamental representation for $\mathfrak{sl}(3,\mathbb{C})$.

which is the $\pi(f_1)$ element of the adjoint representation. The operation that has been made is called *outer product*, see appendix B.3. If we do the same thing for the kaon particles K^0, K^+, $\bar{K}^0$ and $\bar{K}^+$ we will discover that these are the other six step operators in the adjoint representation. The η, η' and the π^0 mesons are the remaining elements in the 9-dimensional representation whereof η and the π^0 are the two Cartan elements in the adjoint representation. η' is thus in the 1-dimensional trivial representation. This 9-dimensional representation is recognized as the nonet mentioned previously, see figure 8.10.

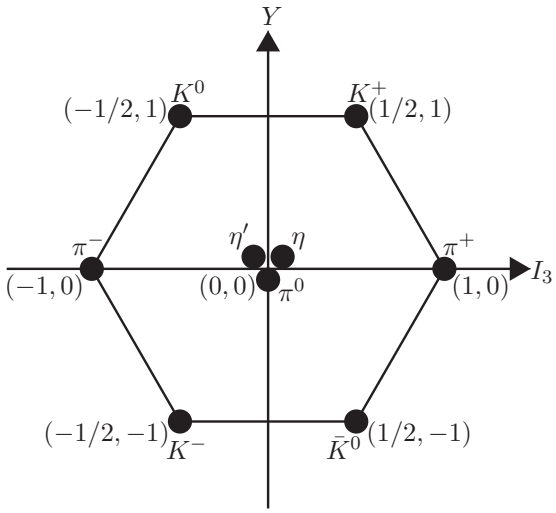


Figure 8.10: The 9-dimensional reducible representation for $\mathfrak{sl}(3,\mathbb{C})$.

In summary, the eight mesons π^+, π^0, π^-, K^0, K^+, $\bar{K}^0$, $\bar{K}^+$ and η consitute the adjoint representation for $\mathfrak{sl}(3,\mathbb{C})$, the *Eightfold Way*. The Eightfold Way is illustrated in figure 8.11. This figure shows the direct sum of the Eightfold Way and the trivial representation, the nonet.

To every root α in 5.2 we associate a meson. Since the mesons sit in the adjoint representation, the weights are the roots. The roots are the eigenvalues to the two observables isospin projection and hypercharge, e.g. π^+ has isospin projection $1/2$ and hypercharge 0. The eight mesons that we have talked about so far are spin 0-particles but the Eightfold Way is equally applicable to the spin 1 mesons.

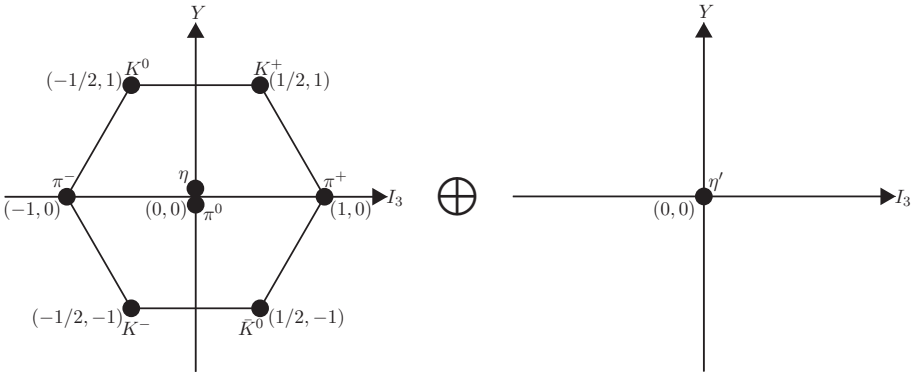


Figure 8.11: The direct sum of the 8-dimensional adjoint representation and the 1-dimensional trivial representation for $\mathfrak{sl}(3,\mathbb{C})$. The result of this sum is the nonet seen in figure 8.10.

Baryons consist of three quarks. This means that the representation of which these particles fall into is 27-dimensional, which is reducible into irreducible representations. Figure 8.6 shows one of the irreducible representations; the 8-dimensional.

In section 4.4 we used the complexified Lie algebra $\mathfrak{sl}(2,\mathbb{C})$ to obtain representations

for $\mathfrak{su}(2)$. The same thing can be done for $SU(3)$, i.e. we use $SL(3,\mathbb{C})$ representations as representations for $SU(3)$. The $SU(3)$ symmetry that we have talked about so far is called the **flavor** $SU(3)$ **symmetry** which means that the physics is invariant under the change of two quarks[4]. This symmetry is not fully met due to the small differences in mass and charge between the quarks which makes flavor $SU(3)$ an *approximate symmetry*.
We mentioned **color** $SU(3)$ **symmetry** in the previous section as the symmetry connected to the color and the strong force. That is, the strong force is invariant under the change of color. This is an *exact symmetry* in contrast to the flavor symmetry. We must be careful to distinguish between *flavor* $SU(3)$ and *color* $SU(3)$. It is the same group but corresponds to different symmetries. From tabel 8.4 and the discussion at the end of the previous section we know that quarks can have 3 different colors. Gluons, on the other hand, can have eight different colors. The three colors r, g and b form the fundamental representation and the anticolors $\bar{r}$, $\bar{g}$ and $\bar{b}$ form the antifundamental representation in complete analogy with the quarks, i.e.

$$r = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad g = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{8.7.9}$$

We can reuse figure 8.8 and 8.9 by just replacing the quarks to colors according to (8.7.9) and change the axis to I_3^C and Y^C. Both mesons and gluons must carry a color and an anticolor charge thus taking the tensor product again will give us the color states for mesons and gluons. Just like before, this 9-dimensional representation is reducible into a 1-dimension and an 8-dimensional representation. In the 8-dimensional representation, there are six different color states, corresponding to the ladder operators, with non-zero color charges. These have to be the gluons' color states since mesons have to be color singlets. The other two are colorless states corresponding to the Cartan elements, but still in the 8-dimensional adjoint representation. So, it is not really sufficient to say that a color singlet is a state with zero color charges. It has to transform trivially, i.e invariant under $SU(3)$. Thus, the color state of meson fall into the 1-dimensional representation and the color states for gluons into the 8-dimensional adjoint representation. Since the color part Ψ_{color} of the wavefuncion is invariant, its matrix representation must be proportional to the unit matrix,

$$\Psi_{color} = \frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} + b\bar{b}). \tag{8.7.10}$$

This is the color-wave function that all the mesons are in. The other color-wave functions are given by $r\bar{b}$, $r\bar{g}$, $g\bar{r}$, $g\bar{b}$, $b\bar{r}$, $b\bar{g}$, $\frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g})$ and $\frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} - 2b\bar{b})$. We obtain the last two by considering a particular linear combination of $r\bar{r}$, $g\bar{g}$ and $b\bar{b}$ that will give us the Cartan elements.

8.8 Concluding Remarks and Future Prospects

As we have discovered in this chapter it is possible to classify particles according to flavor and color invariance with the aid of gauge symmetries. We have also examined approximate symmetries such as isospin for the neutron and proton as well as the Eightfold Way of mesons.

[4]The flavor independence of the strong force that we talked about in the previous section is an example of this.

Hence we may now take the leap and appreciate the implications of the full gauge symmetry group

$$\mathcal{G}_{SM} = U(1) \times SU(2) \times SU(3) \tag{8.8.1}$$

of the Standard Model of particle physics, where we encounter the gauge group $SU(3)$ associated with color and $U(1) \times SU(2)$ which describes the electroweak interaction which we encountered in chapter 7. Combined with the previous remarks about the irreducible representations of the Poincaré group and the Coleman-Mandula theorem we are able to classify elementary particles. They are irreducible representations of the Poincaré group of spacetime symmetries and the gauge group of internal symmetries of the Standard Model. However, the story does not end here... There may be several other particles and symmetry groups out there. The interested reader is referred to discussions about grand unified theories [30] and supersymmetry [23]. A few other topics of interest for future surveys are presented below.

In 1979 the Nobel prize of physics was awarded to Sheldon Glashow, Abdus Salam and Steven Weinberg for their unification of electromagnetism and the weak interaction through the gauge group $U(1) \times SU(2)$. However, at our ordinary energy scales we do not perceive these phenomena in the same way. Beta decay in heavy atomic nuclei (associated with the weak interaction) and the magnetic force of a horseshoe magnet (electromagnetism) show few similarities at a first glance. Yet, if the energy is increased these forces are of the same kind. Moreover, electromagnetism is mediated via massless photons and the weak force by massive $W^{\pm}$- and Z^0-bosons. The weak force also has a very short range (since its gauge bosons are massive) while the electromagnetic force has infinite range. How can these two forces be the same thing? This "broken symmetry" of $U(1) \times SU(2)$ at lower energy scales can be explained with the *Higgs mechanism*, see chapter 20 in [6] and chapter 18 in [4]. Roughly one assumes that there is a non-zero scalar field of the vacuum similar to our first scalar field introduced in chapter 7, *the Higgs field*, which gives mass to the $W^{\pm}$- and Z^0-bosons and this causes a spontaneous breakdown of $U(1) \times SU(2)$ to the $U(1)$ gauge group of electromagnetism on small energy scales. On larger energy scales this mass difference is irrelevant and the forces unite. The Higgs field also gives mass to electrons and quarks mediated via the Higgs boson, which may be found experimentally at CERN in the near future. This is an excellent subject for future students in this field and might result in a few other books.

The latest Nobel prize awarded for works on symmetry in particle physics was announced in 2008 and given to Yoichiro Nambu, Makoto Kobayashi and Toshihide Maskawa as a token of recognition for their work on spontaneously broken symmetry in particle physics. Nambu proposed the color charge of QCD while Kobayashi's and Maskawa's accomplishment lies in the description of *CP-violation* which is a violation of the symmetries of charge (C) and parity (P). Naively, if you perform an experiment with antiparticles with the opposite charge of the ordinary particles and interchanges right and left you should obtain the same result. Yet, experiments during the 1960-1970's showed otherwise. Together with the physicist Nicola Cabibbo they constructed the unitary *Cabibbo-Kobayashi-Maskawa matrix* (CKM-matrix) which provides information on the strength of flavor-altering weak decays, see chapter 20 of [6]. With its aid Kobayashi and Maskawa predicted in 1973 that there should be at least three families of quarks in nature. Their foresight was experimentally verified in 1977 with the discovery of the bottom quark and the finding of the top quark in 1995. This is yet

another important example of how group theory interacts with particle physics which may serve as inspiration for future study and shows the vitality of the field today.

With these concluding remarks, our survey reaches its end. We hope that the reader has gained a valuable insight in the mathematical description of the symmetries of the fundamental laws of nature and wishes to proceed in this field of theoretical physics with scientific endeavor and an open mind.

The End

Appendix A

Some Topological Notions

For the readers who are quite unfamiliar with words such as "manifold" and "topology", here is a brief review. Initially, we will define what a topological space is as well as a manifold. Moreover we will discern the differences between compact and open manifolds. However, be aware that some of the definitions are not complete, some things are intentionally omitted in order to create a manageable text. If one is interested in a more complete picture, *Geometry, topology and physics* by M. Nakahara is an excellent source. [7]

Definition A.0.1. *A* **topological space** *is a set X together with a collection of open subsets T which adhere to the following axioms:*

(i) *The empty set $\emptyset$ is in T.*

(ii) *X is in T.*

(iii) *The intersection of a finite number of sets in T is also in T.*

(iv) *The union of an arbitrary number of sets in T is also in T.*

It is also possible to define it with a collection of closed subsets T by replacing intersection *with* union *in the axiom list above.*

Now, let us have a look at manifolds.

Definition A.0.2. *A* **manifold** *is a topological space of dimension n where every point has a neighborhood which is topologically the same as the open unit ball in $\mathbb{R}^n$. Globally it might have a different structure.*

Definition A.0.3. *A* **differentiable manifold** *is a manifold which has a globally differentiable structure, i.e. such a structure that it is possible to do differential calculus on the manifold. One can describe a manifold locally in terms of* **charts** *whose coordinates are converted to one another through* **transition maps**. *To ensure a global differentiable structure, the coordinates of each chart in the domains where charts overlap must be differentiable with respect to the coordinates defined by each chart in the* **atlas**. *The atlas describes the manifold globally and is patched with charts.*

Moreover, we need to have some idea what a *closed manifold*, *compact manifold* and an *open manifold* are.

Definition A.0.4. *A* **closed manifold** *is a manifold with a compact topological structure without boundary. A* **compact manifold** *on the other hand is defined likewise although it has a boundary, which may be empty. An* **open manifold** *is a manifold with neither boundary nor any compact component.*

Likewise, it is advised to have encountered the concept of *connectedness*.

Definition A.0.5. *A matrix Lie group $\mathcal{G}$ is* **connected** *if there for any two matrices $A_1, A_2 \in \mathcal{G}$ exists a continuous path $A(t)$ in $\mathcal{G}$ where $a \leq t \leq b$ such that $A(a) = A_1$ and $A(b) = A_2$. In topology, this is almost but not quite the same as* **path-connected**. *Moreover, a connected matrix Lie group $\mathcal{G}$ is* **simply connected** *if every loop in $\mathcal{G}$ can be contracted continuously to a point in $\mathcal{G}$. That is, $\mathcal{G}$ is simply connected if given any continuous path in $\mathcal{G}$, $A(t)$ where $a \leq t \leq b$ and $A(a) = A(b)$, there exists a continuous function $A(s,t)$ with $a \leq s, t \leq b$ which takes values in $\mathcal{G}$ such that*

(i) $A(s,a) = A(s,b)\,\forall s$

(ii) $A(a,t) = A(t)$

(iii) $A(b,t) = A(b,a)\,\forall t$

$SU(2)$ is simply connected with regard to the definition above, since we have seen that is isomorphic to the 3-sphere S^3, which can be shrunk to a smaller sphere and ultimately to a point without any discontinuity troubles.

Moreover, a Lie group although itself is a manifold can *act* on another manifold. For instance, we have observed that representations of $SO(3)$ acts on $\mathbb{R}^3$. The definition of action follows below.

Definition A.0.6. *Let $\mathcal{G}$ be a Lie group and $\mathcal{M}$ a manifold. Then the* **action** *of $\mathcal{G}$ on $\mathcal{M}$ is a differential map $\sigma : \mathcal{G} \times \mathcal{M} \to \mathcal{M}$ which satisfies*

(i) $\sigma(e, p) = p$ *for any $p \in \mathcal{M}$ where e is the identity element of $\mathcal{G}$*

(ii) $\sigma(g_1, \sigma(g_2, p)) = \sigma(g_1 g_2, p)$ *for any $p \in \mathcal{M}$ and for any $g_1, g_2 \in \mathcal{G}$.*

With this definition, we can now establish the origin of the term *little group*, which we encounter when examining the Lorentz group.

Definition A.0.7. *Suppose $\mathcal{G}$ is a Lie group and $\mathcal{M}$ is a manifold. The* **little group** *(or* **isotropy/stabilizer group***) of $p \in \mathcal{M}$ is a subgroup $\mathcal{H}(p)$ of $\mathcal{G}$ defined according to*

$$\mathcal{H}(p) = \{g \in \mathcal{G} | \sigma(g, p) = p\}\,. \tag{A.0.1}$$

It is possible to show that $\mathcal{H}(p)$ is a Lie subgroup ([7], p. 181).

Appendix B

Tensors

In this survey we have performed numerous tensor calculations. For the readers who are unfamiliar with such concepts this appendix gives a short introduction to tensors and tensor calculus.

B.1 Definitions and Properties

A tensor is a mathematical object which is associated with a specific vector space and defined by the way it transforms under a change of coordinate system for that vector space. We will soon give the formal definition of a tensor, but let us first explore some of the properties of tensors.

Tensors are denoted by "kernel" symbols such as A, b, ϵ etc. *Specific* components of a tensor are denoted by the tensor's kernel symbol adorned with numerical indices (subscripts and/or superscripts), ${A_1}^2, b_3, \epsilon_{123}$ etc., while a *general* component is denoted by algebraic indices, ${A_i}^j, b_i, \epsilon_{ijk}$ etc. Note though that in most literature a general component of a tensor is identified with the tensor itself, i.e. one speak of the *tensor* ${A_i}^j$, the *tensor* b_i etc.

The number of indices is called the **rank** of the tensor and one usually let the indices range over the values $1, 2, \ldots, N$, where N is the dimension of the associated space. However, in special relativity it is customary to let the indices range from 0 to 3 rather than from 1 to 4. In special relativity one also let lower case Latin letters (i, j, k etc.) range from 1 to 3 and lower case Greek letter (μ, ν, σ etc.) range from 0 to 3. The total number of components of a rank n tensor is N^n.

Tensors of rank zero are called *scalars* while tensors of rank one and two are called *vectors* and *matrices*, respectively (note that these are not the *definitions* of scalars, vectors and matrices).

While the horizontal positions of the indices are important for specifying the components of a tensor, the vertical positions are not. For instance, the element A_{12} of a tensor A is the same element as $A^{12}, {A_1}^2$ and ${A^1}_2$. However, the vertical positions of the indices *are* important in the sense that they decide how the components of a tensor transform under a change of coordinate system (see definition B.1.1). They are also important in connection with **Einstein's summation convention** which states that if an index appears twice in a single term, once as a subscript and once as a superscript, then a summation over the range

of the index is implied. For example, if $i = 1, 2, \ldots, N$ then

$$A_i B^i \stackrel{def}{=} \sum_{i=1}^{N} A_i B^i. \tag{B.1.1}$$

Such double indices are called **dummy indices** while single indices are called **free indices**. If a free index is in a tensor equation then it must appear precisely once in *every* term of the equation. Such an index may be replaced by another index if it is replaced in *every* term. A dummy index pair can be replaced by another dummy index pair in one or several terms of an equation, but no index of any kind may appear more than twice in a single term. Note that one may *never* replace any kind of index by a pre-existing index!

Also note that a tensor equation written in general component form, for instance $A_i{}^j = B_i{}^j$, must hold for all admissible values of the free indices. Moreover, the free indices on the left hand side of the equation must match the free indices on the right hand side of the equation in terms of the number of free indices, their designations and vertical (but not horizontal) positions. The factors of any given term in such an equation can be written in an arbitrary order if the components of the tensors involved are merely numbers. For instance, if A and B are ordinary matrices then $A_i{}^j B_{jk} = B_{jk} A_i{}^j$. This rule follows directly from the commutativity of real and complex numbers. However, if the factors are *operators* then naturally their order may only be reversed if the operators commute with each other.

Two tensors are said to be of the same **type** if they have the same number of free indices (both subscripts as well as superscripts). If two tensors of the same type have the same components in all relevant coordinate systems then they are by definition **equal**.

An **antisymmetric tensor** (or **alternating tensor**) is defined as a tensor for which every component changes sign whenever two indices trade places. For instance, a matrix A is antisymmetric iff for all admissible values of the indices i and j

$$A_{ij} = -A_{ji}. \tag{B.1.2}$$

Note that if $i = j$ then condition (B.1.2) implies that $A_{ii} = 0$ for all i. In general, an antisymmetric tensor of any rank (≥ 2) must have 0-components whenever two indices have the same value. Also note that since a tensor can be antisymmetric in for instance two indices but not all indices, a tensor which satisfies the requirement above is sometimes called a **completely** antisymmetric tensor.

Moreover, a (completely) **symmetric tensor** is analogously defined as a tensor for which every component retains the same value whenever two indices trade places.

Any given tensor A of rank 2 can be uniquely decomposed into a sum of a symmetric tensor B and an antisymmetric tensor C of the same type. For instance: $A_{ij} = B_{ij} + C_{ij}$. The condition that B should be symmetric and C antisymmetric gives the following system of coupled linear equations:

$$\begin{cases} A_{ij} = B_{ij} + C_{ij} \\ A_{ji} = B_{ji} + C_{ji} \end{cases} \Leftrightarrow \begin{cases} A_{ij} = B_{ij} + C_{ij} \\ A_{ji} = B_{ij} - C_{ij} \end{cases} \Leftrightarrow \begin{cases} B_{ij} = (A_{ij} + A_{ji})/2 \\ C_{ij} = (A_{ij} - A_{ji})/2 \end{cases} \tag{B.1.3}$$

The indices of the symmetric tensor B's components are usually written within ordinary parentheses while the indices of the antisymmetric tensor C's components are written within square brackets. So for the example above we would write: $A_{ij} = A_{(ij)} + A_{[ij]}$, where $A_{(ij)} = B_{ij}$ and $A_{[ij]} = C_{ij}$.

Note that we cannot in general write a tensor A of rank 3 or higher as a sum of a symmetric tensor B and an antisymmetric tensor C. The reason for this is that there are $n(n-1)/2$ different ways to select two free indices to switch out of n free indices. The symmetric nature of B and antisymmetric nature of C would therefore lead to a system of $n(n-1)/2+1$ coupled linear equations with only two variables, which cannot in general be solved. For instance, for a rank 3 tensor A with a general component A_{ijk} we would end up with a system of four linear equations

$$\begin{cases} A_{ijk} = B_{ijk} + C_{ijk} \\ A_{ikj} = B_{ijk} - C_{ijk} \\ A_{jik} = B_{ijk} - C_{ijk} \\ A_{kji} = B_{ijk} - C_{ijk} \end{cases} \tag{B.1.4}$$

which is only solvable if $A_{ikj} = A_{jik} = A_{kji}$.

However, one can always decompose any given tensor into a sum of a tensor which is symmetric in two arbitrary indices and a tensor which is antisymmetric in the same indices. For instance:

$$A_{abc}{}^{def} = A_{(ab)c}{}^{def} + A_{[ab]c}{}^{def}, \tag{B.1.5}$$

where

$$A_{(ab)c}{}^{def} = \frac{1}{2}\Big(A_{abc}{}^{def} + A_{bac}{}^{def}\Big) \tag{B.1.6}$$

$$A_{[ab]c}{}^{def} = \frac{1}{2}\Big(A_{abc}{}^{def} - A_{bac}{}^{def}\Big). \tag{B.1.7}$$

Moreover, primed indices are used to distinguish between the same components in different coordinate systems: $b^1, \ldots, b^N$; $b^{1'}, \ldots, b^{N'}$; $b^{1''}, \ldots, b^{N''}$ etc. Sometimes the kernel is primed instead of the indices. For example: A'_{ijk}.

In section B.4 we will see that in special relativity partial derivatives of tensors are themselves tensors. For instance, if $\boldsymbol{x} = (x^0, x^1, x^2, x^3)$ is a *four-tensor* (the definition of four-tensors will soon be given) then

$$\frac{\partial \boldsymbol{x}}{\partial x^\mu} = \left(\frac{\partial x^0}{\partial x^\mu}, \frac{\partial x^1}{\partial x^\mu}, \frac{\partial x^2}{\partial x^\mu}, \frac{\partial x^3}{\partial x^\mu}\right) \tag{B.1.8}$$

$$\frac{\partial \boldsymbol{x}}{\partial x^{\mu'}} = \left(\frac{\partial x^0}{\partial x^{\mu'}}, \frac{\partial x^1}{\partial x^{\mu'}}, \frac{\partial x^2}{\partial x^{\mu'}}, \frac{\partial x^3}{\partial x^{\mu'}}\right) \tag{B.1.9}$$

$$\frac{\partial \boldsymbol{x}'}{\partial x^\mu} = \left(\frac{\partial x^{0'}}{\partial x^\mu}, \frac{\partial x^{1'}}{\partial x^\mu}, \frac{\partial x^{2'}}{\partial x^\mu}, \frac{\partial x^{3'}}{\partial x^\mu}\right) \text{ etc.} \tag{B.1.10}$$

are also tensors, where $\boldsymbol{x}'$ is $\boldsymbol{x}$ in another inertial frame.

In special relativity we may therefore multiply two partial derivatives componentwise, for instance

$$\frac{\partial x^\mu}{\partial x^\sigma}\frac{\partial x^\sigma}{\partial x'^\mu}\,, \tag{B.1.11}$$

and then use Einstein's summation convention on the product.

However, we will also see in chapter B.4 that in general relativity, as well as in many other areas of application of tensor calculus, derivatives of tensors are *not* tensors themselves. Nevertheless, in order to simplify our formulas we will always sum over repeated indices in

partial derivatives (even when the latter are *not* tensors). For the same reason, we will also write partial derivatives in coordinate transformations as

$$p_i^{i'} \stackrel{def}{=} \frac{\partial x^{i'}}{\partial x^i}, \, p_{i'}^{i} \stackrel{def}{=} \frac{\partial x^i}{\partial x^{i'}}, \tag{B.1.12}$$

where p stands for *partial*. Note that the superscripts of p:s have been placed directly over the subscripts as a reminder of the fact that these objects might not be tensors.

Moreover, when we make a change of coordinates we will implicitly assume that the transformation is non-singular and that the functions involved are differentiable.

Let us now define a tensor. Remember that *a tensor is defined by the way it transforms under a change of coordinate system*!

Definition B.1.1. *Since an object may behave as a tensor under a certain set of coordinate transformations, but not in general, we restrict our attention to specific transformations. There are three cases:*

(i) *An object A with the components $A^{ij\ldots n}$ in the $\{x^i\}$ coordinate system and the components $A^{i'j'\ldots n'}$ in the $\{x^{i'}\}$ coordinate system is said to behave as a* **contravariant tensor** *under the transformation $\{x^i\} \to \{x^{i'}\}$ iff*

$$A^{i'j'\ldots n'} = A^{ij\ldots n} p_i^{i'} p_j^{j'} \ldots p_n^{n'}. \tag{B.1.13}$$

(ii) *Similarly, an object A with the components $A_{ij\ldots n}$ in the $\{x^i\}$ coordinate system and the components $A_{i'j'\ldots n'}$ in the $\{x^{i'}\}$ coordinate system is said to behave as a* **covariant tensor** *under the transformation $\{x^i\} \to \{x^{i'}\}$ iff*

$$A_{i'j'\ldots n'} = A_{ij\ldots n} p_{i'}^{i} p_{j'}^{j} \ldots p_{n'}^{n}. \tag{B.1.14}$$

(iii) *Finally, an object A with the components $A^{ij\ldots m}_{no\ldots q}$ in the $\{x^i\}$ coordinate system and the components $A^{i'j'\ldots m'}_{n'o'\ldots q'}$ in the $\{x^{i'}\}$ coordinate system (where the horizontal positions of the subscripts relative to the horizontal positions of the superscripts may be selected freely) is said to behave as a* **mixed tensor** *under the transformation $\{x^i\} \to \{x^{i'}\}$ iff*

$$A^{i'j'\ldots m'}_{n'o'\ldots q'} = A^{ij\ldots m}_{no\ldots q} p_i^{i'} p_j^{j'} \ldots p_m^{m'} p_{n'}^{n} p_{o'}^{o} \ldots p_{q'}^{q}. \tag{B.1.15}$$

If an object is referred to as simply a **tensor** *then it is understood that the object behaves as a tensor under* every *non-singular differentiable coordinate transformation for the associated vector space.*

Note that (i) and (ii) are both special cases of (iii) and that there are multiple summations in equation (B.1.13)–(B.1.15).

Objects which only behave as tensors under a subset of the set of all non-singular differentiable coordinate transformations for the associated vector space are called **qualified tensors**. Such tensors should be referred to by a name which indicates the subset in question. For instance, **four-tensors** (or **Lorentz-tensors**) by definition behave tensorially under Poincaré transformations ([34], p. 138).

The **main theorem of tensor calculus** states that if two tensors of the same type have the same components in *one* coordinate system then they also have the same components

in *every other* coordinate system. The proof follows directly from equation (B.1.15), which shows that tensors of the same type transform in precisely the same way.

A direct consequence of the main theorem is that if a tensor equation is satisfied in one coordinate system then it is also satisfied in every other admissible coordinate system. In special relativity this means that if a physical law can be expressed as a relationship between four-tensors then such a law will automatically be equally valid in every inertial frame.

B.2 Basic Tensors

A tensor B without any indices (a scalar) obeys the transformation rule $B' = B$. Hence, the value of such a tensor depends only on the position in space and not on the choice of coordinate system.

We see in definition B.1.1 that if a tensor of a given type has all components equal to zero in one coordinate system then the same is true in every other coordinate system. Such a tensor is called the **zero tensor** (of its type) and its components are denoted by 0 without any indices.

The most simple nontrivial contravariant tensor is the coordinate differential $\mathrm{d}\boldsymbol{x}$. This *is* a contravariant tensor since

$$\mathrm{d}x^{i'} = \mathrm{d}x^i \frac{\mathrm{d}x^{i'}}{\mathrm{d}x^i} = \mathrm{d}x^i p_i^{i'}. \tag{B.2.1}$$

Moreover, the most basic nontrivial covariant tensor is the gradient $\nabla\phi$. If we denote the components of the gradient by $\phi_{,i}$ then we have that

$$\phi_{,i'} \overset{def}{=} \frac{\partial\phi}{\partial x^{i'}} = \frac{\partial\phi}{\partial x^i}\frac{\partial x^i}{x^{i'}} = \frac{\partial\phi}{\partial x^i}p_{i'}^i = \phi_{,i}p_{i'}^i, \tag{B.2.2}$$

which proves that $\nabla\phi$ *is* a covariant tensor.

The Kronecker delta δ is perhaps the most simple nontrivial example of a mixed tensor. We can prove that it *is* a mixed tensor by first noticing its *index substitution property*

$$\delta_a{}^x A^{abc\ldots}_{def\ldots} = \sum_{a=1}^{N} \delta_a{}^x A^{abc\ldots}_{def\ldots} = A^{xbc\ldots}_{def\ldots}, \tag{B.2.3}$$

where A is an arbitrary tensor and where we have used the fact that $\delta_a{}^x = 0$ if $a \neq x$ and 1 if $a = x$.

With the help of this property we can write

$$\delta_i{}^j p_{i'}^i p_j^{j'} = p_{i'}^j p_j^{j'} = \frac{\partial x^j}{\partial x^{i'}}\frac{\partial x^{j'}}{\partial x^j} = \frac{\partial x^{j'}}{\partial x^{i'}} = \delta_{i'}{}^{j'}, \tag{B.2.4}$$

where we have used the fact that $x^{i'}$ and $x^{j'}$ are independent coordinates when $i' \neq j'$.

Equation (B.2.4) proves that δ is a mixed tensor. Also note that δ is obviously a symmetric tensor since it is the identity matrix.

An important example of an antisymmetric tensor is the **Levi-Civita tensor** (or *permutation symbol*) ϵ, where

$$\epsilon_{ijk} = \begin{cases} 0, & (i=j)\vee(j=k)\vee(i=k) \\ +1, & (i,j,k)\in\{(1,2,3),(3,1,2),(2,3,1)\} \\ -1, & (i,j,k)\in\{(1,3,2),(2,1,3),(3,2,1)\} \end{cases} \tag{B.2.5}$$

B.3 Tensor Algebra

There are four basic arithmetical operations with tensors. They are called *summation*, *outer product*, *contraction* and *index permutation*. All of these operations produce tensors from other tensors.

The **sum** of two tensors A and B of the same but otherwise arbitrary type is defined as the object C with the components

$$C^{abc\ldots}_{efg\ldots} \stackrel{def}{=} A^{abc\ldots}_{efg\ldots} + B^{abc\ldots}_{efg\ldots} \,. \tag{B.3.1}$$

We can prove that C is a tensor by writing

$$\begin{aligned} C^{a'b'c'\ldots}_{e'f'g'\ldots} &= A^{a'b'c'\ldots}_{e'f'g'\ldots} + B^{a'b'c'\ldots}_{e'f'g'\ldots} \\ &= A^{abc\ldots}_{efg\ldots} p^{a'}_{a} p^{e}_{e'} p^{b'}_{b} p^{f}_{f'} \cdots + B^{abc\ldots}_{efg\ldots} p^{a'}_{a} p^{e}_{e'} p^{b'}_{b} p^{f}_{f'} \cdots \\ &= (A^{abc\ldots}_{efg\ldots} + B^{abc\ldots}_{efg\ldots}) p^{a'}_{a} p^{e}_{e'} p^{b'}_{b} p^{f}_{f'} \cdots \\ &= C^{abc\ldots}_{efg\ldots} p^{a'}_{a} p^{e}_{e'} p^{b'}_{b} p^{f}_{f'} \cdots \,. \end{aligned} \tag{B.3.2}$$

Note that we assumed in the final step that the expression $p^{a'}_{a} p^{e}_{e'} p^{b'}_{b} p^{f}_{f'} \cdots$ has the same value for A' and B'. This is always true if we add tensors at the same point in space. However, it is not true in general if we add tensors at *different* points. A sum of such tensors is therefore usually not a tensor.

Next, let A and B be two tensors of arbitrary types. We define the **outer product** C of A and B by simply merging the indices of A and B. For instance:

$$C^{abc\ \ ghi}_{\ def\ \ jkl} \stackrel{def}{=} A^{abc}{}_{def} B^{ghi}{}_{jkl}. \tag{B.3.3}$$

One can easily prove that this procedure gives a new tensor. The same is true for the operations below.

Note that forming the outer product of two tensors of rank n and rank k, respectively, is a simple way of creating a tensor of rank $n + k$.

To **contract** a tensor means to form the components of a new tensor by replacing two *different* free indices, one subscript and one superscript, with a dummy index pair. For instance, if $A^{ab}{}_{cde}$ is an arbitrary component of a tensor A then we can define a new tensor B with the components

$$B^{a}{}_{bc} \stackrel{def}{=} A^{ax}{}_{cdx} = \sum_{x=1}^{N} A^{ax}{}_{cdx}. \tag{B.3.4}$$

The term *contraction* is also used to describe the process of first forming the outer product of two tensors and then summing over the dummy indices. For example:

$$A_{ab} B^{b}{}_{c} = C_{ab}{}^{b}{}_{c} = \sum_{b=1}^{N} C_{ab}{}^{b}{}_{c} \stackrel{def}{=} C_{ac}. \tag{B.3.5}$$

This composition of operations is also referred to as the **inner product** of two tensors. Note that the inner product is not an elementary tensor operation.

Finally, **index permutation** simply means to define the components of a new tensor by rearranging the indices of the components of a given tensor. For example:

$$B_{ab} \stackrel{def}{=} A_{ba}. \tag{B.3.6}$$

Note that the fact that index permutation gives a new tensor means that if a tensor is symmetric/antisymmetric in a given coordinate system then it is also symmetric/antisymmetric in every other coordinate system, since tensor equations such as

$$A_a{}^b = A_b{}^a,\ A_a{}^b = -A_b{}^a \text{ etc.} \tag{B.3.7}$$

are equally valid in every coordinate system.

B.4 Derivatives of Tensors

Let us for the sake of simplicity use the symbol

$$A^{ij\ldots m}_{no\ldots q,\,r} \stackrel{def}{=} \frac{\partial}{\partial x^r}(A^{ij\ldots m}_{no\ldots q}) \tag{B.4.1}$$

to denote the partial derivative of a component of an arbitrary tensor A with respect to x^r.

Moreover, let us define the **partial derivative** $\partial A/\partial x^i$ of a tensor A with respect to x^i as a componentwise operation.

If we calculate the partial derivatives of both sides of equation (B.1.15) with respect to $x^{r'}$ then we obtain the transformation rule

$$A^{i'j'\ldots m'}_{n'o'\ldots q',\,r'} = A^{ij\ldots m}_{no\ldots q,\,r} p^{i'}_i p^{j'}_j \ldots p^{m'}_m p^n_{n'} p^o_{o'} \ldots p^q_{q'} p^r_{r'} + P_1 + P_2 + \ldots, \tag{B.4.2}$$

where P_1, P_2 etc. are terms with partial derivatives of the p:s.

For *linear transformations* the p:s are constants, which implies that the P:s vanish in equation (B.4.2). In this particular case equation (B.4.2) defines a tensor transformation, but the same cannot be said about general coordinate transformations.

By the same argument, higher order partial derivatives such as

$$A^{ij\ldots m}_{no\ldots q,\,rs} \stackrel{def}{=} \frac{\partial^2}{\partial x^r \partial x^s}(A^{ij\ldots m}_{no\ldots q}) \tag{B.4.3}$$

also behave as tensors under linear transformations but not under general transformations.

A similar argument also applies to the derivative $\mathrm{d}A/\,\mathrm{d}c$ of a tensor A with respect to a scalar c.

As a practical example of tensor derivation let us consider partial derivatives of four-tensors. The relevant coordinate transformation is the Poincaré transformation

$$\boldsymbol{x'} = \Lambda \boldsymbol{x} + \boldsymbol{a}, \tag{B.4.4}$$

where Λ is a constant matrix and $\boldsymbol{a}$ is a constant vector.

We can rewrite equation (B.4.4) on component form as

$$x^{\mu'} = \Lambda^{\mu'}{}_{\mu} x^{\mu} + a^{\mu'}. \tag{B.4.5}$$

In particular, four-tensors behave as tensors under *homogenous* Poincaré transformations, which means that

$$x^{\mu'} = \Lambda^{\mu'}{}_{\mu} x^{\mu} \text{ and} \tag{B.4.6}$$

$$x^{\mu'} = p^{\mu'}_{\mu} x^{\mu}. \tag{B.4.7}$$

By comparing the right hand sides of these equations we notice that $\Lambda^{\mu'}{}_{\mu} = p^{\mu'}_{\mu}$. We can therefore write equation (B.4.5) on the form

$$x^{\mu'} = p^{\mu'}_{\mu} x^{\mu} + a^{\mu'}. \tag{B.4.8}$$

Taking the partial derivative of both sides of this equation with respect to (for instance) $x^{\nu'}$ gives the equation

$$p^{\mu'}_{\nu'} = \frac{\partial x^{\mu'}}{\partial x^{\nu'}} = p^{\mu'}_{\mu} \frac{\partial x^{\mu}}{\partial x^{\nu'}} = p^{\mu'}_{\mu} \frac{\partial x^{\mu}}{\partial x^{\nu}} \frac{\partial x^{\nu}}{\partial x^{\nu'}} = \frac{\partial x^{\mu}}{\partial x^{\nu}} p^{\mu'}_{\mu} p^{\nu}_{\nu'} = p^{\mu}_{\nu} p^{\mu'}_{\mu} p^{\nu}_{\nu'}, \tag{B.4.9}$$

which proves that the object $\partial \boldsymbol{x}/\partial x^{\nu}$ transforms as a tensor under Poincaré transformations.

However, note that in *general* relativity the transformation matrix is not constant and partial derivatives are therefore not tensors.

B.5 Metric Spaces

A **metric space** is a set S together with a real valued global "distance" function g (called the **metric**) which for all $x, y, z \in S$ must satisfy the conditions

(i) $g(x, y) = 0 \Leftrightarrow x = y$

(ii) $g(x, y) = g(y, x)$

(iii) $g(x, y) + g(y, z) \geq g(x, z)$ (the triangle inequality).

Note that these conditions together imply that $g(x, y) \geq 0$ for all $x, y \in S$, since

$$g(x,y) = [g(x,y) + g(x,y)]/2 \overset{(2)}{=} [g(x,y) + g(y,x)]/2 \overset{(3)}{\geq} [g(x,x)]/2 \overset{(1)}{=} 0. \tag{B.5.1}$$

More specifically, a set S together with a real *invariant* quadratic form

$$\mathbf{d}\boldsymbol{s}^2 \overset{def}{=} g_{ij}\, \mathrm{d}x^i\, \mathrm{d}x^j, \tag{B.5.2}$$

where g is an invertible matrix and generally a function of position $\boldsymbol{x}$, is called a **properly Riemannian** metric space if $\mathbf{d}\boldsymbol{s}^2 \geq 0$. Note that $\mathbf{d}\boldsymbol{s}^2$ is the metric in such a space ($dx^i = (\boldsymbol{y} - \boldsymbol{x})^i$) and that it fulfills condition $(i), (ii)$ and (iii) above.

If $\mathbf{d}\boldsymbol{s}^2$ is not positive definite then the space is called **pseudo Riemannian**. Moreover, it can be proven that the requirement that $\mathbf{d}\boldsymbol{s}^2$ must be invariant implies that the matrix g in equation (B.5.2) is a tensor, and it is therefore naturally referred to as a **metric tensor**.

Note that we can put components such as g_{12} and g_{21} in the sum on the right hand side of equation (B.5.2) together in terms on the form $(g_{12} + g_{21})x^1x^2$. Since only the sum $g_{12} + g_{21}$ and not the values of the individual terms g_{12} and g_{21} has any influence at all on the value of the metric $\mathbf{d}\boldsymbol{s}^2$, we may add the condition that g should be a symmetric tensor without any loss of generality. For future reference, we will therefore always assume that metric tensors are symmetric.

The **inner product** of any two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ in a properly or pseudo Riemannian space is defined as

$$\boldsymbol{A} \cdot \boldsymbol{B} \overset{def}{=} g_{ij} A^i B^j. \tag{B.5.3}$$

With the inner product (B.5.3) one can also define the **square** $\boldsymbol{A}^2$ and **norm** A of a vector $\boldsymbol{A}$ in a Riemannian space as

$$\boldsymbol{A}^2 \stackrel{def}{=} \boldsymbol{A} \cdot \boldsymbol{A} \tag{B.5.4}$$

$$A \stackrel{def}{=} |\boldsymbol{A}^2|^{1/2}. \tag{B.5.5}$$

Note that in pseudo Riemannian spaces the inner product (B.5.3) is only a **pseudo** inner product since it is not positive definite.

In Riemannian spaces there is also a *fifth basic tensor operation* in addition to the four which were defined in section B.3. It is called **raising and lowering of indices** and it is defined by the following equations:

$$g^{ij} \stackrel{def}{=} (g^{-1})_{ij}, \tag{B.5.6}$$

$$A_{abc\ldots}{}^{m\ o\ldots}_{x} \stackrel{def}{=} g_{nx} A_{abc\ldots}{}^{mno\ldots} \tag{B.5.7}$$

$$A_{a\ c\ldots}^{\ x}{}^{mno\ldots} \stackrel{def}{=} g^{bx} A_{abc\ldots}{}^{mno\ldots}, \tag{B.5.8}$$

where A is a tensor of arbitrary type and the indices b and n were chosen at random.

Note that $g_{nx} = g_{xn}, g^{bx} = g^{xb}$ and $g^{ij} = g^{ji}$ since g is a symmetric tensor. Moreover, the convention of writing subscripts and superscripts in a staggered form is only meaningful in connection with the operation raising and lowering of indices. If one has no intention of raising and lowering indices then an element $A_{ij}{}^{k}$ might as well be written as A^{k}_{ij}. The latter notation is often used in introductory texts on tensor calculus.

The metric of spacetime is the **Minkowski metric**

$$\mathbf{d}\boldsymbol{s}^2 \stackrel{def}{=} -x_0^2 + x_1^2 + x_2^2 + x_3^2 = -c^2t^2 + x^2 + y^2 + z^2 \tag{B.5.9}$$

and the corresponding **Minkowski metric tensor** is denoted as η, where

$$\eta \stackrel{def}{=} \mathrm{diag}(-1, 1, 1, 1) = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{B.5.10}$$

Note that the definition of η follows directly from the definition of the Minkowski metric and equation (B.5.2). Also note that due to different sign conventions for the metric η is sometimes defined as $\mathrm{diag}(1, -1, -1, -1)$.

The raising and lowering of indices is a particularly easy procedure in spacetime since for an arbitrary four-tensor $\boldsymbol{A}$

$$A_0 = \eta_{0\mu} A^\mu = -A^0 \tag{B.5.11}$$

$$A_i = \eta_{i\mu} A^\mu = A^i \ (i = 1, 2, 3). \tag{B.5.12}$$

Let us now prove that preserving the Minkowski metric $\mathbf{d}\boldsymbol{s}^2$ is equivalent with preserving either the **squared displacement** $\boldsymbol{\Delta s}^2$ or the inner product $\boldsymbol{x}\cdot\boldsymbol{y}$ on spacetime. We mentioned in chapter 6 that the transformation matrix Λ of an arbitrary Lorentz transformation satisfies the equation

$$\eta_{\mu\nu} \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma = \eta_{\rho\sigma}. \tag{B.5.13}$$

We therefore have the identities

$$\eta_{\rho\sigma}\,\mathrm{d}x^\rho\,\mathrm{d}y^\sigma \equiv \eta_{\mu\nu}\Lambda^\mu{}_\rho\,\mathrm{d}x^\rho\Lambda^\nu{}_\sigma\,\mathrm{d}y^\sigma \Leftrightarrow \mathbf{d}\boldsymbol{x}^2 \equiv \mathbf{d}\boldsymbol{x}'^2 \tag{B.5.14}$$

$$\eta_{\rho\sigma}(x-y)^\rho(x-y)^\sigma \equiv \eta_{\mu\nu}\Lambda^\mu{}_\rho(x-y)^\rho\Lambda^\nu{}_\sigma(x-y)^\sigma \Leftrightarrow \boldsymbol{\Delta s}^2 \equiv (\boldsymbol{\Delta s}')^2 \tag{B.5.15}$$

$$\eta_{\rho\sigma}x^\rho y^\sigma \equiv \eta_{\mu\nu}\Lambda^\mu{}_\rho x^\rho\Lambda^\nu{}_\sigma y^\sigma \Leftrightarrow \boldsymbol{x}\cdot\boldsymbol{y} \equiv \boldsymbol{x}'\cdot\boldsymbol{y}', \tag{B.5.16}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are two arbitrary elements of $\mathbb{R}^{1,3}$.

Moreover, if either one of the identities (B.5.14)–(B.5.16) is satisfied for all vectors $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^{1,3}$ then equation (B.5.13) must also be satisfied. We therefore conclude that these four statements are in fact equivalent. This is the reason why Lorentz transformations are sometimes said to satisfy equation (B.5.13) and other times to preserve the Minkowski metric, the squared displacement or the inner product on spacetime.

B.6 Formulas and Examples

The best way to learn tensor calculus is to actually perform calculations with tensors. Let us therefore have a look at some simple formulas and examples. We will restrict our attention to four-tensors since most of the calculation in this report involves that type of tensors.

The inner product, square and norm of two arbitrary vectors $\boldsymbol{A},\boldsymbol{B}\in\mathbb{R}^{1,3}$ are given by

$$\boldsymbol{A}\cdot\boldsymbol{B} = \eta_{\mu\nu}A^\mu B^\nu = -A^0B^0 + A^1B^1 + A^2B^2 + A^3B^3 \tag{B.6.1}$$

$$\boldsymbol{A}^2 = \boldsymbol{A}\cdot\boldsymbol{A} = -(A^0)^2 + (A^1)^2 + (A^2)^2 + (A^3)^2 \tag{B.6.2}$$

$$A = |\boldsymbol{A}^2| = |-(A^0)^2 + (A^1)^2 + (A^2)^2 + (A^3)^2|^{1/2}. \tag{B.6.3}$$

Moreover, multiplication between a 4×4 matrix A and a vector b can be written in four different ways as

$$(Ab)_\mu = A_{\mu\nu}b^\nu \text{ or } (Ab)_\mu = A_\mu{}^\nu b_\nu \tag{B.6.4}$$

$$(Ab)^\mu = A^\mu{}_\nu b^\nu \text{ or } (Ab)^\mu = A^{\mu\nu}b_\nu. \tag{B.6.5}$$

Similarly, multiplication between a matrix A and matrix B can be written on eight different forms as

$$(AB)_{\mu\nu} = A_{\mu\rho}B^\rho{}_\nu \text{ or } (AB)_{\mu\nu} = A_\mu{}^\rho B_{\rho\nu} \tag{B.6.6}$$

$$(AB)_\mu{}^\nu = A_{\mu\rho}B^{\rho\nu} \text{ or } (AB)_\mu{}^\nu = A_\mu{}^\rho B_\rho{}^\nu \tag{B.6.7}$$

$$(AB)^\mu{}_\nu = A^\mu{}_\rho B^\rho{}_\nu \text{ or } (AB)^\mu{}_\nu = A^{\mu\rho}B_{\rho\nu} \tag{B.6.8}$$

$$(AB)^{\mu\nu} = A^\mu{}_\rho B^{\rho\nu} \text{ or } (AB)^{\mu\nu} = A^{\mu\rho}B_\rho{}^\nu. \tag{B.6.9}$$

One of the most important aspects of solving a tensor equation is usually the rewriting of expressions on a desirable form (which is known as "index gymnastics"). Let us therefore try to rewrite a product of two matrices A and B and a vector b on component form with the help of only equation (B.6.4):

$$(ABb)_\mu = A_\mu{}^\nu(Bb)_\nu = A_\mu{}^\nu B_\nu{}^\rho b_\rho. \tag{B.6.10}$$

Notice that we only introduced new indices in the form of dummy index pairs.

Finally, as we mentioned in section 6.3 contracting an antisymmetric tensor A of rank 2 with a symmetric tensor B of the same rank always gives a zero result. This is easy to show; consider for instance

$$A_{\mu\nu}B^{\mu\nu} = A_{\nu\mu}(-B^{\nu\mu}) = \{\mu \to \nu, \nu \to \mu\} = -A_{\mu\nu}B^{\mu\nu}, \tag{B.6.11}$$

which implies that $A_{\mu\nu}B^{\mu\nu} = 0$.

Specifically this means that contracting an antisymmetric tensor A of rank 2 with an arbitrary rank 2 tensor C removes the symmetric part of C:

$$A_{\mu\nu}C^{\mu\nu} = A_{\mu\nu}(C^{(\mu\nu)} + C^{[\mu\nu]}) = A_{\mu\nu}C^{[\mu\nu]}. \tag{B.6.12}$$

Appendix C

Kets and Ket Spaces

In this appendix we will give a brief introduction to Hilbert spaces and Dirac notation.

C.1 Hilbert Spaces

Let us begin with the definition of a Hilbert space.

Definition C.1.1. *A* **Hilbert space** $\mathcal{H}$ *is a finite or infinite dimensional vector space together with an inner product* $\langle\cdot,\cdot\rangle$ *such that the norm in* $\mathcal{H}$*, defined by*

$$|f| = \sqrt{\langle f, f\rangle},\ \forall f \in \mathcal{H}, \tag{C.1.1}$$

makes $\mathcal{H}$ *a* complete metric space*.*

If the metric given by the inner product is not *complete* then $\mathcal{H}$ is simply an ordinary inner product space rather than a Hilbert space. Also note that due to the square root in equation (C.1.1) the inner product has to be positive definite and may therefore not be a *pseudo* inner product.

Every inner product space V is also a metric space, where the metric g is given by

$$g(\boldsymbol{u}, \boldsymbol{v}) = \langle \boldsymbol{u} - \boldsymbol{v}, \boldsymbol{u} - \boldsymbol{v}\rangle,\ \boldsymbol{u}, \boldsymbol{v} \in V. \tag{C.1.2}$$

We will now define the specific choice of an inner product used in quantum mechanics before we give the definition of a *complete metric space*.

Definition C.1.2. *An* **Hermitian inner product** $\langle\cdot,\cdot\rangle$ *on a complex vector space* V *is a complex valued function which for all* $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ *and* $c \in \mathbb{C}$ *satisfies the conditions*

(i) $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{u}\rangle^*$

(ii) $\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w}\rangle = \langle \boldsymbol{u}, \boldsymbol{w}\rangle + \langle \boldsymbol{v}, \boldsymbol{w}\rangle$

(iii) $\langle c\boldsymbol{u}, \boldsymbol{v}\rangle = c\langle \boldsymbol{u}, \boldsymbol{v}\rangle$

(iv) $\langle \boldsymbol{u}, \boldsymbol{u}\rangle \geq 0,\ \langle \boldsymbol{u}, \boldsymbol{u}\rangle = 0 \Leftrightarrow \boldsymbol{u} = 0,$

where $\langle \boldsymbol{v}, \boldsymbol{u}\rangle^*$ *denotes the complex conjugate of* $\langle \boldsymbol{v}, \boldsymbol{u}\rangle$*.*

A complex vector space with an Hermitian inner product is called an **Hermitian inner product space***.*

Note that condition (i) and (iii) give that $\langle \boldsymbol{u}, c\boldsymbol{v}\rangle = c^*\langle \boldsymbol{u}, \boldsymbol{v}\rangle$. One say that the Hermitian inner product is linear in its first argument and **antilinear** (or **sesquilinear**) in its second argument.

Definition C.1.3. *A* **complete metric space** *is a metric space in which every* Cauchy sequence *converges. Moreover, a* **Cauchy sequence** *is a sequence* $\{a_n\}_{n=1}^{\infty}$ *such that the metric* g *satisfies the condition*

$$\lim_{min(m,n)\to\infty} g(a_m, a_n) = 0. \tag{C.1.3}$$

C.2 State Vectors

In quantum mechanics the possible states of a physical system are represented by vectors in a complex Hilbert space $\mathcal{H}$. Such vectors are naturally called **state vectors**. State vectors on the form $c\psi$, where $c \neq 0$ is an arbitrary complex number and $0 \neq \psi \in \mathcal{H}$ is a fixed state vector, represent the same physical state and are said to belong to the same **ray** $\mathscr{R} = \{c\psi, c \in \mathbb{C}\}$ in $\mathcal{H}$. Since the constant c is of no physical significance one often normalize state vectors, which also gives simpler formulas.

C.3 Dirac Notation

Let us now say a few words about the notation developed by P. A. M. Dirac called **Dirac notation** (or **bra-ket notation**).

In this notation state vectors, which are called **kets**, are denoted on the form $|\cdot\rangle$. The state ket of a given physical system contains all the information about the state of the system and is an element of the so called **ket space** $\mathcal{H}$ of the system, which is a complex Hilbert space.

Since the ket space is a vector space two kets $|\psi_1\rangle, |\psi_2\rangle \in \mathcal{H}$ may be added

$$|\psi_1\rangle + |\psi_2\rangle \tag{C.3.1}$$

and their sum is another ket in $\mathcal{H}$.

For the same reason, a scalar multiple $c|\psi\rangle$ of any ket in $\mathcal{H}$ with an arbitrary complex number c is another ket in $\mathcal{H}$. We also have that

$$c|\psi\rangle = |\psi\rangle c, \, \forall c \in \mathbb{C} \text{ and } \forall |\psi\rangle \in \mathcal{H}, \tag{C.3.2}$$

even though commutativity for multiplication by scalars is not required by the definition of a vector space.

If $c = 0$ in equation (C.3.2) then the product is the **null ket**, which is denoted by 0.

Moreover, to every ket $|\psi\rangle$ in the ket space corresponds a **bra** $\langle\psi|$ in the associated **bra space** $\mathcal{H}^*$. The bra $\langle\psi|$ is the *dual* to the ket $|\psi\rangle$ and the bra space is the *dual space* to the ket space (see appendix D for a definition of dual spaces). Note that the mapping $|\psi\rangle \to \langle\psi|$ is bijective and that the dual to $c|\psi\rangle$, where $c \in \mathbb{C}$, is $c^*\langle\psi|$ rather than $c\langle\psi|$ ([14], p. 13).

More generally, we have that

$$c|\psi_1\rangle + d|\psi_2\rangle \overset{DC}{\leftrightarrow} c^*\langle\psi_1| + d^*\langle\psi_2|, \, \forall |\psi_1\rangle, |\psi_2\rangle \in \mathcal{H} \text{ and } \forall c, d \in \mathbb{C}, \tag{C.3.3}$$

where DC stands for **dual correspondence**.

An arbitrary operator $\hat{A}$ acting on the ket space acts on the kets from the left, $\hat{A}|\psi\rangle$, while an arbitrary operator $\hat{B}$ acting on the bra space acts on the bras from the right, $\langle\psi|\hat{B}$. Note that in general $\langle\psi|\hat{A}$ is *not* the dual to $\hat{A}|\psi\rangle$.

An operator $\hat{B}$ which satisfies the condition

$$\hat{A}|\psi\rangle \overset{DC}{\leftrightarrow} \langle\psi|\hat{B},\ \forall|\psi\rangle \in \mathcal{H} \tag{C.3.4}$$

is by definition the **Hermitian adjoint** of $\hat{A}$ and it is denoted by $\hat{A}^\dagger$.

Two operators $\hat{A}$ and $\hat{B}$ are said to be **equal** (which is written as $\hat{A} = \hat{B}$) iff

$$\hat{A}|\psi\rangle = \hat{B}|\psi\rangle,\ \forall|\psi\rangle \in \mathcal{H}. \tag{C.3.5}$$

If $\hat{A} = \hat{A}^\dagger$ then $\hat{A}$ is called an **Hermitian** operator.

In Dirac notation the **inner product** of two arbitrary state vectors ψ_1 and ψ_2 is written as

$$\langle\psi_2|\psi_1\rangle \overset{def}{=} \langle\psi_1, \psi_2\rangle. \tag{C.3.6}$$

Note that ψ_1 and ψ_2 traded places. This is necessary in order to make the inner product linear on the "ket-side" and antilinear on the "bra-side" (se definition C.1.2).

The **outer product** of a given ket $|\psi_1\rangle$ in $\mathcal{H}$ and a given bra $\langle\psi_2|$ in $\mathcal{H}^*$ is denoted by $|\psi_1\rangle\langle\psi_2|$ and defined as

$$\Big(|\psi_1\rangle\langle\psi_2|\Big)|\psi\rangle \overset{def}{=} |\psi_1\rangle\langle\psi_2|\psi\rangle,\ \forall|\psi\rangle \in \mathcal{H}. \tag{C.3.7}$$

Moreover, in quantum mechanics operator multiplication is in general associative, i.e. $\hat{A}(\hat{B}\hat{C}) = (\hat{A}\hat{B})\hat{C}$, but noncommutative, i.e. $\hat{A}\hat{B} \neq \hat{B}\hat{A}$.

Appendix D

Dual Spaces

We mentioned in section 5.1 that the set V^* of all linear functions $f : V \to \mathbb{R}$, where V is a vector space, is called the *dual space* to V and that the elements of V^* are called functionals. A more general definition of a dual space is given below[1].

Definition D.0.1. *The* **dual space** V^* *to a given vector space* V *over a field* F *is the set of all linear functions* $f : V \to F$*. Such functions are by definition called* **functionals**. *Note that* V^* *also becomes a vector space over* F *if vector addition and scalar multiplication on* V^* *are defined as:*

$$(f+g)(\boldsymbol{x}) \stackrel{def}{=} f(\boldsymbol{x}) + g(\boldsymbol{x}),\ \forall f, g \in V^* \text{ and } \forall \boldsymbol{x} \in V \tag{D.0.1}$$

$$(cf)(\boldsymbol{x}) \stackrel{def}{=} cf(\boldsymbol{x}),\ \forall f \in V^*, \forall \boldsymbol{x} \in V \text{ and } \forall c \in F. \tag{D.0.2}$$

If V *is a finite dimensional vector space then* V^* *(as a vector space) has the same dimension as* V*. Otherwise* $\dim(V^*) > \dim(V)$*. Moreover, the elements of* V^* *are sometimes called* **covectors** *or* **one-forms**.

Let us now define a field.

Definition D.0.2. *A* **field** *is a set* F *together with two binary operators* $+$ *and* $*$ *which for all* $a, b, c \in F$ *satisfy the conditions:*

Additive associativity: $(a + b) + c = a + (b + c)$

Additive commutativity: $a + b = b + a$

Additive identity: $\exists 0 \in F :\ 0 + a = a + 0 = a$

Additive inverse: $\exists(-a) \in F :\ a + (-a) = (-a) + a = 0$

Multiplicative associativity: $(a * b) * c = a * (b * c)$

Multiplicative commutativity: $a * b = b * a$

Multiplicative identity: $\exists 1 \in F :\ 1 \neq 0$ *and* $1 * a = a * 1 = a$

[1] To be precise, definition D.0.1 is the definition of an **algebraic** dual space. Every vector space has such a dual space, but the dual space to a *topological* vector space also has a subspace of continuous functionals known as a **continuous dual space**.

Multiplicative inverse: $\forall a \neq 0 \; \exists a^{-1} \in F : \; a * a^{-1} = a^{-1} * a = 1$

Left distributivity: $a * (b + c) = (a * b) + (a * c)$

Right distributivity: $(b + c) * a = (b * a) + (c * a)$

Note that since the multiplicative identity must differ from the additive identity every field has at least two elements.

Simple examples of fields include $\mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$, for which the $+$ and $*$ operators are the ordinary addition and multiplication operators. Note that $\mathbb{Z}$ is *not* a field since the elements of $\mathbb{Z}$ have no multiplicative inverses in $\mathbb{Z}$.

To give an example of a dual space let us consider the dot product on $\mathbb{R}^n$. For any fixed vector $\boldsymbol{v} \in \mathbb{R}^n$ we may think of $\boldsymbol{v}\cdot$ as an operator which acts on an arbitrary vector $\boldsymbol{x} \in \mathbb{R}^n$ and gives a real number $\boldsymbol{v} \cdot \boldsymbol{x}$. $\mathbb{R}^n$ is a vector space and we mentioned above that $\mathbb{R}$ is a field. Moreover, the operator $\boldsymbol{v}\cdot$ is clearly linear since

$$\boldsymbol{v} \cdot (a\boldsymbol{x} + b\boldsymbol{y}) = a\boldsymbol{v} \cdot \boldsymbol{x} + b\boldsymbol{v} \cdot \boldsymbol{y}, \; \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \text{ and } \forall a, b \in \mathbb{R}. \tag{D.0.3}$$

This means that $\boldsymbol{v}\cdot$ is a functional and therefore an element of the dual space $(\mathbb{R}^n)^*$ to $\mathbb{R}^n$.

If the elements of $\mathbb{R}^n$ are written as column vectors then their dual elements in $(\mathbb{R}^n)^*$ are the corresponding row vectors, since $\boldsymbol{v} \cdot \boldsymbol{x} = \boldsymbol{v}^T\boldsymbol{x}$ for all $\boldsymbol{v}, \boldsymbol{x} \in \mathbb{R}^n$.

Another example of a dual space is the bra space $\mathcal{H}^*$ associated with the ket space $\mathcal{H}$ of a physical system. The ket space is a complex Hilbert space and therefore a vector space. Any given element $\langle\psi_1|$ in the bra space is an operator which acts on an arbitrary ket $|\psi\rangle \in \mathcal{H}$ and gives a complex number $\langle\psi_1|\psi\rangle$. Since $\mathbb{C}$ is a field and the operator $\langle\psi_1|$ is linear,

$$\langle\psi_1|c\psi_2 + d\psi_3\rangle = c\langle\psi_1|\psi_2\rangle + d\langle\psi_1|\psi_3\rangle, \; \forall|\psi_2\rangle, |\psi_3\rangle \in \mathcal{H} \text{ and } \forall c, d \in \mathbb{C}, \tag{D.0.4}$$

we can conclude that $\langle\psi_1|$ is a functional and indeed an element of $\mathcal{H}^*$.

Let us now consider a general bilinear form

$$\langle\cdot,\cdot\rangle : V \times V \to F, \; (\boldsymbol{x}, \boldsymbol{y}) \mapsto \langle\boldsymbol{x}, \boldsymbol{y}\rangle, \tag{D.0.5}$$

where V is a vector space over a field F.

Let $\boldsymbol{x}$ be a given vector in V and $f_{\boldsymbol{x}}$ be defined by

$$f_{\boldsymbol{x}} \overset{def}{=} \langle\boldsymbol{x},\cdot\rangle : V \to F, \; \boldsymbol{y} \mapsto \langle\boldsymbol{x}, \boldsymbol{y}\rangle. \tag{D.0.6}$$

Then $f_{\boldsymbol{x}}$ is clearly an element of V^*. Moreover, if $\langle\cdot,\cdot\rangle$ is **nondegenerate** then the mapping $\boldsymbol{x} \mapsto f_{\boldsymbol{x}}$ is a by the definition of a nondegenerate bilinear form a group isomorphism from V onto a subspace $\tilde{V}^*$ of V^*, assuming that vector addition and multiplication by scalars have been defined on V^* (every vector space is a group under vector addition). If V is finite dimensional then $\tilde{V}^* = V^*$. Conversely, a given group isomorphism $V \to \tilde{V}^*$, $\boldsymbol{x} \mapsto g_{\boldsymbol{x}}$, from V onto a subspace $\tilde{V}^*$ of V^* uniquely determines a nondegenerate bilinear form

$$\langle\cdot,\cdot\rangle : V \times V \to F, \; (\boldsymbol{x}, \boldsymbol{y}) \mapsto \langle\boldsymbol{x}, \boldsymbol{y}\rangle \overset{def}{=} g_{\boldsymbol{x}}(\boldsymbol{y}). \tag{D.0.7}$$

Hence, to every nondegenerate bilinear form $\langle\cdot,\cdot\rangle$ on V corresponds a uniquely determined isomorphism from V onto a subspace $\tilde{V}^*$ of V^* and vice versa.

The inner product on $\mathbb{R}^n$ is a nondegenerate bilinear form and the *homomorphism* from $\mathbb{R}^n$ to $(\mathbb{R}^n)^*$ simply means that if $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}$ then $\boldsymbol{z}\cdot = (\boldsymbol{x}\cdot) + (\boldsymbol{y}\cdot)$, which is obviously true. Moreover, the mapping $\mathbb{R} \to \mathbb{R}^*$, $\boldsymbol{x} \mapsto \boldsymbol{x}\cdot$ is clearly bijective, which proves that it is a group isomorphism.

Another important example of a nondegenerate bilinear form is the inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = g_{ij} x^i y^j$ on a arbitrary metric space V, where g is the metric tensor. For any given $\boldsymbol{x} \in V$ we know that $\langle \boldsymbol{x}, \cdot \rangle_j = g_{ij} x^i$ is the jth component of an operator $\langle \boldsymbol{x}, \cdot \rangle$ which acts on an arbitrary vector $\boldsymbol{y} \in V$ and gives a real number $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$. Since $\langle \boldsymbol{x}, \cdot \rangle$ is a linear operator it is an element of V^*. Moreover, if we define x_j as

$$x_j \stackrel{def}{=} g_{ij} x^i, \tag{D.0.8}$$

then the inner product on V can simply be written as $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_j y^j$.

This means that the metric tensor g is actually an operator $g : V \to V^*$, $\boldsymbol{x} \mapsto \boldsymbol{x}^*$. So when we for instance use the Minkowski metric to lower the index of a vector in $\mathbb{R}^{1,3}$ then we are actually mapping the element on its dual element in the dual space $(\mathbb{R}^{1,3})^*$ to spacetime.

If V is vector space over $\mathbb{C}$ then the natural choice of an inner product is an Hermitian inner product (defined in section C.1). Such an inner product $\langle \cdot, \cdot \rangle$ satisfies the conditions

$$\langle \boldsymbol{u}, c\boldsymbol{v} + d\boldsymbol{w} \rangle = c^* \langle \boldsymbol{u}, \boldsymbol{v} \rangle + d^* \langle \boldsymbol{u}, \boldsymbol{w} \rangle \tag{D.0.9}$$

$$\langle c\boldsymbol{v} + d\boldsymbol{w}, \boldsymbol{u} \rangle = c \langle \boldsymbol{v}, \boldsymbol{u} \rangle + d \langle \boldsymbol{w}, \boldsymbol{u} \rangle, \tag{D.0.10}$$

for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $c, d \in \mathbb{C}$.

In this case the inner product determines an isomorphism from V to the *complex conjugate* $\bar{V}^*$ of V^*. The **complex conjugate** of V^* is the set of all functionals $f : V \to \mathbb{C}$, $\boldsymbol{x} \to f(\boldsymbol{x})$ such that

$$f(c\boldsymbol{x}) = c^* f(\boldsymbol{x}), \, \forall \boldsymbol{x} \in V \text{ and } \forall c \in \mathbb{C}. \tag{D.0.11}$$

Note that for a ket space the *homomorphism* from V to $\bar{V}^*$ can clearly be seen in equation (C.3.3) in section C.3.

Appendix E

Matrix Lie Groups

In this appendix we will prove theorem 2.2.1 in section 2.2 which states that *every matrix Lie group is a Lie group.*

Let $M_n(\mathbb{C})$ be the set of all $n \times n$ matrices with complex entries. We recall that a matrix Lie group is a closed subgroup of $GL(n, \mathbb{C})$ and that a Lie group is a differentiable manifold which is also a group $\mathcal{G}$, with the properties that the group operation $\mathcal{G} \times \mathcal{G} \to \mathcal{G}$, $(g_1, g_2) \mapsto g_1 g_2$ and the inverse mapping $\mathcal{G} \to \mathcal{G}$, $g \mapsto g^{-1}$ are both differentiable. More specifically, a Lie group is a *smooth* (i.e. infinitely differentiable) manifold!

For any matrix Lie group $\mathcal{L}$ the group operation is matrix multiplication. If l_1 and l_2 are arbitrary elements of the underlying set of $\mathcal{L}$ then the entries of their product $l_1 l_2$ are polynomials in the entries of l_1 and l_2 and therefore smooth. This proves that the group operation $\mathcal{L} \times \mathcal{L} \to \mathcal{L}$, $(l_1, l_2) \mapsto l_1 l_2$ is differentiable.

Moreover, for any given $l \in \mathcal{L}$ the equation

$$l\boldsymbol{x} = \boldsymbol{e}_j \Leftrightarrow \boldsymbol{x} = l^{-1}\boldsymbol{e}_j \stackrel{def}{=} \boldsymbol{l}_j^{-1},\ \boldsymbol{x} \in \mathbb{C}^n \tag{E.0.1}$$

gives the jth column $\boldsymbol{l}_j^{-1}$ of l^{-1}, where $\boldsymbol{e}_j$ is the jth column of the $n \times n$ identity matrix $\mathbb{1}$ in $\mathbb{C}^{n \times n}$.

The ith entry of $\boldsymbol{x} = \boldsymbol{l}_j^{-1}$ is therefore the element $(l^{-1})^{ij}$ of l^{-1}, which is given by Cramér's rule

$$(l^{-1})^{ij} = \frac{\det l_i(\boldsymbol{e}_j)}{\det l}, \tag{E.0.2}$$

where $l_i(\boldsymbol{e}_j)$ is the matrix obtained from l by replacing the ith column of l with the jth column of $\mathbb{1}$.

The determinants in (E.0.2) can be calculated by the method of cofactor expansion and are therefore polynomials in the entries of l (including $\det l_i(\boldsymbol{e}_j)$). This gives that $(l^{-1})^{ij}$ is a rational function and the inverse operation is consequently smooth, which proves that the inverse mapping $\mathcal{L} \to \mathcal{L}$, $l \mapsto l^{-1}$ is differentiable.

We now have to prove that $\mathcal{L}$ is a smooth manifold (the definition of a differentiable manifold is given in chapter A). To do so, it suffices to prove that $\mathcal{L}$ is a *smooth embedded submanifold* of $M_n(\mathbb{C})$ ([35], p. 22).

Definition E.0.3. *A subset S of a smooth manifold M is a n-dimensional* **smooth embedded submanifold** *of M if there exists a covering $\{m_i\}$ of S by open sets (i.e. small neighborhoods) in the ambient smooth manifold M such that the elements of $\{m_i \cap M\}$ are all connected n-dimensional open subsets of M.*

Note that $M_n(\mathbb{C})$ is a smooth manifold. In order to prove that $\mathcal{L}$ is a smooth embedded submanifold of $M_n(\mathbb{C})$ we need the following lemma.

Lemma E.0.1. Let $M_\epsilon = \{m \in m_n(\mathbb{C});\ ||m|| < \epsilon\}$ and $E_\epsilon = \{\exp(m);\ m \in M_\epsilon\}$. If $\mathcal{L}$ is a matrix Lie group with the Lie algebra $\mathfrak{l}$, then there exists an $\epsilon \in (0, \ln 2)$ such that $x \in E_\epsilon \Rightarrow (x \in \mathcal{L} \Leftrightarrow \log(x) \in \mathfrak{l})$.

The proof of this lemma will be omitted here, but may be found in chapter 2 of [35]. Note though that the condition $0 < \epsilon < \ln 2$ guarantees that $\log(\exp x)$ is well-defined and equal to x for all $x \in E_\epsilon$.

Also note that there is nothing obvious about lemma E.0.1 since the exponential mapping for $\mathfrak{l}$ may not cover all of $\mathcal{L}$, i.e. $\mathcal{L}$ may not be an exponential group (see section 2.3). For instance, the matrix

$$l = \begin{pmatrix} -1 & 1 \\ 0 & -1 \end{pmatrix} \tag{E.0.3}$$

is in $SL(2, \mathbb{C})$, and yet one can easily show that there is no matrix $x \in \mathfrak{sl}(2, \mathbb{C})$ such that $\exp(x) = l$ ([35], p. 49).

We will now prove that $\mathcal{L}$ is a smooth embedded submanifold of $M_n(\mathbb{C})$ and therefore a Lie group.

Proof. Let $\epsilon \in (0, \ln 2)$ be such that E.0.1 applies and x_0 any given element of the underlying set of $\mathcal{L}$. Next, consider a neighborhood $x_0 E_\epsilon = \{x_0 x_1;\ x_1 \in E_\epsilon\}$ of x_0 in $M_n(\mathbb{C})$. Let x be an element of $x_0 E_\epsilon$ and note that $x \in x_0 E_\epsilon \Leftrightarrow x_0^{-1} x \in E_\epsilon$. This means that there exists a matrix $X \in M_\epsilon$ such that $x_0^{-1} x = \exp(X)$. Lemma E.0.1 therefore gives that $x_0^{-1} x \in \mathcal{L} \Leftrightarrow \log(x_0^{-1} x) \in \mathfrak{l} \Leftrightarrow \log[\exp(X)] \in \mathfrak{l} \Leftrightarrow X \in \mathfrak{l}$. We also have that $x_0^{-1} x \in \mathcal{L} \Leftrightarrow x \in \mathcal{L}$ since

$$x \in \mathcal{L} \Rightarrow x_0^{-1} x \in \mathcal{L} \tag{E.0.4}$$

$$x_0^{-1} x \in \mathcal{L} \Rightarrow x \in x_0 \mathcal{L} \Rightarrow x \in \mathcal{L}, \tag{E.0.5}$$

where we have used the fact that $x_0^{-1} \in \mathcal{L}$ and that $\mathcal{L}$ is closed under matrix multiplication.

We have thus proven that for any $x \in x_0 E_\epsilon$

$$x \in \mathcal{L} \Leftrightarrow X \in \mathfrak{l}, \tag{E.0.6}$$

for some $X \in M_\epsilon$.

In other words, we can cover $\mathcal{L}$ with small neighborhoods $x_0 E_\epsilon$ to the elements $x_0 \in \mathcal{L}$ such that $x_0 E_\epsilon$ "behaves" as $\mathfrak{l}$, which is naturally a connected open subset of $M_n(\mathbb{C})$. □

Appendix F

Central Charges in the Poincaré Algebra

The following calculation shows how to eliminate central charges from the Poincaré algebra, see ([5], pp. 84-86). This follows approximately the same procedure which was shown for a general case in section 6.7.1. With central charges the commutation relations for the defining representation of the Poincaré algebra become

$$\begin{aligned} i\left[\hat{J}^{\rho\sigma},\, \hat{J}^{\mu\nu}\right] &= \eta^{\sigma\mu}\hat{J}^{\rho\nu} - \eta^{\rho\mu}\hat{J}^{\sigma\nu} - \eta^{\nu\rho}\hat{J}^{\mu\sigma} + \eta^{\nu\sigma}\hat{J}^{\mu\rho} + C^{\mu\nu,\,\rho\sigma} \\ i\left[\hat{P}^{\rho},\, \hat{J}^{\mu\nu}\right] &= \eta^{\rho\mu}\hat{P}^{\nu} - \eta^{\rho\nu}\hat{P}^{\mu} + C^{\mu\nu,\,\rho} \\ \left[\hat{P}^{\rho},\, \hat{P}^{\mu}\right] &= C^{\mu,\,\rho} \end{aligned} \tag{F.0.1}$$

where the central charges C adhere to the following antisymmetric condition, which can easily be found by adding the expressions above to commutation relations where the generators have switched places,

$$\begin{aligned} C^{\mu\nu,\,\rho\sigma} &= -C^{\rho\sigma,\,\mu\nu} \\ C^{\mu\nu,\,\rho} &= -C^{\rho,\,\mu\nu} \\ C^{\mu,\,\rho} &= -C^{\rho,\,\mu}. \end{aligned} \tag{F.0.2}$$

These charges must also obey the Jacobi identity. Writing out this identity for each of the commutation relations mentioned above, we are able to deduce some additional information about the charges. For the pure translational commutation generators, the Jacobi identity with three of the components of $\hat{\boldsymbol{P}}$ is automatically satisfied. Yet, for the others we obtain

$$\begin{aligned} &\left[\hat{J}^{\kappa\lambda},\, [\hat{J}^{\rho\sigma},\, \hat{J}^{\mu\nu}]\right] + \left[\hat{J}^{\mu\nu},\, [\hat{J}^{\kappa\lambda},\, \hat{J}^{\rho\sigma}]\right] + \left[\hat{J}^{\rho\sigma},\, [\hat{J}^{\mu\nu},\, \hat{J}^{\kappa\lambda}]\right] = 0 \\ &\left[\hat{J}^{\kappa\lambda},\, [\hat{J}^{\rho\sigma},\, \hat{P}^{\mu}]\right] + \left[\hat{P}^{\mu},\, [\hat{J}^{\kappa\lambda},\, \hat{J}^{\rho\sigma}]\right] + \left[\hat{J}^{\rho\sigma},\, [\hat{P}^{\mu},\, \hat{J}^{\kappa\lambda}]\right] = 0 \\ &\left[\hat{J}^{\kappa\lambda},\, [\hat{P}^{\rho},\, \hat{P}^{\mu}]\right] + \left[\hat{P}^{\mu},\, [\hat{J}^{\kappa\lambda},\, \hat{P}^{\rho}]\right] + \left[\hat{P}^{\rho},\, [\hat{P}^{\mu},\, \hat{J}^{\kappa\lambda}]\right] = 0. \end{aligned} \tag{F.0.3}$$

Now combining the Jacobi identities with (F.0.1) will result in the following conditions on the central charges

$$
\begin{aligned}
0 &= \eta^{\sigma\mu} C^{\rho\nu,\,\kappa\lambda} - \eta^{\rho\mu} C^{\sigma\nu,\,\kappa\lambda} - \eta^{\nu\rho} C^{\mu\sigma,\,\kappa\lambda} + \eta^{\nu\sigma} C^{\mu\rho,\,\kappa\lambda} + \\
&+ \eta^{\lambda\rho} C^{\kappa\sigma,\,\mu\nu} - \eta^{\kappa\rho} C^{\lambda\sigma,\,\mu\nu} - \eta^{\sigma\kappa} C^{\rho\lambda,\,\mu\nu} + \eta^{\sigma\lambda} C^{\rho\kappa,\,\mu\nu} + \\
&+ \eta^{\nu\kappa} C^{\mu\lambda,\,\rho\sigma} - \eta^{\mu\kappa} C^{\nu\lambda,\,\rho\sigma} - \eta^{\lambda\mu} C^{\kappa\nu,\,\rho\sigma} + \eta^{\lambda\nu} C^{\kappa\mu,\,\rho\sigma} \\
0 &= \eta^{\sigma\mu} C^{\rho\,\kappa\lambda} - \eta^{\rho\mu} C^{\sigma,\,\kappa\lambda} - \eta^{\nu\rho} C^{\mu,\,\kappa\lambda} + \eta^{\nu\sigma} C^{\mu,\,\kappa\lambda} + \\
&+ \eta^{\kappa\sigma} C^{\mu\,\rho\lambda} - \eta^{\lambda\sigma} C^{\mu,\,\rho\kappa} - \eta^{\mu\kappa} C^{\lambda,\,\rho\sigma} + \eta^{\mu\lambda} C^{\kappa,\,\rho\sigma} \\
0 &= \eta^{\sigma\mu} C^{\rho,\,\nu} - \eta^{\rho\mu} C^{\sigma,\,\nu} - \eta^{\sigma\nu} C^{\rho,\,\mu} + \eta^{\rho\nu} C^{\sigma,\,\mu}.
\end{aligned}
\tag{F.0.4}
$$

These relations might seem slightly intimidating at first glance, but if we contract the last line with $\eta_{\sigma\mu}$, then

$$
\begin{aligned}
0 &= \eta_{\sigma\mu} \left(\eta^{\sigma\mu} C^{\rho,\,\nu} - \eta^{\rho\mu} C^{\sigma,\,\nu} - \eta^{\sigma\nu} C^{\rho,\,\mu} + \eta^{\rho\nu} C^{\sigma,\,\mu} \right) = \\
&= C^{\rho,\,\nu} - C^{\nu,\,\rho} - C^{\nu,\,\rho} + C^{\rho,\,\nu} \overset{\text{(F.0.2)}}{=} 4 \cdot C^{\rho,\,\nu} \Rightarrow C^{\rho,\,\nu} = 0.
\end{aligned}
\tag{F.0.5}
$$

For the other two coefficients on the first line and on the second line of equation (F.0.4), respectively, we have insufficient information to determine whether they are identically zero or not. Still, if we contract the two lines with $\eta_{\sigma\mu}$, this will yield

$$
\begin{aligned}
&C^{\rho\nu,\,\kappa\lambda} = \eta^{\lambda\rho} C^{\kappa\nu} - \eta^{\kappa\rho} C^{\lambda\nu} - \eta^{\nu\kappa} C^{\lambda,\rho} - \eta^{\lambda\nu} C^{\kappa\rho} \\
&\text{with } C^{\kappa\nu} \overset{def}{=} \frac{1}{2} \eta_{\sigma\mu} C^{\kappa\sigma,\,\nu\mu}
\end{aligned}
\tag{F.0.6}
$$

for the first line and for the second line the result is

$$
\begin{aligned}
&C^{\rho,\,\kappa\lambda} = \eta^{\rho\lambda} C^{\kappa} - \eta^{\rho\kappa} C^{\lambda} \\
&\text{with } C^{\kappa} \overset{def}{=} \frac{1}{3} \eta_{\mu\sigma} C^{\mu,\,\kappa\sigma}.
\end{aligned}
\tag{F.0.7}
$$

Reaching these expressions we see that we can redefine our operators $\hat{J}$ and $\hat{\boldsymbol{P}}$ to accommodate for these constants

$$
\begin{aligned}
\hat{P}'^{\rho} &\overset{def}{=} \hat{P}^{\rho} + C^{\rho} \\
\hat{J}'^{\rho\sigma} &\overset{def}{=} \hat{P}^{\rho\sigma} + C^{\rho\sigma}
\end{aligned}
\tag{F.0.8}
$$

and replacing the generators in (F.0.1) with the new ones and dropping the $'$:s afterwards will give the familiar commutation relations for the defining representation of the Poincaré algebra without central charges

$$
\begin{aligned}
i \left[\hat{J}^{\rho\sigma},\, \hat{J}^{\mu\nu} \right] &= \eta^{\sigma\mu} \hat{J}^{\rho\nu} - \eta^{\rho\mu} \hat{J}^{\sigma\nu} - \eta^{\nu\rho} \hat{J}^{\mu\sigma} + \eta^{\nu\sigma} \hat{J}^{\mu\rho} \\
i \left[\hat{P}^{\rho},\, \hat{J}^{\mu\nu} \right] &= \eta^{\rho\mu} \hat{P}^{\nu} - \eta^{\rho\nu} \hat{P}^{\mu} \\
\left[\hat{P}^{\rho},\, \hat{P}^{\mu} \right] &= 0.
\end{aligned}
\tag{F.0.9}
$$

Appendix G

Details on Equation (7.1.28)

Here we present the details behind the expansion of the comparator $U(x+\varepsilon n, x)$ in equation (7.1.28) which we restate here for convenience:

$$U(x+\varepsilon n, x) = \exp\left[-i\lambda\varepsilon A_\mu(x+\tfrac{\varepsilon}{2}n) + \mathcal{O}(\varepsilon^3)\right]. \tag{7.1.28}$$

Through the definition of A_μ in equation (7.1.19), the assumption that $U(y,x)$ is a pure phase and that $U(x,x)=1$ we can rewrite U as

$$U(y,x) = \mathrm{P}e^{-i\lambda\int_x^y A_\mu(x')\,\mathrm{d}x'^\mu}. \tag{G.0.1}$$

Note that this is a pathordered exponential as indicated by the path ordering operator P which for a path between x and y is defined as

$$\mathrm{P}e^{-i\lambda\int_x^y A_\mu(x')\,\mathrm{d}x'^\mu} = 1 - i\lambda\int_0^1 \mathrm{d}t A_\mu\dot{\gamma}\int_0^1 \mathrm{d}t + i^2\lambda^2\int_0^t \mathrm{d}t' A_\mu(t)\dot{\gamma}(t)A_\mu(t')\dot{\gamma}(t') + \dots \tag{G.0.2}$$

Using this form for the infinitesimal displacement we obtain

$$U(x+\varepsilon n, x) = \exp\left[-i\lambda\int_0^\varepsilon A_\mu(x+\xi n)\,\mathrm{d}\xi\right] = \exp\left[-i\lambda\int_{-\varepsilon/2}^{\varepsilon/2} A_\mu(x+\tfrac{\varepsilon}{2}n+\xi n)\,\mathrm{d}\xi\right] \tag{G.0.3}$$

and expanding A_μ in ξ yields

$$\begin{aligned}
&\exp\left[-i\lambda\int_{-\varepsilon/2}^{\varepsilon/2}\left(A_\mu(x+\tfrac{\varepsilon}{2}n) + \xi n^\nu\partial_\nu A_\mu(x+\tfrac{\varepsilon}{2}n) + \frac{\xi^2}{2}(n^\nu\partial_\nu)^2 A_\mu(x+\tfrac{\varepsilon}{2}n) + \dots\right)\mathrm{d}\xi\right] = \\
&\qquad = /\text{even integration interval}/ = \\
&\qquad = \exp\left[-i\lambda\left(\varepsilon A_\mu(x+\tfrac{\varepsilon}{2}n) + \frac{2}{6}\left(\frac{\varepsilon}{2}\right)^3(n^\nu\partial_\nu)^2 A_\mu(x+\tfrac{\varepsilon}{2}n)\right) + \mathcal{O}(\varepsilon^5)\right] = \\
&\qquad = \exp\left[-i\lambda\varepsilon A_\mu(x+\tfrac{\varepsilon}{2}n) + \mathcal{O}(\varepsilon^3)\right].
\end{aligned} \tag{G.0.4}$$

This is the expansion in equation (7.1.28).

Appendix H

Calculus of Variation and Least Action Principle

The calculus of variations is a very versatile tool and appears in many applications. The main usage is to find a function or curve which gives a stationary point, usually a minimum, of a quantity expressed as an integral, i.e. a functional. In this appendix we apply this technique to some unspecified problem involving the minimization of the quantity I in the most basic case. The possible generalizations to more variables and constraints are omitted as the relevant parts of these are pretty straightforward and should not cause any problem for the main part of this text. However, the reader is in general encouraged to study the subject more closely. As one of many good presentations [36] is recommended. But for now, suppose we want to find a function $y(x)$ which minimizes I expressed as

$$I = \int_{x_i}^{x_f} f(y, y', x) \, \mathrm{d}x \tag{H.0.1}$$

where $f(y, y', x)$ is a known function of y, its derivative $y' = \frac{\partial y}{\partial x}$ and the independent variable x. y is of course unknown since it is what we seek. To find the minimizing y we let it vary with a parameter α and we define $y(x, \alpha = 0)$ to be the function sought for. We can express this variation as

$$y(x, \alpha) = y(x, 0) + \alpha\eta(x) \tag{H.0.2}$$

where $\eta(x)$ is an arbitrary differentiable function with the requirements that it vanishes at the end points x_i, x_f. Now $I = I(\alpha)$ varies with α and to find a stationary point we derivate and set to zero as in elementary calculus,

$$\frac{\partial I}{\partial \alpha} = \int_{x_i}^{x_f} \left(\frac{\partial f(y, y', x)}{\partial y(x)} \frac{\partial y(x)}{\partial \alpha} + \frac{\partial f(y, y', x)}{\partial y'(x)} \frac{\partial y'(x)}{\partial \alpha} \right) \, \mathrm{d}x = 0. \tag{H.0.3}$$

By equation (H.0.2) we get

$$\frac{\partial I}{\partial \alpha} = \int_{x_i}^{x_f} \left(\frac{\partial f(y, y', x)}{\partial y(x)} \eta(x) + \frac{\partial f(y, y', x)}{\partial y'(x)} \frac{\partial \eta(x)}{\partial x} \right) \, \mathrm{d}x = 0 \tag{H.0.4}$$

and by integrating the second term by part we find

$$\int_{x_i}^{x_f} \frac{\partial f(y, y', x)}{\partial y'(x)} \frac{\partial \eta(x)}{\partial x} \, \mathrm{d}x = \left[\eta(x) \frac{\partial f(y, y', x)}{\partial y'(x)} \right]_{x_i}^{x_f} - \int_{x_i}^{x_f} \left(\frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f(y, y', x)}{\partial y'(x)} \right) \eta(x) \, \mathrm{d}x \tag{H.0.5}$$

where the first term is zero through the vanishing η at the end points. This partial integration is a standard procedure in these calculations. We are left with (H.0.4) in the form

$$\frac{\partial I}{\partial \alpha} = \int_{x_i}^{x_f} \left(\frac{\partial f(y, y', x)}{\partial y(x)} - \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f(y, y', x)}{\partial y'(x)} \right) \eta(x) \, \mathrm{d}x = 0 \tag{H.0.6}$$

and for this to vanish for any arbitrary function $\eta(x)$ we must have

$$\frac{\partial f(y, y', x)}{\partial y(x)} = \frac{\mathrm{d}}{\mathrm{d}x} \frac{\partial f(y, y', x)}{\partial y'(x)}. \tag{H.0.7}$$

This is called the Euler-Lagrange equation and gives the condition which our minimizing function $y(x)$ has to satisfy.

H.0.1 The Principle of Least Action

The calculus of variation has an important application in the analytic formulation of mechanics. Classically the action I is a functional of some chosen coordinates q_i of a mechanical system. It is defined as the time integral of the *Lagrangian* L, a scalar function of the coordinates, their time derivatives and the time,

$$I = \int_{t_i}^{t_f} L(q, \dot{q}, t) \, \mathrm{d}t \tag{H.0.8}$$

where the Lagrangian is the difference between the kinetic energy T and the potential energy V,

$$L(q, \dot{q}, t) = T(q, \dot{q}, t) - V(q, \dot{q}, t). \tag{H.0.9}$$

Varying the action with respect to the coordinates yields the equations of motion, found directly from the Euler-Lagrange equations, i.e. the coordinates describing the systems dynamics are the ones who minimize the action. This is the principle of least action. Quantum mechanics and relativity may use other Lagrangians (or Lagrangian densities also integrated over space) but the principle stays the same.

Appendix I

List of Symbols

The lists below contain the most commonly used symbols in this survey, which have been divided into the categories "Elementary Linear Algebra and Analysis", "Group Theory", "Representation Theory", "Special Relativity", "Gauge Theory" and "Quantum Mechanics". In addition to these symbols there is a list of general notation as well as a list of abbreviations and a list of physical constants. Moreover, since symbols tend to have different meanings in different contexts, a list of ambiguities has been placed after the last list of symbols.

Note that the symbols from section 7.2 and chapter 8 have not been included since these are relatively self-contained parts of this survey.

I.1 General Notation

Symbols	Descriptions
group, **subgroup** etc.	When a concept is defined it is written in upright bold font.
"$\eta = \eta_{\mu\nu}$"	The components of a given tensor are sometimes identified with the tensor itself.
"$O(1,n) = SO^+(1,3)$"	The *Lorentz groups* are the groups $O(1,n)$, but in this survey the "Lorentz group" *usually* means the subgroup $SO^+(1,3)$ of $O(1,n)$.
"$P(1,n) = ISO^+(1,3)$"	The *Poincaré groups* are the groups $P(1,n)$, but in this survey the "Poincaré group" *usually* means the subgroup $ISO^+(1,3)$ of $P(1,n)$.
"$\mathcal{G} = (\mathcal{G}, \star)$"	The underlying set of a group is *usually* identified with the group. This applies to representations of groups and Lie algebras as well.
"$c = \hbar = 1$"	The speed of light in free space and the reduced Planck constant are *usually* set equal to 1.

I.2 Abbreviations

Symbols	Descriptions
O	Orthogonal.

S	Special.
U	Unitary.
SO	Special orthogonal.
SU	Special unitary.
ISO	Isometry.
Sym	Symmetry.
Aut	Automorphism
LHS	Left hand side.
RHS	Right hand side.
P.i.	Partial integration.

I.3 Constants

Symbols	Descriptions
$\hbar = h/(2\pi)$	The reduced Planck constant (or Dirac constant), where h is Planck's constant.
i	The imaginary unit.
e	The elementary electric charge or the base of the natural logarithm.
$c = (\epsilon_0\mu_0)^{-1/2}$	The speed of light in free space.
ϵ_0	The permittivity of free space.
μ_0	The permeability of free space.

I.4 Elementary Linear Algebra and Analysis

Symbols	Descriptions
$\emptyset$	The empty set.
$\mathbb{N}$	The set of all natural numbers (not including 0), which is *not* a field.
$\mathbb{Z}$	The set of all integers, which is a ring but *not* a field.
$\mathbb{Z}_+$	The set of all strictly positive integers (i.e. $\mathbb{N}$), which is *not* a field.
$\mathbb{Z}_-$	The set of all nonpositive integers (including 0), which is *not* a field.
$\mathbb{Q}$	The field of all rational numbers.
$\mathbb{R}$	The field of all real numbers.
$\mathbb{C}$	The field of all complex numbers.
$\mathbb{R}^n$	The set of all ordered sets $(a_1, \ldots, a_n)$ of n real numbers, which are usually written as column vectors.
$\mathbb{C}^n$	The set of all ordered sets $(c_1, \ldots, c_n)$ of n complex numbers, which are usually written as column vectors.
$\mathbb{R}^{n\times n}$	The set of all $n \times n$ matrices with real entries.

$\mathbb{C}^{n\times n}$	The set of all $n \times n$ matrices with complex entries.
S^n	A *topological* n-dimensional hypersphere in $\mathbb{R}^{n+1}$, which is the same as a *geometrical* $(n+1)$-dimensional hypersphere. Note that topologists refer to the number of parameters which describes the sphere, while geometers refer to the dimension of $\mathbb{R}^{n+1}$.
V, W etc.	General vector spaces and other sets are *usually* denoted by upper case Latin letters.
$(a_1, a_2, \ldots, a_n)$	An ordered set.
$\{a_1, a_2, \ldots, a_n\}$	A (disordered) set.
a_n	Abbreviation for $(a_1, a_2, \ldots)$.
$\{a_n\}$	Abbreviation for $\{a_1, a_2, \ldots, a_n\}$.
$\mathrm{span}_{\mathbb{C}}\{t_1, t_2, \ldots, t_n\}$	The set of all complex linear combinations of the elements $t_1, t_2, \ldots, t_n$.
$V_\perp$	The orthogonal complement of a vector space V. Note that this space is usually denoted by $V^\perp$, but for practical reasons the symbol $V_\perp$ is used instead in this survey.
V^*	The *algebraic* dual space to a vector space V over a field.
$A \cup B$	The union of two sets A and B.
$A \cap B$	The intersection of two sets A and B.
$V \oplus W$	The direct sum of two vector spaces V and W.
$\bigoplus_{i=1}^{n} V_i$	Abbreviation for $V_1 \oplus V_2 \oplus \cdots \oplus V_n$.
$V \otimes W$	The tensor product (or *tensor direct product*) of two vector spaces V and W, which is spanned by the set $\{\boldsymbol{v}_i \otimes \boldsymbol{w}_j\}$, where $\{\boldsymbol{v}_i\}$ and $\{\boldsymbol{w}_j\}$ are arbitrary bases for V and W, respectively.
$\mathbb{C}^{\otimes n}$	Abbreviation for $\mathbb{C} \otimes \mathbb{C} \otimes \cdots \otimes \mathbb{C}$.
$W \subseteq V$	W is a subset of V and may be equal to V.
$W \subset V$	W is a subset of V but may not be equal to V.
$v \in V$	v is an element of a set V.
$v \notin V$	v is not an element of a set V.
$\dim(V)$	The dimension of a vector space V.
$f : A \to B$	A function (map) f from a set A to a set B.
$x \mapsto y$	An element x is mapped on an element y.
$\mathbb{1}$	An identity matrix.
R	A matrix for a rotation.
$R(\phi)$	A matrix for a rotation by the angle ϕ.
$R_l(\phi)$	A matrix for a rotation by the angle ϕ about the axis l through the origin. Note that in chapter 4 the same notation is used for the rotation itself.
$R(\boldsymbol{\theta})$	A matrix for a rotation by the angle θ about the axis through the origin with a direction vector $\boldsymbol{\theta}$.

ϕ, θ	Denotes angles.
A, B etc.	General matrices are usually denoted by upper case Latin letters.
A^{-1}	The inverse of a matrix A.
A^T	The transpose of a matrix A.
$A^\dagger$	The Hermitian adjoint (or *conjugate transpose*) of a matrix A.
e^A or $\exp(A)$	The exponential of a matrix A.
$\det(A)$	The determinant of a matrix A.
$\mathrm{Tr}(A)$	The trace of a matrix A.
A_{ij}	An element of a matrix A.
AB	Multiplication of a matrix A with a matrix B.
$A\boldsymbol{v}$	Multiplication of a matrix A with a vector $\boldsymbol{v}$.
$\hat{\boldsymbol{n}}, \hat{\boldsymbol{x}}$ etc.	Unit vectors are adorned with "hats".
$\boldsymbol{v}, \boldsymbol{w}$ etc.	General vectors in $\mathbb{R}^n$ are *usually* denoted by lower case Latin letters in bold font.
$\boldsymbol{v}^T$	The transpose of a vector $\boldsymbol{v}$.
$\boldsymbol{v}^\dagger$	The Hermitian adjoint of a vector $\boldsymbol{v}$.
$\boldsymbol{v} \cdot \boldsymbol{w}$	The ordinary scalar product of two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$.
$\boldsymbol{v}^2$	Abbreviation for $\boldsymbol{v} \cdot \boldsymbol{v} = \lvert\boldsymbol{v}\rvert^2$.
$\boldsymbol{v}_i$	En element of a vector $\boldsymbol{v}$.
z^*	The complex conjugate of a complex number z.
$\lvert z\rvert$	The absolute value of a real or complex number z.
ϵ	The Levi-Civita tensor (or *permutation symbol*).
δ	The Kronecker delta.
$\partial f/\partial_x$ or $\partial_x f$	The partial derivative of a function f of several variables with respect to the variable x.
$\mathrm{d}f/\,\mathrm{d}x$	The derivative of a function f of a single variable x.
$f(x)\lvert_{x=y}$	The value of $f(x)$ at the point $x = y$.
$\langle\cdot,\cdot\rangle$	Inner products and bilinear forms (note that the inner product on $\mathbb{R}^3$ is also a bilinear form, while Hermitian inner products are *antilinear*).
$[\cdot,\cdot]$	The ordinary commutator (and Lie algebras).
$\mathcal{O}$	Big ordo.
lim	Limes.
$\forall$	For all.
∞	Infinity.
$\propto$	Proportional to.

I.5 Group Theory

Symbols	Descriptions
$\mathbb{Z}_n$	Cyclic groups.

$\mathcal{D}_n$	Dihedral groups.
$GL(n, \mathbb{R})$	General linear groups with real entries.
$GL(n, \mathbb{C})$	General linear groups with complex entries.
$GL(V)$	The general linear group acting on a vector space V.
$SL(n, \mathbb{R})$	Special linear groups with real entries.
$SL(n, \mathbb{C})$	Special linear groups with complex entries.
$SL(V)$	The special linear group acting on a vector space V.
$O(n)$	Orthogonal groups.
$SO(n)$	Special orthogonal groups.
$U(n)$	Unitary groups.
$SU(n)$	Special unitary groups.
$O(n, k)$	The generalized orthogonal groups.
$O(1, n)$	The Lorentz groups.
$O(1, 3)$	The "ordinary" (i.e. three spatial dimensions) Lorentz group.
$SO(1, 3)$	The ordinary *proper* Lorentz group.
$SO^+(1, 3)$	The ordinary proper *orthochronous* Lorentz group.
$P(1, n)$	The Poincaré groups.
$P(1, 3)$	The "ordinary" (i.e. three spatial dimensions) Poincaré group.
$ISO(1, 3)$	The ordinary *proper* Poincaré group.
$ISO^+(1, 3)$	The ordinary proper *orthochronous* Poincaré group.
$Spin(1, 3)$	The double cover of $SO(1, 3)$.
$\mathrm{Aut}(\mathcal{G})$	The automorphism group of a group $(\mathcal{G}, \star)$.
$(\mathcal{G}, \star), (\mathcal{H}, \star)$ etc	General groups.
$\star$	A group operation. Such an operation is often called *composition*, *product operation* or simply *multiplication*.
$a \star g = ag$	The star is often omitted.
$\mathcal{G}, \mathcal{H}$ etc.	The underlying sets of groups, which are often identified with the groups themselves.
$\mathcal{G} \times \mathcal{H}$	The underlying set of the direct product of two groups $(\mathcal{G}, \star)$ and $(\mathcal{H}, \star)$.
$\mathcal{G} \rtimes_\varphi \mathcal{H}$	The underlying set of the semidirect product of two groups $(\mathcal{G}, \star_1)$ and $(\mathcal{H}, \star_2)$.
$\mathcal{G}/\mathcal{N}$	The underlying set of the quotient group of two groups $(\mathcal{G}, \star)$ and $(\mathcal{N}, \star)$, where $(\mathcal{N}, \star)$ is a normal subgroup of $(\mathcal{G}, \star)$. Note that $\mathcal{G}/\mathcal{N}$ is the set of all left cosets of $\mathcal{N}$ in $\mathcal{G}$.
$\rtimes$	Semidirect product.
$\mathcal{N} \lhd \mathcal{G}$	$(\mathcal{N}, \star)$ is normal subgroup of a group $(\mathcal{G}, \star)$.

$g\mathcal{H}, \mathcal{H}g$	Left and right coset of $\mathcal{H}$, respectively, where g is an element of the underlying set $\mathcal{G}$ of a group $(\mathcal{G}, \star)$ and $(\mathcal{H}, \star)$ is a subgroup of $(\mathcal{G}, \star)$.
C_a, C_b etc	Conjugacy classes.
$\|\mathcal{G}\|$	The order of a group $(\mathcal{G}, \star)$.
N/n	If $(\mathcal{G}, \star)$ is a group of order N with a subgroup $(\mathcal{H}, \star)$ of order n then N/n is the index of $(\mathcal{H}, \star)$ in $(\mathcal{G}, \star)$.
$< a >$, $< b >$ etc.	$< \cdot >$ is used to denote generators of groups.
e	The unit element of the underlying set of a group.
a, g etc.	General elements of the underlying sets of groups are denoted by lower case Latin letters.
a^{-1}	The inverse of an element a of the underlying set of a group.
$\tilde{g}_1 = g_2 g_1 g_2^{-1}$	If g_1 and g_2 are both elements of the underlying set of a group then $\tilde{g}_1$ is the conjugate of g_1 with respect to g_2.
$\mathfrak{sl}(n, \mathbb{C}), \mathfrak{su}(n)$ etc.	The underlying sets of the Lie algebras of $SL(n, \mathbb{C}), SU(n)$ etc.
$(\mathfrak{g}, [\cdot,\cdot]), (\mathfrak{h}, [\cdot,\cdot])$ etc.	General Lie algebras.
$\mathfrak{g}, \mathfrak{h}$ etc.	The underlying sets of Lie algebras are denoted in the same way as their corresponding Lie groups, except that lower case Gothic letters are used.
$\mathfrak{g}_{\mathbb{C}}$	The complexified equivalence of the underlying set $\mathfrak{g}$ of a Lie algebra $(\mathfrak{g}, [\cdot,\cdot])$.
X, Y etc.	General elements of the underlying sets of Lie algebras are *usually* denoted by upper case Latin letters, but there are plenty of exceptions in this survey.
$\mathfrak{C}$	A Cartan subalgebra.
A	Cartan matrix.
$\boldsymbol{\alpha}$	A root.
e_j, f_j	Root vectors.
$(\cdot,\cdot)$	A Cartan-Killing form.
$\cong$	Isomorphism.
$[\cdot,\cdot]$	A Lie bracket.

I.6 Representation Theory

Symbols	Descriptions
Π	The underlying set of a representation of a Lie group, which is often identified with the representation itself.

$\Pi(\mathcal{G})$	The underlying set of a representation of the Lie group $(\mathcal{G}, \star)$.
$\langle S \vert R \rangle$	A presentation of a group, where S is a set of generators for the group and R is a set of relations between the generators.
$\Pi_1 \oplus \Pi_2$	The direct sum of the underlying sets Π_1 and Π_2 of two representations of Lie groups.
$\Pi_1 \otimes \Pi_2$	The tensor product of the underlying sets Π_1 and Π_2 of two representations of Lie groups.
$\Pi(g)$	The element of Π corresponding to a Lie group element g.
Ad	The adjoint representation of a Lie Group.
3	The underlying set of the fundamental representation of $\mathfrak{sl}(3, \mathbb{C})$.
$\rho(g)$	If g is an element of the underlying set of a Lie group then $\rho(g)$ is a realization of g.
π	The underlying set of a representation of a Lie algebra, which is often identified with the representation itself.
$\pi(\mathfrak{g})$	The underlying set of a representation of the Lie algebra $(\mathfrak{g}, [\cdot,\cdot])$.
$\pi_1 \oplus \pi_2$	The direct sum of the underlying sets π_1 and π_2 of two representations of Lie algebras.
$\pi_1 \otimes \pi_2$	The tensor product of the underlying sets π_1 and π_2 of two representations of Lie algebras.
ad	The adjoint representation of a Lie algebra.
$\rho(X)$	If X is an element of the underlying set of a Lie algebra then $\rho(X)$ is a realization of X.

I.7 Special Relativity

Symbols	Descriptions
$\mathbb{R}^{1,3}$	Minkowski space (or *spacetime*).
$\boldsymbol{A}, \boldsymbol{B}$ etc.	General four-vectors are *usually* denoted by upper case Latin letters in bold font.
$\boldsymbol{A} \cdot \boldsymbol{B} = A_\mu B^\mu$	The inner product of two four-vectors $\boldsymbol{A}$ and $\boldsymbol{B}$. Note that the inner product on spacetime is not the same as the inner product on $\mathbb{R}^n$.
$\hat{\boldsymbol{P}}, \hat{J}$	The generators of the Poincaré group.
$\boldsymbol{P} = (E/c, p_x, p_y, p_z)$	A four-momentum vector, where E is the energy and $\boldsymbol{p} = (p_x, p_y, p_z)$ is the linear momentum vector.
$\boldsymbol{J} = (J_x, J_y, J_z)$	An angular momentum vector, which has only three components and despite its appearance is *not* a four-vector.
$\hat{\boldsymbol{J}} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)$	The operator for angular momentum.

$\hat{\boldsymbol{P}} = (\hat{H}, \hat{p}_x, \hat{p}_y, \hat{p}_z)$	The operator for four-momentum, where $\hat{H}$ is the Hamiltonian operator.
$\hat{\boldsymbol{K}} = (\hat{J}^{10}, \hat{J}^{20}, \hat{J}^{30})$	The operator for boosts (i.e. velocity changes).
$\boldsymbol{x}$	A position vector in spacetime.
$\boldsymbol{x}', \boldsymbol{x}''$ etc.	A position vector $\boldsymbol{x}$ in different inertial frames S', S'' etc.
$\boldsymbol{x}'^{\mu}$ or $\boldsymbol{x}^{\mu'}$	The components of a position vector $\boldsymbol{x}$ in an inertial frame S'.
$\boldsymbol{a}$	Translation vectors in $\mathbb{R}^3$ are denoted by $\boldsymbol{a}$.
Λ	The matrix of a Lorentz transformation.
$\eta_{\mu\nu}$	An element of the Minkowski metric tensor, which are often identified with η.
T	A general symmetry transformation.
T^{-1}	The inverse of a symmetry transformation T.
$T(\Lambda, \boldsymbol{a})$	A Poincaré transformation.
$\hat{U}$	A unitary operator acting on a Hilbert space.
$\hat{U}(\Lambda, \boldsymbol{a})$	The operator for a Poincaré transformation $T(\Lambda, \boldsymbol{a})$.
$A_{(\mu\nu)}, A_{[\mu\nu]}$	An element of the symmetric part and antisymmetric part, respectively, of a matrix A.

I.8 Gauge Theory

Symbols	Descriptions
$\boldsymbol{E}$	Electric field intensity.
V	Electric potential.
$\boldsymbol{B}$	Magnetic flux density.
$\boldsymbol{A}$	Magnetic vector potential.
A_μ	A component of a gauge field (and *not* a component of $\boldsymbol{A}$).
ρ	Total electric volume charge density.
$\boldsymbol{j}$	Total electric current density.
$\boldsymbol{J} = (\rho c, \boldsymbol{j})$	The four-current, where ρ and $\boldsymbol{j}$ are described above.
$\partial_\mu J^\mu$	The conserved current.
U	A comparator.
$\boldsymbol{U}$	A loop of comparators.
$F_{\mu\nu}$	A component of a field strength (or *curvature*).
I	The action.
$\mathcal{L}$	A Lagrangian (or *Lagrangian density*).
ϕ	A complex scalar field.
ϕ^*	The complex conjugate of a complex scalar field.
$f_a{}^{bc}$	A structure constant.
γ^μ	A Dirac matrix.
ψ	A Dirac spinor.

$\epsilon^{\alpha\beta\gamma\mu}$	A component of the totally antisymmetric tensor ϵ, which is the rank four equivalence of the rank three Levi-Civita tensor.
x	In chapter 7 x denotes the position vector in $\mathbb{R}^{1,3}$.
δ	A *functional* derivative.
$D_\mu f$	The covariant derivative of a scalar field f with respect to μ.
∇V	The gradient of a scalar field V.
$\nabla \cdot \boldsymbol{F}$	The divergence of a vector field $\boldsymbol{F}$.
$\nabla \times \boldsymbol{F}$	The curl of a vector field $\boldsymbol{F}$.
$\Box$	The d'Alembertian (or *wave*) operator.

I.9 Quantum Mechanics

Symbols	Descriptions
$\mathcal{H}$	A Hilbert space.
$\mathscr{R}$	A ray in a Hilbert space (i.e. a 1-dimensional subspace).
$\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$	The Pauli vector.
$\sigma_x, \sigma_y, \sigma_z$ or $\sigma_1, \sigma_2, \sigma_3$	The Pauli matrices.
$\boldsymbol{J} = (J_x, J_y, J_z)$	An angular momentum vector.
$\boldsymbol{p} = (p_x, p_y, p_z)$	A linear momentum vector.
$\boldsymbol{x} = (x, y, z)$	A position vector in $\mathbb{R}^3$.
$\hat{A}, \hat{\boldsymbol{b}}$ etc.	The operators for the physical quantities $A, \boldsymbol{b}$ etc.
A	The matrix of an operator $\hat{A}$.
$\hat{A}^\dagger$	The Hermitian adjoint of an operator $\hat{A}$.
$\exp(\hat{A})$	The exponential of an operator $\hat{A}$.
$\langle\psi\|\hat{A}\|\psi\rangle$	The expectation value for a physical quantity A given a state vector ψ.
ψ, Ψ	General state vectors.
$\hat{A}\psi$	An operator $\hat{A}$ acting on a state vector ψ.
$\hat{\mathbb{1}}$ or 1	The identity operator.
$\hat{H}$	The Hamiltonian operator.
$\hat{D}(\hat{\boldsymbol{n}}, \phi)$	A rotation operator, where the rotation is specified by the unit direction vector $\hat{\boldsymbol{n}}$ and angle ϕ.
$\hat{\mathscr{J}}$	An operator for spatial translations.
$\hat{\mathscr{U}}$	An operator for temporal translations.
$\hat{J}_+, \hat{J}_-$	Ladder operators for angular momentum.
$\hat{S}_+, \hat{S}_-$	Ladder operators for spin.
$\hat{\boldsymbol{J}} = (\hat{J}_x, \hat{J}_y, \hat{J}_z)$	The operator for angular momentum.
$\hat{\boldsymbol{p}} = (\hat{p}_x, \hat{p}_y, \hat{p}_z)$	The operator for linear momentum.
$\hat{\boldsymbol{J}}^2$	$\hat{J}_x^2 + \hat{J}_y^2 + \hat{J}_z^2$

$\hat{S}_x, \hat{S}_y, \hat{S}_x$	The operators for the spin's projection in the positive x-, y- and z-direction, respectively.
$\hat{A} \otimes \hat{B}$	The tensor product of two operators $\hat{A}$ and $\hat{B}$.
$\lvert j, m\rangle$	An eigenket of $\hat{\boldsymbol{J}}^2$ and $\hat{J}_z$ corresponding to the eigenvalues j and m, respectively.
$\lvert\cdot\rangle$	Kets.
$\langle\cdot\rvert$	Bras.
$\lvert\cdot\rangle\langle\cdot\rvert$	The outer product.
$\langle\cdot\rvert\cdot\rangle$	The standard Hermitian inner product on a Hilbert space $\mathcal{H}$ in Dirac notation.
$\uparrow$, $\downarrow$	Spin up and spin down, respectively, for a spin 1/2 system.
t	The letter t *usually* denotes time.

I.10 Ambiguities

Symbols	Descriptions
$\mathcal{H}$	In quantum mechanics $\mathcal{H}$ denotes a Hilbert space, while in group theory the same symbol usually denotes the underlying set of general group $(\mathcal{H}, \star)$.
$\boldsymbol{A}, A_\mu$	In chapter 7 $\boldsymbol{A}$ is the magnetic vector potential and A_μ is a component of a general gauge field.
$\boldsymbol{x}, x$	The position vector in spacetime is denoted by $\boldsymbol{x}$ in chapter 6 and by x in chapter 7.
$\hat{\boldsymbol{x}}, \hat{\boldsymbol{p}}$ etc.	Unit vectors in $\mathbb{R}^n$ and quantum mechanical operators are both denoted by "hats".
$R_l(\phi)$	This symbol denotes both a rotation and the matrix of the same rotation.
$[\cdot,\cdot]$	$[\cdot,\cdot]$ sometimes denotes an ordinary commutator and other times a Lie bracket, a general binary operator or a symmetric bilinear operator.
exp()	Natural, matrix or operator exponential function.
ψ	ψ usually denotes a general state vector in a Hilbert space, but in chapter 7 it denotes a Dirac spinor.
$\mathfrak{h}$	In general, $\mathfrak{h}$ denotes an ordinary Lie algebra, but in chapter 5 it denotes an *ideal*.
e	The unit element of a group, the base of the natural logarithm or the elementary electric charge.
ϵ	This symbol usually denotes the rank three Levi-Civita tensor, but in chapter 7 it denotes the rank four equivalence of the Levi-Civita tensor.

References

[1] N. Collins. Spectacular images from 'Big Bang' recreation. *The Daily Telegraph*, August 2011. Courtesy of CERN, P. Loiez. A collection of tracks left by subatomic particles in a bubble chamber, artistically enhanced, from the Big European Bubble Chamber (BEBC), CERN, 1990, `http://www.telegraph.co.uk/science/large-hadron-collider/8675126/Spectacular-images-from-Big-Bang-recreation.html`.

[2] P. Ramond. *Group Theory: A Physicist's Survey*. Cambridge University Press, 2010. ISBN 9780521896030.

[3] R.N. Cahn. *Semi-Simple Lie Algebras And Their Representations*. Dover Books on Mathematics. Dover Publications, 2006. ISBN 9780486449999.

[4] H. Georgi. *Lie Algebras in Particle Physics*. Frontiers in physics. Perseus Books, Advanced Book Program, 1999. ISBN 9780738202334.

[5] S. Weinberg. *The Quantum Theory Of Fields: Foundations*. Number v. 1 in The Quantum Theory of Fields. Cambridge University Press, 2005. ISBN 9780521670531.

[6] M.E. Peskin and D.V. Schroeder. *An Introduction To Quantum Field Theory*. Advanced Book Program. Addison-Wesley Publishing Company, 1995. ISBN 9780201503975.

[7] M. Nakahara. *Geometry, Topology and Physics*. Graduate Student Series in Physics. Institute of Physics Publishing, 2003. ISBN 9780750306065.

[8] H. Weyl. *Symmetry*. Princeton Science Library. Princeton University Press, 1952. ISBN 9780691023748.

[9] F. Gieres. About Symmetries in Physics. In F. Gieres, M. Kibler, C. Lucchesi, and O. Pigue, editors, *Symmetries in Physics*, pages 1–31, Paris, 1998. Proceedings of the fifth "Seminaire Rhodanien de Physique", Dolomieu 1997, Editions Frontières.

[10] R.P. Feynman, R.B. Leighton, and M.L. Sands. *The Feynman Lectures on Physics*. Number v. 1 in The Feynman Lectures on Physics. Addison-Wesley, 1965. Chapter 52, Symmetry in physical laws.

[11] B.C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Graduate Texts in Mathematics. Springer-Verlag, 2003. ISBN 9780387401225.

[12] D.C. Lay. *Linear Algebra and Its Applications*. Linear Algebra and Its Applications. Pearson/Addison-Wesley, 3rd edition, 2006.

[13] T. Rowland. Orthogonal Matrix, 2012. URL http://mathworld.wolfram.com/OrthogonalMatrix.html. From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein.

[14] J.J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Addison-Wesley, 2010. ISBN 9780805382914.

[15] J. M. Finn. *Classical Mechanics*. Jones and Bartlett Publishers, Inc., Sudbury, 2009.

[16] W. Greiner and B. Müller. *Quantum Mechanics: Symmetries*. Number v. 2 in Theoretical physics. Springer-Verlag, 1994. ISBN 9783540580805.

[17] M.A. Srednicki. *Quantum Field Theory*. Cambridge University Press, 2007. ISBN 9780521864497.

[18] S. Weinberg. *Gravitation and cosmology: principles and applications of the general theory of relativity*. Wiley, 1972. ISBN 9780471925675.

[19] M. Henningson. *Börja med Kvantfysik*. Institutionen för Fundamental fysik, Chalmers Tekniska Högskola, 2011.

[20] J. Maciejko. Representations of Lorentz and Poincaré Groups, September 2011. URL http://www.stanford.edu/~maciejko/groups.pdf.

[21] G.M. Tuynman and W.A.J.J. Weigerinck. Central Extensions and Physics. *Journal of Geometry and Physics*, 4(2):207 – 258, 1987. ISSN 0393-0440.

[22] V. Bargmann. On Unitary Ray Representations of Continuous Groups. *The Annals of Mathematics*, 59(1):pp. 1–46, 1954. ISSN 0003486X.

[23] M. Bertolini. Lectures on Supersymmetry, February 2012. URL http://people.sissa.it/~bertmat/susycourse.pdf. Lecture notes for the course "Supersymmetry" in the Elementary Particle Physics PhD-student curriculum at Scuola Internazionale Superiore di Studi Avanzati (SISSA), Trieste.

[24] A.J. Macfarlane. On the Restricted Lorentz Group and Groups Homomorphically Related to It. *Journal of Mathematical Physics*, 3(6):1116–1129, 1962.

[25] S. Coleman and J. Mandula. All Possible Symmetries of the S-Matrix. *Phys. Rev.*, 159: 1251–1256, 1967.

[26] S. Weinberg. *The Quantum Theory of Fields: Supersymmetry*. Number v. 3 in The Quantum Theory of Fields. Cambridge University Press, 2000. ISBN 9780521660006.

[27] R.A. Hulse. Hulse's Nobel Lecture, December 1993. URL http://www.nobelprize.org/nobel_prizes/physics/laureates/1993/hulse-lecture.pdf.

[28] B. R. Martin. *Nuclear and Particle Physics – An Introduction*. Wiley-Blackwell, Chichester, 2nd edition, 2009.

[29] K. Nakamura and others (Particle Data Group). 2011 Review of Particle Physics. *Journal of Physics G: Nuclear and Particle Physics*, 37(075021), 2010. and 2011 partial update for the 2012 edition.

[30] J. Baez and J. Huerta. The Algebra of Grand Unified Theories. *Bulletin of the American Mathematical Society*, 47(3):483–552, 2010.

[31] Y. Hideki. Yukawa's Nobel Lecture, December 1949. URL `http://www.nobelprize.org/nobel_prizes/physics/laureates/1949/yukawa-lecture.pdf`.

[32] F. Close et al. *The Particle Odyssey – A Journey to the Heart of Matter*. Oxford University Press, Oxford, 2004.

[33] M. F. L'Annunziata. *Radioactivity – Introduction and History*. Elsevier, Amsterdam, 2007.

[34] W. Rindler. *Relativity – Special, General, and Cosmological*. Oxford University Press, Oxford, 2nd edition, 2006.

[35] B. C. Hall. *Lie Groups, Lie Algebras, and Representations – An Elementary Introduction*. Springer-Verlag, New York, 2004.

[36] G. Arfken and H.J. Weber. *Mathematical Methods for Physicists*. Academic Press, 2003. ISBN 9780120598762.

Index