A monster tale: a review on Borcherds' proof of monstrous moonshine conjecture

by

Alan Gerardo Reyes Figueroa

Instituto de Matemática Pura e Aplicada

IMPA

A monster tale: a review on Borcherds' proof of monstrous moonshine conjecture

by

Alan Gerardo Reyes Figueroa

Dissertação

Presented in partial fulfillment of

the requirements for the degree of

Mestre em Matemática

Instituto de Matemática Pura e Aplicada

IMPA

Maio de 2010

A monster tale:
a review on Borcherds' proof of
monstrous moonshine conjecture

Alan Gerardo Reyes Figueroa

Instituto de Matemática Pura e Aplicada

APPROVED: 26.05.2010

Hossein Movasati, Ph.D.

Henrique Bursztyn, Ph.D.

Amílcar Pacheco, Ph.D.

Frédéric Paugam, Ph.D.

Advisor:
Hossein Movasati, Ph.D.

Abstract

The Monster $\mathbb{M}$ is the largest of the sporadic simple groups. In 1979 Conway and Norton published the remarkable paper *'Monstrous Moonshine'* [38], proposing a completely unexpected relationship between finite simple groups and modular functions, in which related the Monster to the theory of modular forms. Conway and Norton conjectured in this paper that there is a close connection between the conjugacy classes of the Monster and the action of certain subgroups of $SL_2(\mathbb{R})$ on the upper half plane $\mathbb{H}$. This conjecture implies that extensive information on the representations of the Monster is contained in the classical picture describing the action of $SL_2(\mathbb{R})$ on the upper half plane. **Monstrous Moonshine** is the collection of questions (and few answers) that these observations had directly inspired.

In 1988, the book *'Vertex Operator Algebras and the Monster'* [67] by Frenkel, Lepowsky and Meurman appeared. This book gave an explanation of why the Monster is related to the theory of modular forms, by showing that it acts as an automorphism group of a vertex algebra $V^\natural$ of central charge 24, constructed in terms of the Leech lattice. Vertex operators also arise in mathematical physics contexts, such as conformal field theory and string theory. Although Frenkel *et al.* did not prove the Conway-Norton conjecture, they provided the raw material with which this conjecture could be approached. The main conjecture was eventually proved by Borcherds in 1992 in his paper *'Monstrous moonshine and monstrous Lie superalgebras'* [12]. A key role in the proof was played by a class of infinite dimensional Lie algebras —called by Borcherds— generalized Kac-Moody algebras. Borcherds proved properties of the monster Lie algebra which turned out to be sufficient to complete the Conway-Norton conjecture. A key step in the proof was an application of the *'no-ghost'* theorem from string theory.

Borcherds also obtained many other remarkable connections between sporadic finite simple groups and modular forms in this and other papers. He was awarded a Fields medal in 1998 for his contributions.

The principal interest of Moonshine is that it constitutes a new bridge between algebraic structures and modular apparatus, and a new era of collaboration between mathematics and physics. This work is a modest review of the Moonshine phenomena: the main conjectures, Borcherds' proof, open problems and actual areas of research.

Resumo

O Monstro $\mathbb{M}$ é o maior dos grupos simples esporádicos. Em 1979 Conway e Norton publicaram um remarcável artigo *'Monstrous Moonshine'* [38], onde proporam uma relação inesperada entre grupos finitos simples e funções modulares, na qual relacionaram o Monstro à teoria de formas modulares. Conway e Norton conjeturaram neste artigo que existe uma forte conexão entre as classes de conjugação do Monstro e a ação de certos subgrupos de $SL_2(\mathbb{R})$ no semi-plano superior $\mathbb{H}$. Tal conjetura implica que muita informação das representações do Monstro está contida na imagem que descreve a ação de $SL_2(\mathbb{R})$ no semi-plano superior. **Monstrous Moonshine** é a coleção de questões (e umas poucas respostas) diretamente inspiradas por essas observações.

Em 1988, apareceu o livro *'Vertex Operator Algebras and the Monster'* [67] de Frenkel, Lepowsky and Meurman. Este livro deu uma explicação de por que o Monstro está relacionado com a teoria de formas modulares, mostrando que $\mathbb{M}$ age como um grupo de automorfismos de uma álgebra de vêrtices $V^\natural$ de carga central 24, construida em termos do retículo de Leech. Os operadores vêrtice ocorrem também em contextos da física matemática, tais como teoria de campos conformes e teoria das cordas. Embora Frenkel *et al.* não provaram a conjetura de Conway-Norton, eles proporcionaram a ferramenta básica com a qual esta conjetura pode-se provar. A conjetura principal eventualmente foi provada por Borcherds em 1992 no seu artigo *'Monstrous moonshine and monstrous Lie superalgebras'* [12]. Um papel chave na prova é feito por uma classe de álgebras de Lie infinito dimensionales —chamadas por Borcherds— de álgebras de Kac-Moody geralizadas. Borcherds provou propriedades da álgebra de Lie monstro que foram suficentes para completar a conjetura Conway-Norton. Um papel fundamental na prova foi uma aplicação do teorema *'no-ghost'* da teoria das cordas.

Borcherds obteve também muitas outras conexões entre grupos simples esporádicos e formas modulares em este e outros artigos. Ele foi galardoado com a medalha Fields em 1998 pelas suas contribuções.

O interés principal de Moonshine é que ele estabelece uma nova ponte entre as estruturas algébricas e as modulares, e começa uma nova era de cooperação entre a matemática e a física. Esta dissertação é uma revisão somera do fenómeno Moonshine: as conjeturas principais, a prova de Borcherds, problemas abertos e áreas atuais de pesquisa.

Preface

Monstrous Moonshine is probably one of the most esoteric achievements arising in mathematics. The fact that the Monster has connections to other parts of mathematics shows that there is something very deep going on here. No one fully understands it, and the links to the fields of physics and geometry are tantalizing. The Moonshine connections have spawned a lot of work by a several mathematicians and mathematical physicists recently, and have opened intriguing conjectures, most of them remaining open until today.

This work grew out in a attempt to explain the mysteries about the Monster and its relation with number theory, specially the j modular invariant. The idea about work with this topic was given to me by Prof. Hossein Movasati, with the purpose to have a better understand on some connections occurring between automorphic functions, geometry and physics. The original idea was to write a complete and detailed proof of the Moonshines conjectures, but by obviously reasons, a complete proof was somewhat voluminous to be given here. I have decided to restrict attention to the original Conway-Norton conjecture. Other aspects of Moonshine are mentioned only as a matter of general culture on the subject.

I have written with special attention to the non initiated. In fact, I have included so many theory on the first chapters in order to give sufficient background to understand all material in the later ones, and to make easier most of the ideas of Borcherds' proof. For the sake of background material, the proofs of theorems and propositions are omitted. The idea is to give only the proof of the Conway-Norton conjecture, making the material accessible to the level of a second year graduate student. Unfortunately, this work is by no means complete. There are several references included, with the purpose of encourage readers to specialize topics of their interest. Perhaps, the recent [72] is the most complete review of the Moonshine phenomenon. A good resume of [72] is the paper [71]. For the non-mathematician reader, [160] is good source, basically for its divulgation style. Main differences between this work and [72, 71] is the devoted attention to the number theory of moonshine, and the detailed exposition of the proof of the Conway-Norton conjecture.

In Chapter 1 I present an historical introduction, in order to facilitate a quickly access to the main conjecture. Several concepts appear, most of them explained more detailed in subsequent chapters.
Chapters 2 and 3 give background material on algebra, particularly on Lie algebras. Chapter 2 also includes other classic material on algebra, such as representation of finite groups, modules, tensor products and some constructions. The main topic of Chapter 2 is to explain affine Lie algebras (Section 2.8), which appear later. Chapter 3 introduce important concepts in Lie algebra theory, such as the Cartan subalgebras, the root systems and the

Weyl group. Main topics of this chapter are the Theorem of Highest Weights (Section 3.7) and the Weyl Character formula (Section 3.8).

Chapter 4 introduces the concept of vertex operator algebras, and relate it to the representation theory of Kac-Moody algebras, including a bit of lattice theory. At the final of Chapter 4, I give some interesting data relating Moonshine to the E_8 classical Lie algebra and the Leech lattice.

Chapter 5 is devoted exclusively to a partial construction of the Moonshine module $V^\natural$ given by Frenkel-Lepowsky-Meurman. It also includes some aspects of other 'unusual' algebraic structures, such as the Golay code $\mathscr{C}_{24}$, the Leech lattice Λ_{24} and the Griess algebra $\mathscr{B}$.

In Chapters 6 and Chapter 7, I explain in more detail the j-function and the original Conway-Norton conjecture. This is the only chapter where the proofs are given (mainly due to the fact that my principal area is number theory, and I have decided give more attention to this). Chapter 7 includes important useful material to understand the ideas of Borcherds' proof: congruence subgroups of $SL_2(\mathbb{R})$, replicable functions and the so called replication formulae.

Finally, Chapter 8 is devoted to the proof of the Conway-Norton conjecture given by Borcherds (here is where a lot of concepts introduced in previous chapters main their contribution). I introduce here the Borcherds Lie algebras (Section 8.1) and subsequently construct the Monster Lie algebra $\mathbb{M}$ (Section 8.3), and the denominator formulas (Section 8.4). At the end, the proof is concluded by using properties of the replication formulae.

Chapter 9 presents the actual status of Moonshine. It includes some generalizations of the Moonshine conjectures, and explores deep connections between Moonshine, number theory, geometry and physics. I have added a lot of references to facilitate future work. Most of the material of this chapter is taken from [72].

The order of the chapters is by no means strict. In fact, most of the material can be omitted in a rapid lecture. For example, the advanced reader may omit Chapters 2 and 3. Similarly, reader not interested in number theory can omit Chapter 6. A quickly reading of Borcherds proof could be:

Chapter 1; Sections 4.1, 4.3, 4.5, 5.3; Sections 7.3, 7.4, 7.5; and Chapter 8

Obviously the final chapter is interesting by its own.

Acknowledgments

I would like to express my deep-felt gratitude to Hossein Movasati, for his advice, encouragement, enduring patience and constant support. He was never ceasing in his belief in me, always providing clear explanations, expert in indicate the simplest way to do the job, constantly driving me with energy and always, *always*, giving me his time, in spite of anything else that was going on. I wish all students the honor and opportunity to experience

his ability to perform at that job.

I also wish to thank the other members of my committee, Prof. Henrique Bursztyn from Impa, Prof. Amilcar Pacheco from UFRJ, and Prof. Frédéric Paugam from Université Pierre et Marie Curie - Paris VI. Their suggestions, comments and additional guidance were invaluable to the completion of this work. Professor Don Zagier, from Max Plank Institute of Mathematics, give some useful references about equation (7.16) and made important comments about future work, so I am in debt with him. Also, Professor Michel Waldschmidt from Université Pierre et Marie Curie - Paris VI, made additional useful comments. I also would tank to Professor Alfredo Iusem. He was extremely helpful and provided all the additional support I needed when things were going not so well.

Additionally, I want to thank all my colleagues: Altemar, Antonio, Arturo, Bruno, Cadú, Carlos, Carol, Cristiane, Daniel, Diogo, Guillermo, José Vítor, Juan, Leonardo, Lucas, Marco, Mauricio, Phillip, Rafael, Ricardo, Susana and Tiane. They were always expressing best wishes and advices. In particular, I appreciate the conversations (and e-mails) with Arturo, Karla and Mabel, who were always giving their moral support. I must also thank to Juan, Juan Pablo, Cristian, Mauricio, Pedro and Sergio by their constant advices, and to Marlon for his motivation, constant interest and help in the progress of my work. Also, I must thank Angel, by some suggestions on future work. Thanks to Alvaro for his comments and references in the mathematical physics topics. I would like to thank also María Jose for her encouragement and her valuable indications about how to use the PsTricks package for the drawings. María Jose and Carol also give some helpful tips in how to use Beamer. As a special note, I would like to express my gratitude to Altemar, who's patient and helpful indications conduced me to complete the proof of the replication formulae.

Finally, I must thank Impa professors and staff for all their hard work and dedication, providing me the means to complete this degree and prepare for a career. I express my gratitude to CNPq and Impa for their financial support.

A. G. Reyes Figueroa
Rio de Janeiro, maio de 2010.

Table of Contents

Appendix

List of Tables

List of Figures

Chapter 1

A historical crash course on Monstrous Moonshine

1.1 Introduction

In 1978, John McKay made the observation that

$$196,884 = 196,883 + 1. \tag{1.1}$$

Here, the number 196,884 refers to the first nontrivial coefficient of the automorphic form or normalized j-function

$$J(z) = q^{-1} + 196,884q + 21,493,760q^2 + \dots$$

associated to the modular group $SL_2(\mathbb{Z})$, that appears in number theory. On the other hand, the number 196,883 refers to the dimension of the minimal faithful representation of the Monster group.

The central question of McKay's equation (1.1) is: What does the j-function (the left side) have to do with the Monster finite group (the right side)? In general, specialists on the subject agree that we still do not understand completely this phenomenon. Today we say that there exists a vertex operator algebra, called the *Moonshine module* $V^\natural$, which interpolates between left and right sides of (1.1): its automorphisms group is the Monster and its graded dimension is the normalized j-function. Since the original conjecture appeared, the Moonshine question triggered new developments in theory and pushed out the boundary of mathematical knowledge, bringing with itself more questions than answers. This thesis work consists mainly in a modest review of the original problem and the main proof of Moonshine conjecture, a briefly look of how this theme is related to other areas of mathematics, and what are the unsolved questions that remains until today.

The purpose of this chapter is to give a quickly understanding of the original Moonshine conjecture. Because Moonshine is a subject that mixes several branches of mathematics and physics, the reader is warned that this work is not self-contained and requires extensive mathematical baggage. There are several references included, some of them by historical reasons, others only with the intention of conduce the interested reader to specialize particular topics.

1.2 Finite simple groups

The discovery of the Monster was preceded by a long history of development of the theory of finite groups. It has been of interest to ask for the classification of all finite groups, yielding the enumeration of all kind of symmetries. Remember that a group G is called *simple* if it has no non trivial normal subgroups. Basically, the importance of finite simple groups is that by Jordan-Hölder theorem, they constitute the *'primes'* of all finite groups, that is, they are the constructive blocks of all finite groups. Thus, the core of the problem is the classification of finite simple groups.

This classification was announced in 1981 [77]. This result was the culmination of an intense effort involving several hundred of mathematicians over a period of some 25 years. The complete classification theorem requires between 9,000 and 10,000 pages of mathematical work, although a more streamlined proof is now on progress. By the end of the nineteenth century, thanks to works of Jordan, Dickson, Chevalley and others, several families of simple groups were known. In addition, by 1861, Mathieu discovered five strange finite simple groups [141]. This groups were first called *sporadic* in the book of Burnside [23]. Most of the finite simple groups are now called of Lie type or Chevalley groups, and admit a uniform construction in terms of simple Lie algebras, via a systematic treatment discovered in [27].

The modern classification race started with the paper of Feit and Thompson in 1963 [61], showing that every non-cyclic finite simple group has even order. This gave the greatest impetus to the effort to classify the finite simple groups. Thompson followed this up by another lengthy paper in which he classified all the minimal simple groups. This papers made feasible the classification project. In 1972, Gorenstein proposed a strategy for the classification involving a detailed 16 point programme. Progress was rapid from this stage, with Gorenstein and Aschbacher playing leading roles in the project. It was finally shown in 1981 that every finite simple group is isomorphic to one of the following:

- A cyclic group $\mathbb{Z}_p$ of prime order p.
- An alternating group Alt_n for $n \geq 5$.
- A simple group of Lie type over a finite field, *e. g.*, $PSL_n(\mathbb{F}_q)$.
- Some one of the 26 sporadic simple groups (see Table 1.1).

There are 26 sporadic simple groups, not definitively organized by any simple theme. Further details about these groups and their classification programme can be found in [77].

The largest of the sporadic simple groups is called *the Monster*. We shall denote this group by $\mathbb{M}$. The Monster contains among its subquotients twenty of the sporadic simple

Group	Order	Name/Discoverer
M_{11}	$2^4\cdot 3^2\cdot 5\cdot 11$	Mathieu
M_{12}	$2^6\cdot 3^3\cdot 5\cdot 11$	Mathieu
M_{22}	$2^7\cdot 3^2\cdot 5\cdot 7\cdot 11$	Mathieu
M_{23}	$2^7\cdot 3^2\cdot 5\cdot 7\cdot 11\cdot 23$	Mathieu
M_{24}	$2^{10}\cdot 3^3\cdot 5\cdot 7\cdot 11\cdot 23$	Mathieu
J_1	$2^3\cdot 3\cdot 5\cdot 7\cdot 11\cdot 19$	Janko
$\mathrm{J}_2 = \mathrm{HJ}$	$2^7\cdot 3^3\cdot 5^2\cdot 7$	Hall-Janko
J_3	$2^7\cdot 3^5\cdot 5\cdot 17\cdot 19$	Higman-Janko-McKay
HS	$2^9\cdot 3^2\cdot 5^3\cdot 7\cdot 11$	Higman-Sims
McL	$2^7\cdot 3^6\cdot 5^3\cdot 7\cdot 11$	McLaughlin
Suz	$2^{13}\cdot 3^7\cdot 5^2\cdot 7\cdot 11\cdot 13$	Suzuki
He	$2^{10}\cdot 3^3\cdot 5^2\cdot 7^3\cdot 17$	Held
Ru	$2^{14}\cdot 3^3\cdot 5^3\cdot 7\cdot 13\cdot 29$	Rudvalis
Co_1	$2^{21}\cdot 3^9\cdot 5^4\cdot 7^2\cdot 11\cdot 13\cdot 23$	Conway
Co_2	$2^{18}\cdot 3^6\cdot 5^3\cdot 7\cdot 11\cdot 23$	Conway
Co_3	$2^{10}\cdot 3^7\cdot 5^3\cdot 7\cdot 11\cdot 23$	Conway
Fi_{22}	$2^{17}\cdot 3^9\cdot 5^2\cdot 7\cdot 11\cdot 23$	Fischer
Fi_{23}	$2^{18}\cdot 3^{13}\cdot 5^2\cdot 7\cdot 11\cdot 13\cdot 17\cdot 23$	Fischer
$\mathrm{Fi'}_{24}$	$2^{21}\cdot 3^{16}\cdot 5^2\cdot 7^3\cdot 11\cdot 13\cdot 17\cdot 23\cdot 29$	Fischer
O'N	$2^9\cdot 3^4\cdot 5\cdot 7^3\cdot 11\cdot 19\cdot 31$	O'Nahn
Ly	$2^8\cdot 3^7\cdot 5^6\cdot 7\cdot 11\cdot 31\cdot 37\cdot 67$	Lyons
J_4	$2^{21}\cdot 3^3\cdot 5\cdot 7\cdot 11^3\cdot 23\cdot 29\cdot 31\cdot 37\cdot 43$	Janko
HN	$2^{14}\cdot 3^6\cdot 5^6\cdot 7\cdot 11\cdot 19$	Harada-Norton
Th	$2^{15}\cdot 3^{10}\cdot 5^3\cdot 7^2\cdot 13\cdot 19\cdot 31$	Thompson
$\mathbb{B}$	$2^{41}\cdot 3^{13}\cdot 5^6\cdot 7^2\cdot 11\cdot 13\cdot 17\cdot 19\cdot 23\cdot 31\cdot 47$	Baby Monster
$\mathbb{M}$	$2^{46}\cdot 3^{20}\cdot 5^9\cdot 7^6\cdot 11^2\cdot 13^3\cdot 17\cdot 19\cdot 23\cdot 29\cdot 31\cdot 41\cdot 47\cdot 59\cdot 71$	The Monster

Table 1.1: All 26 finite sporadic simple groups.

groups, except for J_3, Ru, O'N, Ly and J_4. These twenty constitute the *Happy Family*, and they occur naturally in three generations: the family of Mathieu groups, the family of Conway groups, and the so called Third Generation. Each one of these is related to an algebraic structure, in particular, the last one is associated to the Monster. The other five sporadic groups are called the *pariah*. Fischer and Griess independently predicted in 1973 the existence and properties of $\mathbb{M}$ as the largest of the sporadic groups. It has 194 conjugacy classes and irreducible characters. The character table of $\mathbb{M}$ was determined by Fischer, Livingstone and Thorne in 1978 [62] and can be found in the Atlas of Finite Groups [36]. The degree of the smallest irreducible characters of $\mathbb{M}$ are:

$$d_0 = 1, \quad d_1 = 196883, \quad d_2 = 21296876, \quad d_3 = 842609326, \ \ldots$$

In particular, the minimal faithful representation of the Monster would have dimension

$$d_1 = 196,883. \tag{1.2}$$

A lot of information about $\mathbb{M}$ can be found in [36].

As a remarkable fact, we mention that the greatest of the Mathieu sporadic groups, M_{24} was constructed by Mathieu as the group of symmetries of the Golay code $\mathscr{C}_{24}$ (a 12-dimensional subspace of a 24-dimensional vector space over $\mathbb{F}_2$). Similarly, Conway constructed his 3 sporadic groups in 1968, and the greater of them, Co_3, is realized as the automorphism group of the Leech lattice Λ_{24} (a 24-dimensional subspace of the 26-dimensional Lorentzian vector space $\mathbb{R}^{25,1}$, with norm $||x||^2 = x_1^2 + \ldots + x_{25}^2 - x_{26}^2$). Norton observed that the minimal representation of $\mathbb{M}$ would have the structure of a real commutative non-associative algebra with an associative form. Finally, the Monster group was first constructed by Griess in 1980 as a group of automorphisms of a commutative non-associative algebra of dimension 196,883 over $\mathbb{Q}$ with an associative form [80]. Griess' construction has been simplified in works of Tits [174, 175, 176] and Conway [34, 35]; and Tits has in fact proved that $\mathbb{M}$ is the full group of automorphisms of the Griess algebra $\mathscr{B}$. Unfortunately, even this fine version of Griess algebra does not appear as elegant as Golay code $\mathscr{C}_{24}$ or Leech lattice Λ_{24}, which have simple characterizations. But, there was some hints that the Monster could be associated with an elegant canonical structure. We shall discuss this structure in Chapter 5.

1.3 The discovery of $\mathbb{M}$

We have already mention that Feit and Thompson's paper [61] was the starting point on the systematic search of sporadic groups, or sometimes called *symmetry atoms*, leading to a whole project known as 'the Classification' [77]. The idea: to compiling a list of all finite symmetry atoms, and showing that the list was complete (like a periodic table of symmetries). Recall that there were already know some of this exceptional the sporadic simple groups. Conway constructed his three sporadic groups Co_1, Co_2 and Co_3 in 1968, by looking various mirror symmetries occurring within Leech lattice Λ_{24}, and this lattice had yielded a total of 12 sporadic groups, nine of which had already appeared elsewhere. These 12 groups, along with Janko's groups J_1 and J_3 —which had nothing to do with the Leech lattice— brought the total number of exceptions to 14. By the end of 1972 there were six more: three found by Fischer; one by Dieter Held; one by Richard Lyons; and one by Arunas Rudvalis. Both Held and Lyons used the cross-section method. Rudvalis used permutations. The total number of exceptions was now 20. With so much activity and so much information coming in, it made excellent sense to collect it all, correct errors, and present it in a form that was easy to read and readily available. This was 'the Atlas' [36] (exceptions to periodic table).

In fact, Fischer constructed his sporadic groups Fi_{22}, Fi_{23} and $\mathrm{Fi'}_{24}$ by using the method of transpositions (adding some new symmetries to the Mathieu groups M_{22}, M_{23}, M_{24} respectively), leading him to the discovery of the Baby Monster $\mathbb{B}$ in 1973. Another new group of symmetries was found by Michael O'Nan the same year. This brought the total to 22, though not all these groups were yet known to exist. After that, it was expected that no other new sporadics would appear. Fortunately, Fischer and Griess noticed independently that this Fischer's huge new group could appear as a cross-section in a larger group. If a group emerged from the cross-section method —like $\mathbb{B}$— then a great deal of information needed to be calculated before a construction was possible. Most of this data was encoded in the form of a square array of numbers called a *character table* (a character table is a square array of numbers that that express the fundamentally ways the group can operate in multidimensional space) (see Section 2.1). In fact, a finite group can have a huge character table, but sporadic groups usually have a low number of rows/columns in their character tables.

In late 1973, knowing only that the Monster had two cross-sections, using a procedure called *Thompson order formula*, the size of the whole thing was within reach. Thompson's technique needed detailed computations on how the two cross-sections could intersect. Fischer used it to show that the size could not be greater than a certain number. Further calculations made by Conway, led Fischer to four new sporadic groups: $\mathbb{B}$, Th, HN, and $\mathbb{M}$, the last one with size

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$$

Having the size of the Monster was essential before working out its character table. Norton and others established that the Monster would have a representation of dimension $196,883 = 47 \cdot 59 \cdot 71$. Only with this information, Fischer, Livigstone and Thorne computed the entire character table of $\mathbb{M}$, in 1974; and then that of the Baby Monster. Finally, Sims and Leon conclude the construction of $\mathbb{B}$ in 1977, by creating it on a computer as a group of permutations on 13,571,955,000 mirrors.

After Sims and Leon construction, it was natural to ask whether the Monster could be constructed in a similar way. Unfortunately this seemed to be out of sight, so an alternative method was needed. Perhaps one could use multidimensional space. Similar methods had been applied to other exceptional groups, such as Janko's first group J_1. But, as J_1 needed seven dimensions, the Monster needed 196,883, which means that a single operation in the Monster would appear as a matrix with nearly 196,883 rows and as many columns.

Suddenly in 1980, Bob Griess announced a construction. Already, Norton had figure out that the Monster must preserve an algebra structure in 196,884 dimensions. This structure would allow any two points to be multiplied together to give a third point. Griess' first task was to construct a suitable multiplication, solving some problems about signs on the algebra structure, by tracking them back into the group. This group Griess is referring to here is a huge subgroup of the Monster, extended by a factor of over 32 million. This group needs 96,308 dimensions, and is one of the two cross-sections of the Monster (the other one involves the Baby Monster). Fischer used it earlier in helping to build the character table of the Monster, and Griess now used it to help to construct the Monster in 196,884 dimensions. He knew that the action of this huge subgroup must split the space into three subspaces of the following dimensions (see (5.19) and (5.20)):

$$98,304 + 300 + 98,280 = 196,884$$

The first number is $98,304 = 2^{12} \times 24$. This is the space needed for the cross-section mentioned above. The second number is $300 = 24 + 23 + 22 + ... + 3 + 2 + 1$. This comes from a triangular arrangement of numbers with 24 in the first row, 23 in the second row, 22 in the third row, and so on. The third number is $98,280 = \frac{1}{2}(196,560)$. This comes from the Leech lattice (see Chapter 5), where there are 196,560 points closest to a given point, and they come in 98,280 diametrically opposite pairs. Each pair yields an axis through the given point, and in the 196,884-dimensional space these axes become independent of one another. Griess [80] finally sort out the sign problem for the multiplication, and proved that the group of symmetries contains the Monster, by adding one extra permutation to create a larger group.

Two other mathematicians got deeply involved in looking at the Griess construction. One was Jacques Tits, who found a way of avoiding the sign problems [175]. Tits also found a number of other improvements and simplified Griess's construction. The other person who took a detailed interest in the Griess construction was Conway, giving a construction of his own, and a very elegant proof of finiteness [34, 35]. In a sense, Tits avoided all calculations in the Monster, while Conway made easier to perform such calculations. Conway's construction is similar to Griess's in the sense that they both use the same large cross-section of the Monster to get started. This splits the 196,884-dimensional space into three subspaces. A more detailed (with a lot of historical notes) on the construction of the sporadic groups can be found in [160]. The interested reader may also see [72] to have a complete context on the Moonshine phenomenon. The complete construction of the Monster group is too voluminous to be explained here. We will present a sketched partial construction of $\mathbb{M}$ in Section 5.4.

For now, we describe a remarkably simple representation of $\mathbb{M}$. As with any noncyclic finite simple group, it is generated by its involutions (*i. e.*, elements of order 2) and so will be an image of a Coxeter group. Let $\mathcal{G}_{pqr}$, $p \geq q \geq r \geq 2$, be the graph consisting of three strands of lengths $p+1$, $q+1$, $r+1$, sharing a common endpoint. Label the $p+q+r+1$ nodes as in Figure 1.1. Given any graph $\mathcal{G}_{pqr}$, define Y_{pqr} to be the group consisting of a generator for each node, obeying the usual Coxeter group relations (*i. e.*, all generators are involutions, and the product gg' of two generators has order 2 or 3, depending on whether or not the two nodes are adjacent), together with one more relation:

$$(ab_1b_2ac_1c_2ad_1d_2)^{10} = 1.$$

The groups Y_{pqr}, for $p \leq 5$, are all identified (see [102]). Conway conjectured and, building on work by Ivanov [103], Norton proved [156] that $Y_{555} \cong Y_{444}$ is the *Bimonster*, the wreathed-square $\mathbb{M} \wr \mathbb{Z}_2$ (or $(\mathbb{M} \times \mathbb{M}).2$ in Conway's notation), that is, a group with $\mathbb{M} \times \mathbb{M}$ as normal subgroup and $\mathbb{Z}_2$ as quotient, with order $2|\mathbb{M}|^2$. A closely related presentation of the Bimonster has 26 involutions as generators and has relations given by the incidence graph of the projective plane of order 3; the Monster itself arises from 21 involutions and the affine plane of order 3. Likewise, $Y_{553} \cong Y_{443} \cong \mathbb{M} \times \mathbb{Z}_2$. Other sporadic groups arise in *e. g.*, $Y_{533} \cong Y_{433} \cong \mathbb{B}$ is the Baby Monster; $Y_{552} \cong Y_{442}$ is the Fischer group Fi'$_{24}$; and $Y_{532} \cong Y_{432}$ is the Fischer group Fi$_{23}$. The Coxeter groups associated to the other graphs $\mathcal{G}_{pqr}$, $p \leq 5$ are all finite groups of hyperbolic reflections in $\mathbb{R}^{17,1}$ and contain copies of the Weyl group E_8, giving a rich underlying geometry.

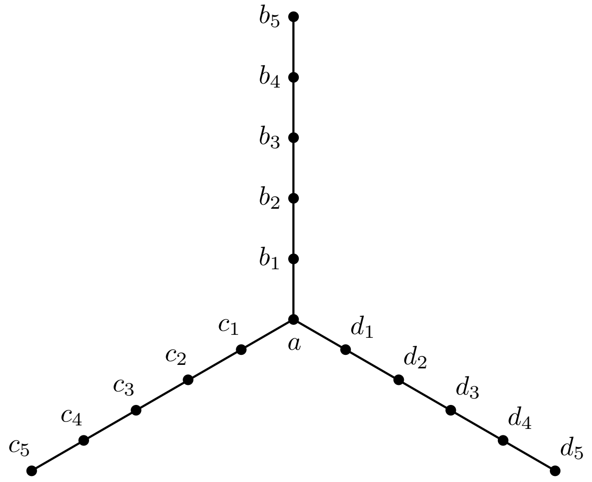

Figure 1.1: The graph $\mathcal{G}_{555}$ representing the Bimonster.

1.4 Modular functions

Although the Monster group $\mathbb{M}$ was discovered within the context of finite simple groups, hints later began to emerge that it might be strongly related to other branches of mathematics. One of these is the theory of modular functions and modular forms.

Consider the action of the group $SL_2(\mathbb{R})$ on the upper half-plane $\mathbb{H}$. Let

$$SL_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{R}) : ad - bc = 1 \right\}.$$

$SL_2(\mathbb{R})$ acts on $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d},$$

(note that $\mathrm{Im}(z) > 0 \Rightarrow \mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \mathrm{Im}(z) > 0$). In particular, the subgroup $SL_2(\mathbb{Z})$ acts on $\mathbb{H}$ discontinually.

Since $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \cdot z = \frac{-z}{-1} = z$, we see that $PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I\}$ acts on $\mathbb{H}$. We call $\Gamma = PSL_2(\mathbb{Z})$ the *modular group* and we denote by $\mathbb{H}/SL_2(\mathbb{Z})$ the set of orbits. Since

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in SL_2(\mathbb{Z}) \text{ and } \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot z = z + 1,$$

then z and $z+1$ are in the same orbit. Thus, each orbit intersects

$$\{z \in \mathbb{H} : -1/2 \leq \mathrm{Re}(z) \leq 1/2\}.$$

Similarly,

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in SL_2(\mathbb{Z}) \text{ and } \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \cdot \sigma = -\frac{1}{\sigma},$$

so the elements σ and $-1/\sigma$ are in the same orbit. Thus, each orbit intersects

$$\{\sigma \in \mathbb{H} : |\sigma| \geq 1\}.$$

In particular, if $|\sigma| = 1$, then $\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \cdot \sigma = -\frac{1}{\sigma} = -\bar{\sigma}$. In fact, Theorem 6.1.2 shows that we obtain a fundamental domain D for the action of $SL_2(\mathbb{Z})$ on $\mathbb{H}$ taking the region

$$D = \{z \in \mathbb{H} : -1/2 \leq \mathrm{Re}(z) \leq 1/2, |z| \geq 1\},$$

and identifying z with $z+1$, for $\mathrm{Re}(z) = -1/2$; and σ with $-\bar{\sigma} = -\frac{1}{\sigma}$, for $|\sigma| = 1$ (see Figure 1.2).

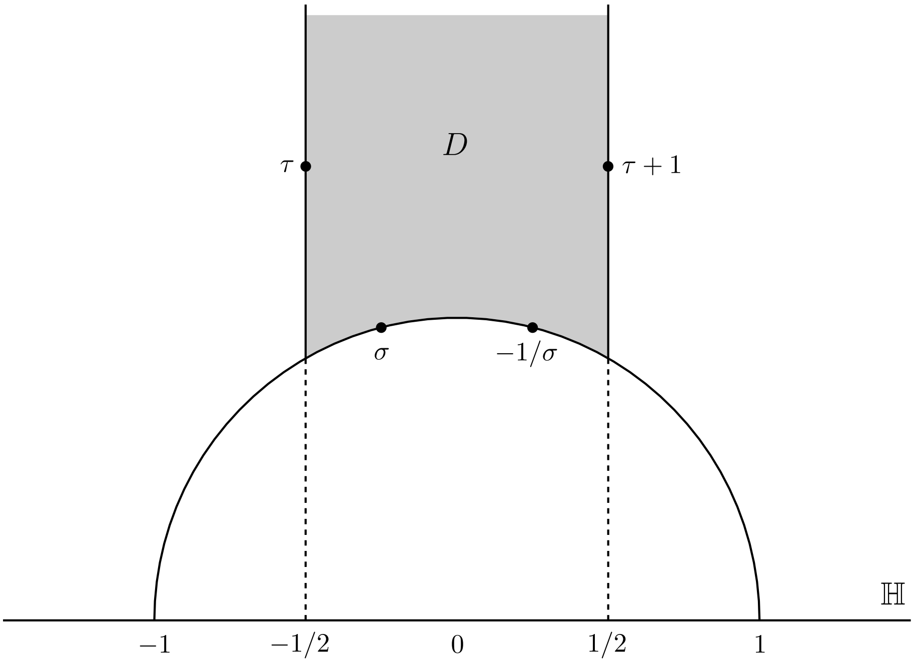

Figure 1.2: Fundamental domain D for the action of $PSL_2(\mathbb{Z})$ on $\mathbb{H}$.

With these identifications we obtain a set D intersecting each orbit just at one point. In fact, the canonical map $D \to \mathbb{H}/SL_2(\mathbb{Z})$ is surjective and its restriction to the interior of D is injective. We can prove also that the modular group Γ is generated by S and T (see Theorem 6.1.3). The set $\mathbb{H}/SL_2(\mathbb{Z})$ of orbits has the structure of a Riemann surface with one point removed. This is a Riemann surface of genus 0. When we remove one point of it, we obtain a set which can be identified with $\mathbb{C}$. Thus, we have an isomorphism of Riemann surfaces

$$\mathbb{H}/SL_2(\mathbb{Z}) \to \mathbb{C}.$$

This map can be extended to an isomorphism between compact Riemann surfaces

$$\mathbb{H}/SL_2(\mathbb{Z}) \cup \{i\infty\} \to \overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\} \cong \mathbb{CP}^1,$$

which maps $i\infty \mapsto \infty$. Such an isomorphism is not unique. However, when we operate with Riemann surfaces of genus 0, we can prove that if j is just one of these isomorphisms, then any other is of the form $a(j+b)$, where a, b are constants and $a \neq 0$.

Let $j : \mathbb{H}/SL_2(\mathbb{Z}) \to \overline{\mathbb{C}}$ be one of such isomorphisms. Then, j defines a map $j : \mathbb{H} \to \overline{\mathbb{C}}$ that is constant on orbits. Since z and $z+1$ lies on the same orbit, then we have $j(z) = j(z+1)$. Thus, j is periodic (of period 1). This implies that j has a Fourier expansion

$$j(z) = \sum_{n\in\mathbb{Z}} c_n e^{2\pi i n z}.$$

Writing $q = e^{2\pi iz}$, then we have

$$j(z) = \sum_{n\in\mathbb{Z}} c_n q^n.$$

Let k be an integer. We say that a meromorphic function $f : \mathbb{H} \to \mathbb{C}$ is a *modular function* of weight $2k$ if[1]

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^{2k} f(z), \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}).$$

If f is holomorphic everywhere —including at infinity—, we say that f is a *modular form*. We give some examples (see Section 6.3):

Example 1.4.1. Let $\sigma_3(n) = \sum_{d|n} d^3$. We define the *Eisenstein series*

$$E_4(z) = 1 + 240 \sum_{n\geq 1} \sigma_3(n) q^n = 1 + 240q + 2160q^2 + \dots$$

Thus, $E_4(z)$ is a modular form of weight 4.

Example 1.4.2. The *Dedekind's function*

$$\eta(z) = q \prod_{d\geq 1} (1-q^n)^{24} = q - 24q^2 + 252q^3 - \dots$$

is a modular form of weight 12.

We now define $j : \mathbb{H} \to \mathbb{C}$ by

$$j(z) = \frac{(E_4(z))^3}{\eta(z)}.$$

This is a modular form of weight 0, so it is constant on orbits of $PSL_2(\mathbb{Z})$ on $\mathbb{H}$. In fact,

$$j(z) = q^{-1} + 744 + 196,884q + 21,493,760q^2 + \dots \tag{1.3}$$

Such maps are usually called *fundamental functions* or *Hauptmodul*.

Observe that $j(z)$ is holomorphic on $\mathbb{H}$ and that $j(z)$ has a simple pole at $z = i\infty$ (*i. e.*, at $q = 0$). Then, j gives an holomorphic isomorphism between $\mathbb{H}/SL_2(\mathbb{Z})$ and $\mathbb{C}$ that can be extended to a meromorphic isomorphism of compact Riemann surfaces $\overline{\mathbb{H}}/SL_2(\mathbb{Z}) \to \overline{\mathbb{C}}$, where $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{Q} \cup \{i\infty\}$.
The $SL_2(\mathbb{Z})$-orbits of $\mathbb{Q} \cup \{i\infty\}$ are called *cusps* and their role is to fill the punctures

[1] Some authors say that f is of weight k.

of $\mathbb{H}/SL_2(\mathbb{Z})$, compactifying the surface, as there are much fewer meromorphic functions on compact surfaces than on noncompact ones. Since any other isomorphism is of the form $a(j(z)+b)$, with a, b constants and $a \neq 0$, in particular there is just one of such isomorphisms with leading coefficient $c_1 = 1$ and constant term $c_0 = 0$:

$$J(z) = j(z) - 744 = q^{-1} + 196,884q + 21,493,760q^2 + \ldots \tag{1.4}$$

We shall call this function $J(z)$ the *canonical isomorphism* or the *normalized Hauptmodul* of the modular group.

The expansion coefficients of $J(z)$, which are all positive integers (except for the constant term), might appear unattractive. As we shall see, it took many years and an accident before their meaning was finally found.

1.5 Some on Kleinian groups

Even before the discovery of the modular invariant $j(z)$ was made, it was observed that certain characteristics of elliptic functions with periods ω_1, ω_2 were invariant only under a certain subgroup $\Gamma(2)$ of $\Gamma = SL_2(\mathbb{Z})$. This and other facts led Klein to the creation of the theory of congruence subgroups [115]. He introduced a class of principal congruence subgroups

$$\Gamma(n) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod{n} \right\}, \text{ for } n > 0,$$

and notions of congruence subgroups Γ' of level n. An example of a congruence subgroup of level n is the class

$$\Gamma_0(n) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : c \equiv 0 \pmod{n} \right\}.$$

At the same time, Poincaré, influenced by a paper of Fuchs [69], launched a program to study discrete subgroup of $PSL_2(\mathbb{R})$ and their corresponding automorphic functions (analogous to modular functions), and this laid to the theory of Fuchsian groups.

One of the basics results of this theory is that for any Kleinian group Γ', a suitable compactification of $\mathbb{H}/\Gamma'$ has the structure of a compact Riemann surface. In the special case when the genus of $\mathbb{H}/\Gamma'$ is zero the theory of automorphic functions becomes specially simple: the field of automorphic functions is generated by only one function $J_{\Gamma'}(z)$, called the *Hauptmodul* of Γ'. In the particular case of the modular group $\Gamma = PSL_2(\mathbb{Z})$, the surface —the Riemann sphere— has genus zero, and the Hauptmodul of Γ is just $J(z)$ in (1.4).

Fricke [68] investigated the surfaces associated with $\Gamma_0(n)$. In particular, congruence subgroups $\Gamma_0(p)$, for p prime, provide examples of genus zero surfaces if and only if $p - 1 \mid 24$.

One can obtain more examples adjoining to $\Gamma_0(n)$ the Fricke involution $w_n(z) = -1/nz$, which of course can be realized as an element of $PSL_2(\mathbb{R})$. That is

$$\Gamma_0(n)^+ = \left\langle \Gamma_0(n),\ \frac{1}{\sqrt{n}} \begin{pmatrix} 0 & -1 \\ n & 0 \end{pmatrix} \right\rangle.$$

The normalizer of $\Gamma_0(n)$ in $PSL_2(\mathbb{R})$ was fully described by Atkin and Lehner [3]. When n is a prime p, it is just the group $\Gamma_0(p)^+$ generated by $\Gamma_0(p)$ and the Fricke involution w_p. Ogg [158] completed Fricke's proof [68] that for a prime p, $\Gamma_0(p)^+$ has the genus zero property if and only if

$$p = 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 41, 47, 59, 71. \tag{1.5}$$

This strange set of primes could be passed to history as another mathematical fact without any special significance. It happened however that Ogg attended a talk of Tits mentioning a certain sporadic simple group predicted —but not proved— to exist by Fischer and Griess, of order

$$|\mathbb{M}| = 2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71,$$

and noticed that the primes appearing in list (1.5) are exactly the prime divisors of the order of the Monster group $\mathbb{M}$. What a coincidence! It was not realized at that time, at 1975, that this coincidence was the tip of an iceberg.

1.6 McKay's observation

Hints that the Monster might in fact be associated with an elegant structure had appeared before Griess announced his construction. We had mention already that Ogg, who was working on the field of modular functions, came with some coincidences. From the other side, McKay, who was working in finite group theory, noticed another relation between the Monster and modular functions: the near coincidence of the minimal possible representation of $\mathbb{M}$ (1.2) and the first non trivial coefficient c_1 of the Hauptmodul $J(z)$ in (1.4).

$$196,884 = 196,883 + 1.$$

Soon, McKay and Thompson found similar relations [172] including another dimensions of irreducible representations of $\mathbb{M}$, for example:

$$\begin{aligned} 196,884 &= 196,883 + 1, \\ 21,493,760 &= 21,296,876 + 196,883 + 1, \\ 864,299,970 &= 842,609,326 + 21,296,876 + 2 \cdot 196,883 + 2 \cdot 1, \end{aligned} \tag{1.6}$$

If we interpret $d_0 = 1$ as the dimension of the trivial representation of $\mathbb{M}$, then we have

$$c_1 = d_0 + d_1.$$

where d_1 is the dimension of the next irreducible representation of the Monster. In fact, if we compute some more few coefficients of $J(z)$

$$\begin{aligned} J(z) &= q^{-1} + 196,884q + 21,493,760q^2 + 864,299,970q^3 + \\ &\quad + 20,245,856,256q^4 + 333,202,640,600q^5 + 4,252,023,300,096q^6 + \dots \end{aligned}$$

and some more few dimensions of irreducible characters of $\mathbb{M}$

$d_0 = 1$	$d_1 = 196,883$
$d_2 = 21,296,876$	$d_3 = 842,609,326$
$d_4 = 18,538,750,076$	$d_5 = 19,360,062,527$
$d_6 = 293,553,734,298$	...

Table 1.2: First dimensions of irreducible characters of $\mathbb{M}$.

then equations (1.6) are extended:

$$\begin{aligned} c_1 &= d_0 + d_1, \\ c_2 &= d_0 + d_1 + d_2, \\ c_3 &= 2d_0 + 2d_1 + d_2 + d_3, \\ c_4 &= 2d_0 + 3d_1 + 2d_2 + d_3 + d_5, \\ c_5 &= 4d_0 + 5d_1 + 3d_2 + 2d_3 + d_4 + d_5 + d_6, \qquad (1.7) \\ &\dots \end{aligned}$$

and further relations of this sort. Based on these observations, McKay and Thompson conjectured the existence of a natural infinite-dimensional representation of the Monster

$$V = V_{-1} \oplus V_1 \oplus V_2 \oplus V_3 \oplus \dots,$$

such that $\dim V_n = c_n$, for $n = -1, 1, 2, 3, \dots$. That is, this suggests that there is a graded vector space acted on by the Monster, and that our Hauptmodul $J(z)$ is in fact what we call the *graded dimension* of V:

$$J(z) = \sum_{n\geq -1} c_n q^n = q^{-1} + \sum_{n\geq 1} (\dim V_n) q^n. \qquad (1.8)$$

McKay, also gave a relation between the j-function and the classical Lie algebra E_8 (see Section 4.7). A more elementary observation concerns the Leech lattice Λ_{24}. The Leech lattice is a particularly special one in 24 dimensions. It has some special features. The book by Conway and Sloane [39] includes a lot of interesting details relating the Leech lattice. Another useful reference is [86]. For example, 196,560 is the number of vectors in the Leech lattice with (squared) norm equal to 4, and note that this number is also close to 196,884: In fact, the first coefficients of the Hauptmodul J (1.4) are related to the Leech lattice

$$\begin{aligned} 196,884 &= 196,560 + 324 \cdot 1, \\ 21,493,760 &= 16,773,120 + 24 \cdot 196,560 + 3,200 \cdot 1, \\ 864,299,970 &= 398,034,000 + 24 \cdot 16,773,120 + 324 \cdot 196,560 + 25,650 \cdot 1, \end{aligned}$$

where 16,773,120 and 398,034,000 are the numbers of 6-norm and 8-norm vectors in Λ_{24}. This may not seem as convincing as (1.6), but the same equations hold for any of the 24-dimensional even self-dual lattices, apart from an extra term on the right sides corresponding to 2-norm vectors (there are none of these in the Leech).

What conceptually does the Monster, the Leech lattice have to do with the j-function? Is there a common theory explaining this numerology? The answer is yes. In fact, as we shall see, there is a relation between E_8 to the j-function. In the late 1960's, Victor Kac [110] and Robert Moody [149] independently defined a new class of infinite-dimensional Lie algebras. A Lie algebra is a vector space with a bilinear vector-valued product that is both anti-commutative and anti-associative (Chapter 2). The familiar vector-product $u \cdot v$ in three dimensions defines a Lie algebra, called $\mathfrak{sl}_2(\mathbb{R})$, and in fact this algebra generates all Kac-Moody algebras. Within a decade it was realised that the graded dimensions of representations of the affine Kac-Moody algebras are (vector-valued) modular functions for $SL_2(\mathbb{Z})$ (see Theorem 3.2.3 in [72]).

1.7 The Monstrous Moonshine conjecture

Recall the strong connections of previous section, that suggested the existence of the graded algebra V. Thompson [171] also proposed considering, for any element $g \in \mathbb{M}$, the modular properties of the series

$$T_g(z) = \sum_{n\in\mathbb{Z}} \mathrm{tr}(g|V_n)q^n = q^{-1} + \mathrm{tr}(g|V_1)q + \mathrm{tr}(g|V_2)q^2 + \ldots, \tag{1.9}$$

where $q = e^{2\pi iz}$ and V_n is the n-th graded component of V. This graded trace above is called the *McKay-Thompson series* of g, and generalizes our Hauptmodul on (1.4). For example, if $g = 1$ (the identity element of $\mathbb{M}$), then $T_g(z)$ is just $J(z)$.

Working with data from the table character of the Monster, remarkable numerology concerning these graded traces was collected in the paper *Monstrous Moonshine* [38]. It has been pointed by Conway that he found some of these interesting series in the classical book of Jacobi [106].

Influenced by Ogg's observation, Thompson, Conway and Norton realized that all the series they were discovering (proceeding experimentally by the first few coefficients) were normalized generators of genus zero function fields arising from certain discrete subgroups of $PSL_2(\mathbb{R})$, that is the McKay-Thompson series they were discovering behave as 'mini j-functions', for certain other subgroups of $PSL_2(\mathbb{R})$. They were led to conjecture that there exists a graded representation V of the Monster —in fact a double-graded space— with all the functions $T_g(z)$ having this genus zero property. Knowing the functions $T_g(z)$ determines the $\mathbb{M}$-module V uniquely, and the question was whether it existed, given the list of proposed functions $T_g(z)$, for each of the 194 conjugacy classes of $\mathbb{M}$. Such graded module was subsequently discovered by Frenkel, Lepowsky and Meurman [66], [67]. It is called the *monster vertex algebra* or the *moonshine module* $V^\natural$. We shall study it in Chapter 5.

Instead of consider only the modular subgroup $PSL_2(\mathbb{Z})$ of $PSL_2(\mathbb{R})$, we shall consider other subgroups. A subgroup G of $PSL_2(\mathbb{R})$ is called *commensurable* with $PSL_2(\mathbb{Z})$ if:

- $[PSL_2(\mathbb{Z}) : PSL_2(\mathbb{Z}) \cap G]$ is finite;
- $[G : PSL_2(\mathbb{Z}) \cap G]$ is finite.

We may consider the action of such a group on the upper half-plane $\mathbb{H}$. We know that if G is a subgroup of $PSL_2(\mathbb{R})$ commensurable with $PSL_2(\mathbb{Z})$, then the set of orbits $\mathbb{H}/G$ has the structure of a compact Riemann surface, with finitely many points removed. We write

$$\mathbb{H}/G \subseteq \overline{\mathbb{H}}/G,$$

where $\overline{\mathbb{H}}/G$ is a compact Riemann surface. Let $\overline{\mathbb{H}}/G$ a compact Riemann surface of genus g. Then, $\overline{\mathbb{H}}/G$ is homeomorphic to a g-torus. In the case $g = 0$, $\overline{\mathbb{H}}/G$ is homeomorphic to the Riemann sphere $\mathbb{CP}^1$. A subgroup G is called of *genus* 0, if the Riemann surface $\overline{\mathbb{H}}/G$ has genus 0. We have already pointed in Section 1.5 that under these particular circumstances, the field of automorphic functions of $\overline{\mathbb{H}}/G \to \mathbb{CP}^1$ is generated by just one element, and we can take this element as the unique isomorphism of Riemann surfaces $\overline{\mathbb{H}}/G \to \mathbb{CP}^1$ with leading coefficient 1 and constant term 0. We denote this by $J_G(z)$, and is called the *canonical isomorphism* or *normalized Hauptmodul* of G. So J_G, plays exactly the same role for G that the J-function plays for $SL_2(\mathbb{Z})$. For example, in the case of $\Gamma_0(2)$, $\Gamma_0(13)$ and $\Gamma_0(25)$ —are all genus 0 subgroups of $PSL_2(\mathbb{R})$—, we obtain Hauptmoduls

$$\begin{aligned} J_{\Gamma_0(2)}(z) &= q^{-1} + 276q - 2048q^2 + 11202q^3 - 49152q^4 + 184024q^5 + \dots, && (1.10)\\ J_{\Gamma_0(13)}(z) &= q^{-1} - q + 2q^2 + q^3 + 2q^4 - 2q^5 - 2q^7 - 2q^8 + q^9 + \dots, && (1.11)\\ J_{\Gamma_0(25)}(z) &= q^{-1} - q + q^4 + q^6 - q^{11} - q^{14} + q^{21} + q^{24} - q^{26} + \dots && (1.12) \end{aligned}$$

We now state formally an initial form of the Conway-Norton conjecture, or Moonshine conjecture.

Conjecture 1.7.1 (**Moonshine Conjecture**). *(Conway-Norton, 1979).*
For each $g \in \mathbb{M}$, the McKay-Thompson series $T_g(z)$ is the normalized Hauptmodul $J_G(z) : \overline{\mathbb{H}}/G \to \mathbb{CP}^1$ for some subgroup G of $SL_2(\mathbb{R})$ commensurable with $PSL_2(\mathbb{Z})$.

The first major step in the proof of Monstrous Moonshine was accomplished in the mid 1980's with the construction by Frenkel-Lepowsky-Meurman [67] of the Moonshine module $V^\natural$, and its interpretation by Richard Borcherds [8] as a vertex operator algebra (VOA). In 1992, Borcherds [12] completed the proof of the original Monstrous Moonshine conjectures by showing that the graded characters T_g of $V^\natural$ are indeed the Hauptmoduls identified by Conway and Norton, and hence that $V^\natural$ is indeed the desired representation $V^\natural$ of $\mathbb{M}$ conjectured by McKay and Thompson. The algebraic structure appearing in moonshine typically arises as the symmetry group of the associated vertex operator algebra, for example, that of $V^\natural$ is the Monster $\mathbb{M}$. By Zhu's Theorem (Theorem 9.2.1), the modular functions appear as graded dimensions of the (possibly twisted) modules of a vertex operator algebra. In particular, the answer that this framework provides for what $\mathbb{M}$, E_8 and the Leech lattice have to do with the j-function is that they each correspond to a vertex operator algebra with a single simple module; their relation to j is then an immediate corollary to the much more general Zhu's Theorem.

Moonshine is also profoundly connected with geometry and physics (namely conformal field theory and string theory). String theory proposes that the elementary particles (electrons, photons, quarks, etc.) are vibrational modes on a string of length about 10^{-33} cm. These strings can interact only by splitting apart or joining together as they evolve through time, these (classical) strings will trace out a surface called the *world-sheet*. Quantum field theory tells us that the quantum quantities of interest (amplitudes) can be perturbatively computed as weighted averages taken over spaces of these world-sheets. Conformally equivalent world-sheets should be identified, so we are led to interpret amplitudes as certain integrals over moduli spaces of surfaces. This approach to string theory leads to a conformally invariant quantum field theory on two-dimensional space-time, called conformal field theory (CFT). The various modular forms and functions arising in Moonshine appear as integrands in some of these genus 1 surfaces appearing in these conformal theories. We shall explain a bit of these relation between Moonshine and physics in Chapter 9. Thus, the actual importance of the Moonshine phenomenon: It proposes a conexion between four branches of mathematics, namely algebraic structures, modular structures, geometry and mathematical physics.

Chapter 2

Some basics on Lie algebras

In this chapter we introduce Lie algebras termed affine, of which Virasoro algebra plays a central role throughout this work. We also present standard constructions of important classes of modules and algebras and we discuss the concept of graded dimension. We have introduced a number of elementary concepts in first sections to provide sufficient background for understanding the main concepts. For an extensive exposition of basic algebra, the reader may refer to [107, 108] and [127]. A detailed reference on Lie groups and Lie algebras can be found in [100] and [179], or the more recent [91].

2.1 Representations of finite groups

Let V be a vector space over the field $\mathbb{C}$ of complex numbers and let $GL(V)$ be the group of isomorphisms of V onto itself. An element $a \in GL(V)$ is, by definition, a linear mapping of V into V which has an inverse a^{-1}; this inverse is linear. When V has a finite basis (of n elements), each linear map $a : V \to V$ is defined by a square matrix $[a_{ij}]$ of order n. The coefficients a_{ij} are complex numbers, and we usually write in this case $GL(V)$ as $GL_n(\mathbb{C})$.

Suppose now G is a finite group, with identity element 1 and with composition $(s, t) \mapsto st$. A *linear representation* of G in V is a homomorphism π from the group G into the group $GL(V)$. In other words, we associate with each element $s \in G$ an element $\pi(s)$ of $GL(V)$ in such a way that we have the equality

$$\pi(st) = \pi(s) \cdot \pi(t), \quad \text{for } s, t \in G.$$

Observe that the preceding formula implies the following: $\pi(1) = 1$, and $\pi(s^{-1}) = \pi(s)^{-1}$. When π is given, we say that V is a representation space of G (or even simply, a representation of G). We restrict ourselves to the case where V has finite dimension. When $\dim V = n$, we say that a representation $\pi : G \to GL(V)$ has *degree* n.

Let $\pi : G \to GL(V)$ be a linear representation and let W be a vector subspace of V. We say that W is *stable* under the action of G (we say also invariant), if $\pi(s)W \subseteq W$, for all $s \in G$. Given a linear representation $\pi : G \to GL(V)$, we say that it is *irreducible* or *simple* if V is not 0 and if V has no nontrivial stable subspaces under G. This condition is equivalent to saying V is not the direct sum of two representations (except for the trivial $V = 0 \oplus V$). We have the following result [166, p.7]

Theorem 2.1.1. *Every representation is a direct sum of irreducible representations.*

Let V be a vector space having a basis $e_1, \ldots, e_n$, and let a be a linear map of V into itself, with matrix $[a_{ij}]$. By the *trace* of a we mean the scalar

$$\operatorname{tr} a = \sum_{1 \leq i \leq n} a_{ii}.$$

It is the sum of the eigenvalues of a (counted with their multiplicities), and does not depend on the choice of basis $\{e_i\}$. Now let $\pi : G \to GL(V)$ be a linear representation of a finite group G in the vector space V. For each $s \in G$, we put

$$\chi_\pi(s) = \operatorname{tr}(\pi(s)).$$

The complex valued function χ_π on G thus obtained is called the *character* of the representation π; the importance of this function comes primarily from the fact that it characterizes the representation π.

Proposition 2.1.2. *If χ is the character of a representation π of degree n, we have:*

1. $\chi(1) = n$.
2. $\chi(s^{-1}) = \overline{\chi(s)}$, *for all* $s \in G$
3. $\chi(tst^{-1}) = \chi(s)$, *for all* $s, t \in G$.

Proof. We have $\pi(1) = I$, and $\operatorname{tr}(I) = n$ since V has dimension n; hence (1). For (2) we observe that $\pi(s)$ has finite order; consequently the same is true of its eigenvalues $\lambda_1, \ldots, \lambda_n$ and so these have absolute value equal to 1 (this is also a consequence of the fact that π can be defined by a unitary matrix. Thus

$$\overline{\chi(s)} = \overline{\operatorname{tr}(\pi(s))} = \sum \overline{\lambda_i} = \sum \lambda_i^{-1} = \operatorname{tr}(\pi(s)^{-1}) = \operatorname{tr}_{\pi^{-1}}(s) = \operatorname{tr}_\pi(s^{-1}).$$

Formula (3) can also be written $\chi(vu) = \chi(uv)$, putting $u = ts$, $v = t^{-1}$; hence it Follows from the well known formula $\operatorname{tr}(ab) = \operatorname{tr}(ba)$, valid for two arbitrary linear mappings a and b of V into itself. □

A function f on G satisfying identity (3) above, is called a *class function.* We shall see later that each class function is a linear combination of characters. We will use the following result [166, p.11]

Proposition 2.1.3. *Let $\pi_1 : G \to GL(V_1)$ and $\pi_2 : G \to GL(V_2)$ be two linear representations of G, and let χ_1 and χ_2 be their characters. Then:*

1. *The character χ of the direct sum representation $V_1 \oplus V_2$ is equal to $\chi_1 + \chi_2$.*
2. *The character ψ of the tensor product representation $V_1 \otimes V_2$ is equal to $\chi_1 \cdot \chi_2$.*

If ϕ and ψ are functions on G, set

$$\langle \phi, \psi \rangle = \frac{1}{|G|} \sum_{g \in G} \phi(g^{-1}) \psi(g) = \frac{1}{|G|} \sum_{g \in G} \phi(g) \psi(g^{-1}).$$

We have $\langle \phi, \psi \rangle = \langle \psi, \phi \rangle$. Moreover $\langle \phi, \psi \rangle$ is linear in ϕ and in ψ. Consider the following notation. If ϕ and ψ are two complex-valued functions on G, put

$$\langle \phi | \psi \rangle = \frac{1}{|G|} \sum_{g \in G} \phi(g) \overline{\psi(g)}.$$

This is a scalar product; it is linear in ϕ, semilinear in ψ; and we have $\langle \phi | \phi \rangle > 0$, for all $\phi \neq 0$. If $\check{\psi}$ is the function defined by the formula $\check{\psi}(g) = \overline{\psi(g^{-1})}$, then we have

$$\langle \phi | \psi \rangle = \frac{1}{|G|} \sum_{g \in G} \phi(g) \check{\psi}(g^{-1}) = \langle \phi, \check{\psi} \rangle.$$

In particular, if χ is the character of a representation of G, we have $\check{\chi} = \chi$ (Proposition 2.1.2), so that $\langle \phi | \chi \rangle = \langle \phi, \chi \rangle$, for all functions ϕ on G. So we can use at will $\langle \phi | \chi \rangle$ or $\langle \phi, \chi \rangle$, so long as we are concerned with characters. We have an important result on orthogonality of characters [166, p.15]:

Theorem 2.1.4. 1. *If χ is the character of an irreducible representation, we have $\langle \chi | \chi \rangle = 1$ (i. e., χ is of norm 1).*

2. *If χ and χ' are the characters of two nonisomorphic irreducible representations, we have $\langle \chi | \chi' \rangle = 0$ (i. e., χ and χ' are orthogonal).*

A character of an irreducible representation is called an *irreducible character*. Thus, Theorem 2.1.4 shows that the irreducible characters form an orthonormal system.

Theorem 2.1.5. *Let V be a linear representation of G, with character ϕ, and suppose V decomposes into a direct sum of irreducible representations:*

$$V = W_1 \oplus W_2 \oplus \cdots \oplus W_k.$$

Then, if W is an irreducible representation with character χ, the number of W_i's isomorphic to W is equal to the scalar product $\langle \phi | \chi \rangle$.

Proof. Let χ_i be the character of W_i. By Proposition 2.1.3, we have $\phi = \chi_1 + \ldots + \chi_k$. Thus, $\langle \phi | \chi \rangle = \langle \chi_1 | \chi \rangle + \ldots + \langle \chi_k | \chi \rangle$. But, according to the preceding Theorem 2.1.4, $\langle \chi_i | \chi \rangle$ is equal to 1 or 0, depending on whether W_i is, or is not, isomorphic to W. The result follows. □

Corollary 2.1.6. *The number of W_i isomorphic to W does not depend on the chosen decomposition.*

This number is called the 'number of times that W occurs in V', or the 'number of times that W is contained in V'. It is in this sense that one can say that there is uniqueness in the decomposition of a representation into irreducible representations.

Corollary 2.1.7. *Two representations with the same character are isomorphic.*

The above results reduce the study of representations to that of their characters. If $\chi_1, \ldots, \chi_k$ are the distinct irreducible characters of G, and if $W_1, \ldots, W_k$ denote corresponding representations, each representation V of G is isomorphic to a direct sum

$$V = m_1 W_1 \oplus m_2 W_2 \oplus \ldots \oplus m_k W_k,$$

with $m_1, \ldots, m_k \geq 0$ integers. The character ϕ of V is equal to $m_1\chi_1 + \ldots + m_k\chi_k$ and we have $m_i = \langle \phi | \chi_i \rangle$. We obtain thus a very convenient irreducibility criterion [166, p.17]:

Theorem 2.1.8. *If χ is the character of a representation V, $\langle \phi | \phi \rangle$ is a positive integer and we have $\langle \phi | \phi \rangle = 1$ if, and only if, V is irreducible.*

Example 2.1.9 (The regular representation)**.** Let n be the order of G, and let V be a vector space of dimension n, with a basis $\{e^g\}_{g \in G}$ indexed by the elements g of G. For $s \in G$, let $\pi(s)$ be the linear map of V into V which sends $e^g \mapsto e^{sg}$; this defines a linear representation, which is called the *regular representation* of G. Its degree is equal to the order of G. Note that $e^g = \pi_g(e^1)$; hence note that the images of e^1 form a basis of V. Conversely, let W be a representation of G containing a vector w such that the $\pi_g(w)$, $g \in G$, form a basis of W; then W is isomorphic to the regular representation (an isomorphism $\tau : V \to W$ is defined by putting $\tau(e^g) = \pi_g(w)$.

For the rest of this section, the irreducible characters of G are denoted $\chi_1, \ldots, \chi_k$; their degrees are written $n_1, \ldots, n_k$, where $n_i = \chi_i(1)$ (Proposition 2.1.2). Let R be the regular representation of G. Recall that it has a basis $\{e^g\}_{g \in G}$ such that $\pi_s(e^g) = e^{sg}$. If $s \neq 1$, we have $sg \neq s$ for all g, which shows that the diagonal terms of the matrix of π_s are zero; in particular we have $\mathrm{tr}(\pi_s) = 0$. On the other hand, for $s = 1$, we have $\mathrm{tr}(\pi_s) = \mathrm{tr}(I) = \dim R = n$. Whence:

Proposition 2.1.10. *The character r_G of the regular representation is given by the formulas:*

$$r_G(1) = n = |G|, \quad r_G(s) = 0, \quad \textit{if } s \neq 1.$$

Corollary 2.1.11. *Every irreducible representation W_i of G is contained in the regular representation with multiplicity equal to its degree n_i.*

Proof. According to Theorem 2.1.4, this number is equal to $\langle r_G | \chi_i \rangle$, and we have

$$\langle r_G | \chi_i \rangle = \frac{1}{|n|} \sum_{g \in G} r_G(s^{-1}) \chi_i(s) = \frac{1}{n}\big(r_G(1)\chi_i(1)\big) = \frac{1}{n}\big(n \cdot \chi_i(1)\big) = \chi_i(1) = n_i. \ \square$$

Corollary 2.1.12. *1. The degrees n_i satisfy the relation $\sum_i n_i^2 = n$.*

2. If $s \in G$ is different from 1, we have $n_i \chi_i(s) = 0$.

Proof. By Corollary 2.1.11, we have $r_G(s) = \sum_i n_i \chi_i(s)$, for all $s \in G$. Taking $s = 1$ we obtain (1), and taking $s \neq 1$, we obtain (2). □

Note that the above result can be used in determining the irreducible representations of a group G: suppose we have constructed some mutually non isomorphic irreducible representations of degrees $n_1, \ldots, n_k$; in order that they be all the irreducible representations of G (up to isomorphism), it is necessary and sufficient that $n_1^2 + \ldots + n_k^2 = n$. Also, we will see below that the degrees n_i divide the order $n = |G|$.

Recall that a function f on G is called a class function if $f(tst^{-1}) = f(s)$ for all $s, t \in G$. Denote by H the space of class functions on G. In particular, the irreducible characters $\chi_1, \ldots, \chi_k$ belong to H. In fact, [166, p.19]

Theorem 2.1.13. *The irreducible characters $\chi_1, \ldots, \chi_k$ form an orthonormal basis of H.*

Recall that two elements g and g' of G are said to be conjugate if there exists $s \in G$ such that $g' = sgs^{-1}$; this is an equivalence relation, which partitions G into classes (also called *conjugacy classes*). We have the following (see [166, p.19–20])

Theorem 2.1.14. *The number of irreducible representations of G (up to isomorphism) is equal to the number of conjugacy classes of G.*

Proposition 2.1.15. *Let $s \in G$, and let $c(s)$ be the number of elements in the conjugacy class of s. Then, $\sum_i \overline{\chi_i(s)} \chi_i(s) = \frac{n}{c(s)}$, and for $g \in G$ not conjugate to s, we have $\sum_i \overline{\chi_i(s)} \chi_i(g) = 0$.*

We now give an example of a character table of a finite group.

Example 2.1.16 (Table of characters of S_3)**.** Take for G the group of symmetric group of three elements S_3. We have $n = 6$, and there are three conjugacy classes: the element $1 = (1)$; the three transpositions (1 2), (2 3), (3 1); and the two cyclic permutations (1 2 3), (1 3 2). Let t be a transposition and c a cyclic permutation. We have $t^2 = 1$, $c^3 = 1$, and $tc = c^2 t$ (equivalently, $ctct = 1$). Whence there are just two characters of degree 1: the unit character χ_1 and the character χ_2 giving the signature of a permutation (that is 1 for the even permutations and -1 for the odd ones). Theorem 2.1.14 above shows that there exists one other irreducible character χ_3; if n is its degree we must have $1 + 1 + n^2 = 6$, hence $n = 2$.

The values of χ_3 can be deduced from the fact that $\chi_1 + \chi_2 + 2\chi_3$ is the character of the regular representation r_{S_3} of S_3 (Proposition 2.1.10). We thus get the character table of S_3:

character	1	t	c
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

Table 2.1: Character table for the symmetric group S_3.

2.2 Lie groups

Consider the *general linear group* over the real numbers, denoted $GL_n(\mathbb{R})$, *i. e.*, the group of all $n \times n$ invertible matrices with real entries. Similarly, we consider the the general linear group over the complex numbers of all $n \times n$ invertible matrices with complex entries, denoted $GL_n(\mathbb{C})$. The general linear groups are indeed groups under the operation of matrix multiplication: the product of two invertible matrices is invertible, the identity matrix is an identity for the group, an invertible matrix has (by definition) an inverse, and matrix multiplication is associative.
Let $\mathbb{C}^{n\times n}$ denote the space of all $n \times n$ matrices with complex entries, and let $\{A_k\}_k$ be a sequence of complex matrices in $\mathbb{C}^{n\times n}$. We say that $\{A_k\}_k$ *converges* to a matrix A if each entry of $(A_k)_{ij}$ of A_k converges to the corresponding entry of A_{ij} of A.

Definition 2.2.1. A *matrix Lie group* is any subgroup G of $GL_n(\mathbb{C})$ with the following property: If $\{A_k\}_k$ is any sequence of matrices in G, and $\{A_k\}$ converges to some matrix A then either $A \in G$, or A is not invertible.

The condition on G amounts to saying that G is a closed subgroup of $GL_n(\mathbb{C})$ (although it does not necessarily mean that G is closed as a subset of $\mathbb{C}^{n\times n}$. Most of the 'interesting' subgroups of $GL_n(\mathbb{C})$ have this property, we give some classical examples:

Example 2.2.2. The general linear groups $GL_n(\mathbb{R})$ and $GL_n(\mathbb{C})$ are themselves matrix Lie groups.

Example 2.2.3. (The special linear groups $SL_n(\mathbb{R})$ and $SL_n(\mathbb{C})$)
The *special linear group* (over $\mathbb{R}$ or $\mathbb{C}$) is the group of $n \times n$ invertible matrices (with real or complex entries) having determinant one. That is $SL_2(\mathbb{C}) = \{A \in GL_2(\mathbb{C}) : \det A = 1\}$. Both of these are subgroups of $GL_n(\mathbb{C})$. Furthermore, if $\{A_k\}_k$ is a sequence of matrices with determinant one and $\{A_k\}$ converges to A, then continuity of determinant function implies that A also has determinant one.

Example 2.2.4. (The orthogonal and special orthogonal groups $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$)
An $n \times n$ real matrix A is said to be *orthogonal* if the column vectors that make up A are

orthonormal, that is, if

$$\sum_{\ell=1}^{n} A_{\ell j} A_{\ell k} = \delta_{jk}, \quad \text{for all } 1 < j, k < n.$$

Equivalently, A is orthogonal if it preserves the inner product, namely if $\langle x, y\rangle = \langle Ax, Ay\rangle$, for all vectors $x, y \in \mathbb{R}^n$. Since $\det A^{\mathrm{tr}} = \det A$, we see that if A is orthogonal, then $(\det A)^2 = \det(A^{\mathrm{tr}} A) = \det I = 1$. Hence, $\det A = \pm 1$, for all orthogonal matrices A. In particular, every orthogonal matrix must be invertible. However, if A is an orthogonal matrix, then

$$\langle A^{-1}x, A^{-1}y\rangle = \langle A(A^{-1})x, A(A^{-1})y\rangle = \langle x, y\rangle.$$

Thus, the inverse of an orthogonal matrix is orthogonal. Furthermore, the product of two orthogonal matrices is orthogonal, since if A and B both preserve inner products, then so does AB. Thus, the set of orthogonal matrices forms a group.
The set of all $n \times n$ real orthogonal matrices is the *orthogonal group* $O_n(\mathbb{R})$, and it is a subgroup of $GL_n(\mathbb{C})$. The limit of a sequence of orthogonal matrices is orthogonal, because the relation $A^{\mathrm{tr}} A = I$ is preserved under taking limits. Thus, $O_n(\mathbb{R})$ is a matrix Lie group. The set of $n \times n$ orthogonal matrices with determinant one is the *special orthogonal group* $SO_n(\mathbb{R})$. This is also a matrix Lie group. Sometimes $O_n(\mathbb{R})$ and $SO_n(\mathbb{R})$ are simply denoted by $O(n)$ and $SO(n)$, respectively.

Example 2.2.5. (The complex orthogonal groups $O_n(\mathbb{C})$ and $SO_n(\mathbb{C})$)
Similarly to the example above, we can define orthogonal groups of complex matrices. Consider the bilinear form $\langle\cdot,\cdot\rangle$ on $\mathbb{C}^n$ given by $\langle x, y\rangle = \sum_k x_k y_k$. We define the *complex orthonormal group* as

$$O_n(\mathbb{C}) = \{A \in GL_n(\mathbb{C}) : \langle Ax, Ay\rangle = \langle x, y\rangle\}.$$

The *complex special orthogonal group* $SO_n(\mathbb{C})$ is defined to be the set of all matrices $A \in O_n(\mathbb{C})$ with $\det A = 1$. These groups are also matrix Lie groups.

Example 2.2.6. (The unitary groups $U(n)$ and $SU(n)$)
An $n \times n$ complex matrix A is said to be *unitary* if the column vectors of A are orthonormal, that is, if

$$\sum_{\ell=1}^{n} \overline{A_{\ell j}} A_{\ell k} = \delta_{jk}, \quad \text{for all } 1 < j, k < n.$$

Equivalently, A is unitary if it preserves the inner product $\langle x, y\rangle_{\mathbb{C}} = \sum_k x_k \bar{y}_k$, namely if $\langle Ax, Ay\rangle_{\mathbb{C}} = \langle x, y\rangle_{\mathbb{C}}$ for all vectors $x, y \in \mathbb{C}^n$. Still another equivalent definition is that A is unitary if $A^*A = \mathrm{I}$, where, $A^* = \overline{A}^{\mathrm{tr}}$ is the adjoint (hermitian) of A. Since $\det A^* = \overline{\det A}$, we see that if A is unitary, then $\det(A^*A) = |\det A|^2 = 1$. Hence, $|\det A| = 1$, for all unitary matrices A. In particular, this shows that every unitary matrix is invertible. The same argument as for the orthogonal group shows that the set of unitary matrices forms a group.

The set of all $n \times n$ unitary matrices is the unitary group $U(n)$, and it is a subgroup of $GL_n(\mathbb{C})$. The limit of unitary matrices is unitary, so $U(n)$ is a matrix Lie group. The set of unitary matrices with determinant one is the special unitary group $SU(n)$, and it can be proved that $SU(n)$ is a matrix Lie group.

Example 2.2.7. (The symplectic groups $Sp_n(\mathbb{R})$, $Sp_n(\mathbb{C})$, and $Sp(n)$)
Consider the skew-symmetric bilinear form $[\cdot,\cdot]$ on $\mathbb{R}^{2n}$ defined as follows:

$$[x,y] = \sum_{k=1}^{n} (x_k y_{n+k} - x_{n+k} y_k). \tag{2.1}$$

The set of all $2n \times 2n$ matrices A which preserve $[\cdot,\cdot]$ is called the *real symplectic group* $Sp_n(\mathbb{R})$, and it is a subgroup of $GL_{2n}(\mathbb{C})$. In fact, $Sp_n(\mathbb{R})$ is a matrix Lie group. This group arises naturally in the study of classical mechanics: If J is the $2n \times 2n$ matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

then $[x,y] = \langle x, Jy \rangle$, and it is possible to check that a $2n \times 2n$ real matrix A is in $Sp_n(\mathbb{R})$ if and only if $A^{\mathrm{tr}} J A = J$. Taking the determinant of this identity gives $(\det A)^2 \det J = \det J$, or $(\det A)^2 = 1$. This shows that $\det A = \pm 1$, for all $A \in Sp_n(\mathbb{R})$. In fact, $\det A = 1$ for all $A \in Sp_n(\mathbb{R})$.

One can define a bilinear form on $\mathbb{C}^{2n}$ by the same formula (2.1), since this form involves no complex conjugates. The set of $2n \times 2n$ complex matrices which preserve this form is called the *complex symplectic group* $Sp_n(\mathbb{C})$. Finally, we have the *compact symplectic group* $Sp(n)$ defined as

$$Sp(n) = Sp_n(\mathbb{C}) \cap U(2n).$$

Example 2.2.8. (The Heisenberg group H)
The set of all 3×3 real matrices A of the form

$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{where } a, b, c \in \mathbb{R},$$

is called the *Heisenberg group H*. In fact, H is a matrix Lie group. The reason for the name Heisenberg group is that the Lie algebra of H gives a realization of the Heisenberg commutation relations of quantum mechanics.

Example 2.2.9. The groups $\mathbb{R}^\times$, $\mathbb{C}^\times$, S^1, $\mathbb{R}$, and $\mathbb{R}^n$ can be thought as subgroups of matrices. For example, the group $\mathbb{R}^\times$ of non-zero real numbers under multiplication is isomorphic to $GL_1(\mathbb{R})$. Thus, we will regard $\mathbb{R}^\times$ as a matrix Lie group. Similarly, the group $\mathbb{C}^\times$. The group S^1 of complex numbers module one is isomorphic to $U(1)$.
The additive group $\mathbb{R}$ is isomorphic to $GL_1(\mathbb{R})^+$ (1×1 real matrices with positive determinant) via the map $x \mapsto [e^x]$. In a similar way, the additive group $\mathbb{R}^n$ is isomorphic to the group of diagonal real matrices with positive diagonal entries, via the map

$$(x_1, \dots, x_n) \mapsto \begin{pmatrix} e^{x_1} & & 0 \\ & \ddots & \\ 0 & & e^{x_n} \end{pmatrix}.$$

Now we discuss some notions of compactness and connectedness for matrix Lie groups.

Definition 2.2.10. A matrix Lie group G is said to be *compact* if the following two conditions are satisfied:
1. [closeness] If $\{A_k\}_k$ is any sequence of matrices in G and $\{A_k\}_k$ converges to a matrix A, then A is in G.
2. [boundedness] There exists a constant C such that for all $A \in G$, $|A_{ij}| < C$ for all $1 < i, j < n$.

This is not the usual topological definition of compactness. Thinking the set of all $n \times n$ complex matrices as $\mathbb{C}^{n^2}$, the above definition says that G is compact if it is a closed, bounded subset of $\mathbb{C}^{n^2}$. It is a standard theorem from elementary analysis that a subset of $\mathbb{C}^{n^2}$ is compact if and only if it is closed and bounded. All of our examples of matrix Lie groups except $GL_n(\mathbb{R})$ and $GL_n(\mathbb{C})$ have property (1). Thus, it is the boundedness condition (2) that is most important.

The groups $O_n(\mathbb{C})$ and $SO_n(\mathbb{C})$ are compact. Property (1) is satisfied because the limit of orthogonal matrices is orthogonal and the limit of matrices with determinant one has determinant one. Property (2) is satisfied because if A is orthogonal, then the column vectors of A have norm one, and hence $|A_{ij}| < 1$, for all $1 < i, j < n$. A similar argument shows that $U(n)$, $SU(n)$, and $Sp(n)$ are compact, this includes the unit circle, $S^1 \cong U(1)$. On the other hand, the groups $GL_n(\mathbb{R})$ and $GL_n(\mathbb{C})$ are noncompact, since a limit of invertible matrices could be noninvertible. The groups $SL_n(\mathbb{R})$ and $SL_n(\mathbb{C})$ violate boundedness, except in the trivial case $n = 1$. The following groups also violate (2), and hence are noncompact: $O_n(\mathbb{C})$ and $SO_n(\mathbb{C})$; the Heisenberg group H; $Sp_n(\mathbb{R})$ and $Sp_n(\mathbb{C})$; $\mathbb{R}$, $\mathbb{C}$, $\mathbb{R}^\times$ and $\mathbb{C}^\times$.

Definition 2.2.11. A matrix Lie group G is said to be *connected* if given any two matrices A and B in G, there exists a continuous path $\gamma : [0, 1] \to G$, with $\gamma(0) = A$ and $\gamma(1) = B$.

This property is what is called path-connected in topology, which is not (in general) the same as connected. However, it is a fact that a matrix Lie group is connected if and only if it is path-connected. A matrix Lie group G which is not connected can be decomposed (uniquely) as a union of several pieces, called *components*, such that two elements of the same component can be joined by a continuous path, but two elements of different components cannot.

It is a known result that the general linear group $GL_n(\mathbb{C})$ is connected for all $n \geq 1$, but the group $GL_n(\mathbb{R})$ is not. In Table 2.2 we list the examples of matrix Lie groups above, indicating their connectedness properties.

Group	Connected?	Number of Components
$GL_n(\mathbb{C})$	yes	1
$SL_n(\mathbb{C})$	yes	1
$GL_n(\mathbb{R})$	no	2
$SL_n(\mathbb{R})$	yes	1
$O_n(\mathbb{C})$	yes	1
$SO_n(\mathbb{C})$	yes	1
$O_n(\mathbb{R})$	no	2
$SO_n(\mathbb{R})$	yes	1
$U(n)$	yes	1
$SU(n)$	yes	1
Heisenberg	yes	1

Table 2.2: Connectedness properties of some classical matrix Lie groups.

In a similar manner, we can study the notion of simply connected for matrix Lie groups:

Definition 2.2.12. A matrix Lie group G is said to be *simply connected* if it is connected and, in addition, every loop in G can be shrunk continuously to a point in G. More precisely, if given any continuous path $\gamma : [0, 1] \to G$, with $A(0) = A(1)$, there exists a continuous function $H(s, t) : [0, 1] \times [0, 1] \to G$ having the following properties:
(a) $H(s, 0) = H(s, 1)$ for all $s \in [0, 1]$.
(b) $H(0, t) = \gamma(t)$, for all $t \in [0, 1]$.
(c) $H(1, t) = H(1, 0)$, for all $t \in [0, 1]$.

Table 2.3 resumes the examples o matrix Lie groups above, indicating their simply connectedness properties.

We conclude this section with the definition of a Lie group:

Definition 2.2.13. A *Lie group* is a differentiable manifold G which is also a group and such that the group product

$$G \times G \to G$$

Group	Simply connected?	Fundamental group
$SO_2(\mathbb{R})$	no	$\mathbb{Z}$
$SO_n(\mathbb{R})$, $n \geq 3$	no	$2\mathbb{Z}$
$U(n)$	no	$\mathbb{Z}$
$SU(n)$	yes	$\{1\}$
$Sp(n)$	yes	$\{1\}$
$GL_n(\mathbb{C})$	no	$\mathbb{Z}$
$SL_n(\mathbb{C})$	yes	$\{1\}$
$GL_n(\mathbb{R})^+$, $n \geq 2$	no	same as $SO_n(\mathbb{R})$
$SL_n(\mathbb{R})$	no	same as $SO_n(\mathbb{R})$
$SO_n(\mathbb{C})$	no	same as $SO_n(\mathbb{R})$
$Sp_n(\mathbb{C})$	yes	1
$Sp_n(\mathbb{R})$	no	$\mathbb{Z}$

Table 2.3: Simply connectedness properties of some classical matrix Lie groups.

and the inverse map $g \mapsto g^{-1}$ are differentiable.

A manifold is an object that looks locally like a piece of $\mathbb{R}^n$, that is, a topological space that is Hausdorff, second countable, and locally Euclidean. An example would be a torus. For a precise definition, see [128] or [129]. We conclude with a result that establishes whether every matrix Lie group is a Lie group [91, p.52].

Theorem 2.2.14. *Every matrix Lie group is a smooth embedded submanifold of $\mathbb{C}^{n\times n}$ and is thus a Lie group.*

2.3 Lie algebras

A *nonassociative algebra* is a vector space A (over a field $\mathbb{F}$ of characteristic 0) equipped with a bilinear map, called *product* from $A \times A$ to A. The algebra A is called *associative* if it contains an identity element 1 for multiplication, so that

$$1a = a = a1, \ \forall a \in A,$$

and if the associative law $a(bc) = (ab)c$ holds, for all $a, b, c \in A$. The algebra A is said to be *commutative* if the commutative law $ab = ba$ holds, $\forall a, b \in A$.
For subspaces B, C of an algebra A, we write BC for the subspace of A spanned by the products bc, for $b \in B, c \in C$. Similarly, we denote by $[B, C]$ the subspace $BC - CB =$

$\{[b,c] : b \in B, c \in C\}$ spanned by all the commutators $[b,c]$. A *subalgebra* of A is a subspace B of A such that $B^2 \subseteq B$, and $1 \in B$ in the associative case. Equivalently, a subalgebra of A is a subset B of A which is an algebra under the linear and product structures induced from A.

For algebras A and B, a linear map $f : A \to B$ is an *homomorphism* if $f(ab) = f(a)f(b)$, for $a, b \in A$, and if in addition $f(1) = 1$ in the associative case. The homomorphism f is an *isomorphism* if it is a linear isomorphism. In case $A = B$, an homomorphism is called an *endomorphism*, and an isomorphism is called an *automorphism*. Two algebras A and B are said *isomorphic* if there is an isomorphism $f : A \to B$. In this case, we write $A \cong B$. A linear endomorphism $d : A \to A$ of an algebra A is called a *derivation* if for all $x, y \in A$ we have $d(xy) = d(x)y + xd(y)$.

Definition 2.3.1. A *Lie algebra* is a nonassociative algebra $\mathfrak{g}$ such that:
(i) (alternate axiom) $[x,x] = 0$, $\forall x \in \mathfrak{g}$.
(ii) (Jacobi's identity) $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$, $\forall x, y, z \in \mathfrak{g}$.

The alternate property is equivalent to the *skew-symmetry* condition $[x,y] + [y,x] = 0$, $\forall x, y \in \mathfrak{g}$. The Jacobi identity is equivalent to the condition that the *adjoint map*

$$\operatorname{ad} x : \mathfrak{g} \to \mathfrak{g}, \quad \text{given by} \quad \operatorname{ad} x(y) = [x,y]$$

be a derivation, for all $x \in \mathfrak{g}$.

A Lie algebra $\mathfrak{g}$ is called *commutative* or *abelian* if $[\mathfrak{g},\mathfrak{g}] = 0$. Two elements x, y of a Lie algebra $\mathfrak{g}$ are said to *commute* if $[x,y] = 0$. In particular, all pairs of elements of an abelian Lie algebra commute. For Lie algebras $\mathfrak{g}$ and $\mathfrak{k}$, a linear map $f : \mathfrak{g} \to \mathfrak{k}$ is an *homomorphism* if f is an algebra homomorphism and it satisfies

$$f([x,y]) = [f(x), f(y)], \ \forall x, y \in \mathfrak{g}.$$

We define in obvious way the term isomorphism, endomorphism and automorphism.

Example 2.3.2. An associative algebra is a Lie algebra with the bracket $[x,y] = xy - yx$.

Example 2.3.3. Let B be a nonassociative algebra, and set $A = \operatorname{End} B$. Then, the space of derivations of B forms a Lie subalgebra of A.

Example 2.3.4. Every one-dimensional Lie algebra is abelian.

Lie algebras are simpler than matrix Lie groups, because (as we have seen) the Lie algebra is a linear space. Thus, we can understand much about Lie algebras just by doing linear algebra. In fact, to each matrix Lie group G we can associate a Lie algebra, just by considering the exponencial of a matrix. Let X be an $n \times n$ real or complex matrix. We define the *exponential matrix* of X, denoted e^X or $\exp X$, by the usual power series

$$e^X = \sum_{m \geq 0} \frac{X^m}{m!}. \tag{2.2}$$

We know from the differential equation theory that the exponencial series of any matrix $A \in GL_n(\mathbb{C})$ converges uniformly in every compact subset of $\mathbb{C}$, since it is the solution of the homogeneous linear equation given by $X' = AX$, $X(0) = I$ (see [168] for a proof). Also, the exponencial satisfies the usual properties:

1. $\frac{d}{dt}e^{tA} = Ae^{tA}$;
2. $e^{(s+t)A} = e^{sA}e^{tA}$, for all $s, t \in \mathbb{R}$;
3. $e^{tA} = \sum_{m\geq 0} \frac{t^m A^m}{m!}$;
4. $BC = CA$ implies $e^{tB}C = Ce^{tA}$;
5. $e^{t(A+B)} = e^{tA}e^{tB}$ if and only if $AB = BA$. In particular, $(e^{tA})^{-1} = e^{-tA}$;

for all $A, B, C \in GL_n(\mathbb{C})$ and all $s, t \in \mathbb{C}$ (see [168] or [26]).

Definition 2.3.5. Let G be a matrix Lie group in $\mathbb{C}^{n\times n}$. The *Lie algebra* of G, denoted by $\mathfrak{g}$, is the set of all matrices $X \in \mathbb{C}^{n\times n}$ such that e^{tX} is in G for all real numbers t.

This means that X is in $\mathfrak{g}$ if and only if the one-parameter subgroup generated by X lies in G. Note that even though G is a subgroup of $GL_n(\mathbb{C})$ (and not necessarily of $GL_n(\mathbb{R})$), we do not require that e^{tX} be in G for all complex numbers t, but only for all real numbers t. Also, it is not enough to have just $e^X \in G$. It can be given an example of an X and a G such that $e^X \in G$, but such that $e^{tX} \notin G$ for some real values of t. Such an X is not in the Lie algebra of G.
On the other hand, since all matrix Lie group G is itself a Lie group, and thus a smooth manifold (Theorem 2.2.14), we can obtain the tangent space of G on the point I, the identity matrix of G, namely T_IG. It is a well known result that, if X is on T_IG, then $e^{tX} \in G$, for all $t \in \mathbb{R}$ (see [134] for a proof).

For $A, B \in T_IG$, we define the bracket $[\cdot, \cdot]$ by $[A, B] = AB - BA$. Since we can write

$$\begin{aligned}
e^{\sqrt{t}A} &= I + \sqrt{t}A + \tfrac{1}{2}tA^2 + o(t),\\
e^{\sqrt{t}B} &= I + \sqrt{t}B + \tfrac{1}{2}tB^2 + o(t),\\
e^{-\sqrt{t}A} &= I - \sqrt{t}A + \tfrac{1}{2}tA^2 + o(t),\\
e^{-\sqrt{t}B} &= I - \sqrt{t}B + \tfrac{1}{2}tB^2 + o(t),
\end{aligned}$$

and set $\lambda(t) = e^{\sqrt{t}A}e^{\sqrt{t}B}e^{-\sqrt{t}A}e^{-\sqrt{t}B}$, then we have

$$\begin{aligned}\lambda(t) &= e^{\sqrt{t}A}e^{\sqrt{t}B}e^{-\sqrt{t}A}e^{-\sqrt{t}B}\\ &= I + \sqrt{t}\big((A+B)-(A+B)\big) + t\big(A^2+2AB+B^2-(A+B)^2\big)+o(t)\\ &= I + t[A,B]+o(t),\end{aligned}$$

with $\lim\limits_{t\to 0}\frac{o(t)}{t} = 0$. Then, $\lambda'(0) = [A,B]$. Since $\lambda(t) \in G$, for all $t \in \mathbb{R}$, we see that $[A,B] \in T_IG$. We have also proved that T_IG admits a Lie algebra structure. In fact, the Lie algebra $\mathfrak{g}$ of a matrix Lie group G is given by

$$\mathfrak{g} = T_IG,$$

and it directly follows that $T_XG = T_IG \cdot X$, for any $X \in G$ (see [129] or [134]).

Note that we can also see the Lie algebra $\mathfrak{g}$ of a Lie group G as the set of all left-invariant vector fields $X : G \to TG$ (or linear derivations $X : C^\infty(G) \to \mathbb{R}$), with some bracket $[X,Y]$ (see [91]). Because $\mathfrak{g}$ is a real subalgebra of the space $GL_n(\mathbb{C})$ we have the following results, [91, p.55] and [179, p.237]

Proposition 2.3.6. *The Lie algebra $\mathfrak{g}$ associated to the Lie group G is a real Lie algebra.*

Theorem 2.3.7 (Ado)**.** *Every finite-dimensional real Lie algebra is isomorphic to a subalgebra of $\mathfrak{gl}_n(\mathbb{R})$. Every finite-dimensional complex Lie algebra is isomorphic to a subalgebra of $\mathfrak{gl}_n(\mathbb{C})$.*

We present some examples of Lie algebras associated to some Lie groups studied in previous section:

Example 2.3.8 (The general linear groups)**.** If X is any $n \times n$ complex matrix, then e^{tX} is invertible. Thus, the Lie algebra of $GL_n(\mathbb{C})$ is the space of all $n \times n$ complex matrices. This Lie algebra is denoted $\mathfrak{gl}_n(\mathbb{C})$. If X is any $n\times n$ real matrix, then e^{tX} will be invertible and real. On the other hand, if e^{tX} is real for all real numbers t, then $X = \frac{d}{dt}\big(e^{tX}\big)\big|_{t=0}$ will also be real. Thus, the Lie algebra of $GL_n(\mathbb{R})$ is the space of all $n \times n$ real matrices, and is denoted $\mathfrak{gl}_n(\mathbb{R})$.

Example 2.3.9 (The special linear groups)**.** It is a well known restult that $\det e^X = e^{\operatorname{tr} X}$. Thus, if $\operatorname{tr} X = 0$, then $\det e^{tX} = 1$ for all real numbers t. On the other hand, if X is any $n \times n$ matrix such that $\det e^{tX} = 1$ for all t, then $e^{t \operatorname{tr} X} = 1$ for all t. This means that $t \operatorname{tr} X$ is an integer multiple of $2\pi i$ for all t, which is only possible if $\operatorname{tr} X = 0$. Thus, the Lie algebra of $SL_n(\mathbb{C})$ is the space of all $n \times n$ complex matrices with trace zero, denoted $\mathfrak{sl}_n(\mathbb{C})$.
Similarly, the Lie algebra of $SL_n(\mathbb{R})$ is the space of all $n \times n$ real matrices with trace zero, denoted $\mathfrak{sl}_n(\mathbb{R})$.

Example 2.3.10 (The unitary groups)**.** Recall that a matrix U is unitary if and only if $U^* = U^{-1}$. Thus, e^{tX} is unitary if and only if

$$(e^{tX})^* = (e^{tX})^{-1} = e^{-tX}. \tag{2.3}$$

Since $(e^{tX})^* = e^{tX^*}$, then (2.3) becomes $e^{tX^*} = e^{-tX}$. Clearly, a sufficient condition for this last identity to hold is that $X^* = -X$. On the other hand, if $e^{tX^*} = e^{-tX}$ holds for all t, then by differentiating at $t = 0$, we see that $X^* = -X$ is necessary. Thus, the Lie algebra of $U(n)$ is the space of all $n \times n$ complex matrices X such that $X^* = -X$, and is denoted by $\mathfrak{u}(n)$.

Combining the two previous computations, we see that the Lie algebra of $SU(n)$ is the space of all $n \times n$ complex matrices X such that $X^* = -X$ and $\operatorname{tr} X = 0$, denoted $\mathfrak{su}(n)$.

Example 2.3.11 (The orthogonal groups)**.** The identity component of $O(n)$ is just $SO(n)$. Since the exponential of a matrix in the Lie algebra is automatically in the identity component, the Lie algebra of $O(n)$ is the same as the Lie algebra of $SO(n)$. Now, an $n \times n$ real matrix R is orthogonal if and only if $R^{\text{tr}} = R^{-1}$. So, given an $n \times n$ real matrix X, e^{tX} is orthogonal if and only if $(e^{tX})^{\text{tr}} = (e^{tX})^{-1}$, or

$$e^{tX^{\text{tr}}} = e^{-tX}. \tag{2.4}$$

Clearly, a sufficient condition for this to hold is that $X^{\text{tr}} = -X$. If (2.4) holds for all t, then by differentiating at $t = 0$, we must have $X^{\text{tr}} = -X$. Thus, the Lie algebra of $O(n)$, as well as the Lie algebra of $SO(n)$, is the space of all $n \times n$ real matrices X with $X^{\text{tr}} = -X$, denoted by $\mathfrak{so}_n(\mathbb{R})$ (or simply $\mathfrak{so}(n)$). Note that the condition $X^{\text{tr}} = -X$ forces the diagonal entries of X to be zero, and so, necessarily the trace of X is zero.

The same argument shows that the Lie algebra of $SO_n(\mathbb{C})$ is the space of $n \times n$ complex matrices satisfying $X^{\text{tr}} = -X$, denoted by $\mathfrak{so}_n(\mathbb{C})$. This is not the same as $\mathfrak{su}(n)$.

Example 2.3.12 (The symplectic groups)**.** These are denoted $\mathfrak{sp}_n(\mathbb{R})$, $\mathfrak{sp}_n(\mathbb{C})$, and $\mathfrak{sp}(n)$. The calculation of these Lie algebras is similar to that of the generalized orthogonal groups, and we will just record the result here. Let J be the matrix in the definition of the symplectic groups. Then, $\mathfrak{sp}_n(\mathbb{R})$ is the space of $2n \times 2n$ real matrices X such that $JX^{\text{tr}}J = X$, $\mathfrak{sp}_n(\mathbb{C})$ is the space of $2n \times 2n$ complex matrices satisfying the same condition, and $\mathfrak{sp}(n) = \mathfrak{sp}_n(\mathbb{C}) \cap \mathfrak{u}(2n)$.

Example 2.3.13 (The Heisenberg group)**.** Recall that the Heisenberg group H is the group of all 3×3 real matrices A of the form

$$\begin{pmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{where } a, b, c \in \mathbb{R}.$$

Computing the exponential of any matrix of the form

$$X = \begin{pmatrix} 0 & \alpha & \beta \\ 0 & 0 & \gamma \\ 0 & 0 & 0 \end{pmatrix}, \quad \text{where } \alpha, \beta, \gamma \in \mathbb{R}, \tag{2.5}$$

we can see that e^{tX} is in H. On the other hand, if X is any matrix such that e^{tx} is in H for all t, then all of the entries of $X = \frac{d}{dt}e^{tX}\big|_{t=0}$ which are on or below the diagonal must be zero, so that X is of form (2.5). Thus, the Lie algebra of the Heisenberg group is the space of all 3 × 3 real matrices that are strictly upper triangular.

Definition 2.3.14. If V is a finite-dimensional real vector space, then the *complexification* of V, denoted $V_{\mathbb{C}}$, is the space of formal linear combinations

$$v_1 + iv_2, \quad \text{with } v_1, v_2 \in V.$$

This becomes a real vector space in the obvious way and becomes a complex vector space if we define $i(v_1 + iv_2) = -v_2 + iv_1$. We will regard V as a real subspace of $V_{\mathbb{C}}$ in the obvious way. Now let $\mathfrak{g}$ be a finite-dimensional real Lie algebra and $\mathfrak{g}_{\mathbb{C}}$ its complexification (as a real vector space). Then, the bracket operation on $\mathfrak{g}$ has a unique extension to $\mathfrak{g}_{\mathbb{C}}$ which makes $\mathfrak{g}_{\mathbb{C}}$ into a complex Lie algebra. In fact, the uniqueness of the extension is obvious, since if the bracket operation on $\mathfrak{g}_{\mathbb{C}}$ is to be bilinear, then it must be given by

$$[x_1 + ix_2, y_1 + iy_2] = \big([x_1, y_1] - [x_2, y_2]\big) + i\big([x_1, y_2] + [x_2, y_1]\big). \tag{2.6}$$

This bracket defined above is really bilinear and skew symmetric and it satisfies the Jacobi identity. It is clear from (2.6) is real bilinear, and skew-symmetric. The skew symmetry means that if (2.6) is complex linear in the first factor, it is also complex linear in the second factor. Thus, we need only show that

$$[i(x_1 + ix_2), y_1 + iy_2] = i[x_1 + ix_2, y_1 + iy_2]. \tag{2.7}$$

The left-hand side of (2.7) is $[-x_2 + ix_1, y_1 + iy_2] = (-[x_2, y_1] - [x_1, y_2]) + i([x_1, y_1] + [x_2, y_2])$, whereas the right-hand side of (2.7) is just

$$i\big([x_1, y_1] - [x_2, y_2]\big) + i\big([x_1, y_2] + [x_2, y_1]\big) = (-[x_2, y_1] - [x_1, y_2]) + i([x_1, y_1] + [x_2, y_2]),$$

and, indeed, these are equal. To check the Jacobi identity, note that the Jacobi identity holds if x, y, and z are in $\mathfrak{g}$. However, observe that the expression on the left-hand side of the Jacobi identity in Definition 2.3.1 is (complex!) linear in x for fixed y and z. It follows that the Jacobi identity holds if x is in $\mathfrak{g}_{\mathbb{C}}$, and y and z are in $\mathfrak{g}$. The same argument shows that we can extend to y in $\mathfrak{g}_{\mathbb{C}}$, and then to z in $\mathfrak{g}_{\mathbb{C}}$. Thus, the Jacobi identity holds in $\mathfrak{g}_{\mathbb{C}}$. Thus, we have

Proposition 2.3.15. *Let $\mathfrak{g}$ be a finite-dimensional real Lie algebra and $\mathfrak{g}_{\mathbb{C}}$ its complexification (as a real vector space). Then, the bracket operation on $\mathfrak{g}$ has a unique extension to $\mathfrak{g}_{\mathbb{C}}$ which makes $\mathfrak{g}_{\mathbb{C}}$ into a complex Lie algebra.*

Definition 2.3.16. The complex Lie algebra $\mathfrak{g}_{\mathbb{C}}$ is called the *complexification* of the real Lie algebra $\mathfrak{g}$.

We give the complexifications of some real Lie algebras in Table 2.4

Lie algebra	Complexification
$\mathfrak{gl}_n(\mathbb{R})$	$\mathfrak{gl}_n(\mathbb{R})_{\mathbb{C}} = \mathfrak{gl}_n(\mathbb{C})$
$\mathfrak{u}(n)$	$\mathfrak{u}(n)_{\mathbb{C}} = \mathfrak{gl}_n(\mathbb{C})$
$\mathfrak{su}(n)$	$\mathfrak{su}(n)_{\mathbb{C}} = \mathfrak{sl}_n(\mathbb{C})$
$\mathfrak{sl}_n(\mathbb{R})$	$\mathfrak{sl}_n(\mathbb{R})_{\mathbb{C}} = \mathfrak{sl}_n(\mathbb{C})$
$\mathfrak{so}(n)$	$\mathfrak{so}(n)_{\mathbb{C}} = \mathfrak{so}_n(\mathbb{C})$
$\mathfrak{sp}_n(\mathbb{R})$	$\mathfrak{sp}_n(\mathbb{R})_{\mathbb{C}} = \mathfrak{sp}_n(\mathbb{C})$
$\mathfrak{sp}(n)$	$\mathfrak{sp}(n)_{\mathbb{C}} = \mathfrak{sp}_n(\mathbb{C})$

Table 2.4: Complexifications of some classical real Lie algebras.

2.4 Modules

Definition 2.4.1. Let A be an associative algebra and let V be a vector space. We say that V is an A-*module* if there is a bilinear map $A \times V \to V$ (denoted by a dot $(a, v) \mapsto a \cdot v$) such that
(i) [identity] $1 \cdot v = v$, for all $v \in V$.
(ii) [associativity] $(ab) \cdot v = a \cdot (b \cdot v)$, for all $a, b \in A$, and all $v \in V$.

For $a \in A$, let π_a the corresponding linear endomorphism of V, such that $\pi_a(v) = a \cdot v$. Then, the map

$$\pi : A \to \operatorname{End} V, \quad \text{such that } \pi(a) = \pi_a$$

is an homomorphism of associative algebras. Such an homomorphism is called a *representation* of A on V. Sometimes, V is called a *representation* of A.

Example 2.4.2. Note that the associative algebra A has a natural representation on itself, given by the left multiplication action: $a \cdot b = ab$, $\forall a, b \in A$.

Definition 2.4.3. Analogously, let $\mathfrak{g}$ be a Lie algebra and let V be a vector space. Then V is called a $\mathfrak{g}$-*module* if there is a bilinear map $\mathfrak{g} \times V \to V$ (denoted by a dot $(x, v) \mapsto x \cdot v$) such that
(i) (bracket preserving) $[x, y] \cdot v = x \cdot (y \cdot v) - y \cdot (x \cdot v)$, for all $x, y \in \mathfrak{g}$, and all $v \in V$.

If for $x \in \mathfrak{g}$, we denote π_x the corresponding linear endomorphism of V, such that

$$\pi_x(v) = x \cdot v.$$

Then, the map

$$\pi : \mathfrak{g} \to \operatorname{End} V, \quad \text{such that } \pi(x) = \pi_x$$

is a Lie algebra homomorphism. Such an homomorphism is called a *representation* of $\mathfrak{g}$ on V (sometimes V is called a *representation* of $\mathfrak{g}$). The notions of $\mathfrak{g}$-module and representation

of $\mathfrak{g}$ are equivalent.
The Lie algebra $\mathfrak{g}$ has a natural representation on itself, called the *adjoint representation*, by the map

$$\operatorname{ad} : \mathfrak{g} \to \operatorname{End} \mathfrak{g}, \quad \text{such that} \quad x \mapsto \operatorname{ad} x.$$

In particular, taking $\mathfrak{g} = \operatorname{End} V$, every Lie subalgebra of $\operatorname{End} V$ has a natural representation on V.

Let $\mathfrak{g}$ be an associative or Lie algebra, and let V be a $\mathfrak{g}$-module. For subspaces $\mathfrak{h}$ of $\mathfrak{g}$ and W of V, we denote $\mathfrak{h} \cdot W$ the linear span of all $x \cdot w$, for $x \in \mathfrak{h}, w \in W$. A *submodule* of V is a subspace W of V such that $\mathfrak{g} \cdot W \subseteq W$, or equivalently, a subset W of V which is a $\mathfrak{g}$-module under the linear structure and $\mathfrak{g}$-module action induced from V. A subspace W of V is *invariant* (under $\mathfrak{g}$) if it is a submodule. The module V is said *irreducible* or *simple* if $V \neq 0$ and if V has no proper nonzero invariant subspaces. The module V is called *indecomposable* if it cannot be decomposed as a direct sum of two nonzero submodules. Clearly, an irreducible module is indecomposable. Let V and W be $\mathfrak{g}$-modules. A linear map $f : V \to W$ is called a $\mathfrak{g}$-*module homomorphism* or $\mathfrak{g}$-*module map* if

$$f(x \cdot v) = x \cdot f(v), \;\; \forall x \in \mathfrak{g}, \forall v \in V.$$

Such a map f is called a $\mathfrak{g}$-*module isomorphism* or $\mathfrak{g}$-*module equivalence* if it is a linear isomorphism. Two modules V and W are *isomorphic* or *equivalent* if there is an isomorphism $f : V \to W$. In that case we write $V \cong W$.

Definition 2.4.4. A subspace $\mathfrak{a}$ of a Lie algebra $\mathfrak{g}$ is called an *ideal* of $\mathfrak{g}$ if $[\mathfrak{g}, \mathfrak{a}] \subseteq \mathfrak{a}$. Equivalently, an ideal of $\mathfrak{g}$ is a submodule under the adjoint representation.

An ideal is a subalgebra. Given an ideal $\mathfrak{a}$ of $\mathfrak{g}$, the quotient vector space $\mathfrak{g}/\mathfrak{a}$ becomes a Lie algebra, called the *quotient Lie algebra*, by means of the well defined nonassociative product

$$[x + \mathfrak{a}, y + \mathfrak{a}] = [x, y] + \mathfrak{a}, \;\; \forall x, y \in \mathfrak{g}.$$

The canonical map $\pi : \mathfrak{g} \to \mathfrak{g}/\mathfrak{a}$ is an homomorphism, and we have an exact sequence of Lie algebras

$$0 \longrightarrow \mathfrak{a} \longrightarrow \mathfrak{g} \xrightarrow{\pi} \mathfrak{g}/\mathfrak{a} \longrightarrow 0.$$

A Lie algebra $\mathfrak{g}$ is said to be *simple* if $\mathfrak{g}$ is nonzero and has no proper nonzero ideals (equivalently, the adjoint representation is simple), and if $\dim \mathfrak{g} > 1$ (*i. e.*, $\mathfrak{g}$ is not abelian).

Remember that a subspace I of an associative algebra A is called a *left ideal* (respectively *right ideal*) of A if $AI \subseteq I$ (respectively $IA \subseteq A$), and an *ideal* if is both left and right ideal. Since an ideal I need not to contain 1, it need not be a subalgebra. Given and ideal I of A, the quotient vector space A/I becomes an associative algebra —the *quotient algebra*— in an obvious way.

Example 2.4.5. The kernel of any homomorphism of a Lie (respectively associative) algebra $\mathfrak{g}$ is an ideal.

In particular, the kernel of the adjoint representation of a Lie algebra $\mathfrak{g}$ is an important ideal called the *center* of $\mathfrak{g}$ and denoted by $\operatorname{Cent}\mathfrak{g}$:

$$\operatorname{Cent}\mathfrak{g} = \{x \in \mathfrak{g} : [x, \mathfrak{g}] = 0\}.$$

Any subspace of $\operatorname{Cent}\mathfrak{g}$ is an ideal in $\mathfrak{g}$ and is said to be a *central ideal.*

Definition 2.4.6. Given the Lie algebras $\mathfrak{a}$ and $\mathfrak{b}$, an *extension of* $\mathfrak{a}$ *by* $\mathfrak{b}$ is a Lie algebra $\mathfrak{g}$ together with an exact sequence

$$0 \longrightarrow \mathfrak{b} \longrightarrow \mathfrak{g} \longrightarrow \mathfrak{a} \longrightarrow 0,$$

(note that $\mathfrak{b}$ is an ideal and $\mathfrak{g}/\mathfrak{b} \cong \mathfrak{a}$). This extension is said to be *central* if $\mathfrak{b}$ is a central ideal of $\mathfrak{g}$.

Two extensions $\mathfrak{g}_1$ and $\mathfrak{g}_2$ of $\mathfrak{a}$ by $\mathfrak{b}$ are *equivalent* if there is an isomorphism $\mathfrak{g}_1 \cong \mathfrak{g}_2$, making the following diagram commute

$$\begin{tikzcd}
 & & \mathfrak{g}_1 \arrow[dd, "\cong"] \arrow[dr] & & \\
0 \arrow[r] & \mathfrak{b} \arrow[ur] \arrow[dr] & & \mathfrak{a} \arrow[r] & 0. \\
 & & \mathfrak{g}_2 \arrow[ur] & &
\end{tikzcd}$$

If $\mathfrak{a}$ and $\mathfrak{b}$ are ideals of a Lie algebra $\mathfrak{g}$, then $\mathfrak{a} + \mathfrak{b}$, $\mathfrak{a} \cap \mathfrak{b}$ and $[\mathfrak{a}, \mathfrak{b}]$ are ideals also. In particular, $[\mathfrak{g}, \mathfrak{g}]$ is an ideal of $\mathfrak{g}$, called the *commutator ideal*, and is denoted by

$$\mathfrak{g}' = [\mathfrak{g}, \mathfrak{g}].$$

Given two Lie algebras $\mathfrak{a}$ and $\mathfrak{b}$, their *direct product* is the Lie algebra $\mathfrak{a} \times \mathfrak{b}$, which is $\mathfrak{a} \oplus \mathfrak{b}$ as a vector space, with $\mathfrak{a}$ and $\mathfrak{b}$ retaining their original bracket structures and commuting with one another. In particular, $\mathfrak{a}$ and $\mathfrak{b}$ are ideals of $\mathfrak{a} \times \mathfrak{b}$. We define the direct product of finitely many Lie algebras analogously.

More generally, suppose that we have a representation $\pi : \mathfrak{a} \to \operatorname{End}\mathfrak{b}$ of a Lie algebra $\mathfrak{a}$ on a Lie algebra $\mathfrak{b}$ by derivations, *i. e.*, $\pi(x) = \pi_x$ is a derivation of $\mathfrak{b}$, for all $x \in \mathfrak{a}$. Then, the vector space $\mathfrak{a} \oplus \mathfrak{b}$ carries a unique Lie algebra structure such that $\mathfrak{a}$ and $\mathfrak{b}$ are subalgebras and $[x, y] = \pi_x(y)$, for all $x \in \mathfrak{a}, y \in \mathfrak{b}$ (observe that $\mathfrak{b}$ is an ideal, but not necessarily $\mathfrak{a}$). This Lie algebra is called the *semidirect product* of $\mathfrak{a}$ and $\mathfrak{b}$, and we denote this by $\mathfrak{a} \ltimes \mathfrak{b}$. A Lie algebra is a semidirect product whenever it is a vector space direct sum of a subalgebra and a ideal. In fact, $\mathfrak{a} \ltimes \mathfrak{b} = \mathfrak{a} \times \mathfrak{b}$ if and only if $\pi \equiv 0$. A semidirect product $\mathfrak{a} \ltimes \mathfrak{b}$ is an extension of $\mathfrak{a}$ by $\mathfrak{b}$. An extension of a Lie algebra $\mathfrak{a}$ by a Lie algebra $\mathfrak{b}$ is said *trivial* if it is equivalent to a semidirect product $\mathfrak{a} \ltimes \mathfrak{b}$.

Example 2.4.7. We mention a particular kind of semidirect product. Given a Lie algebra $\mathfrak{g}$ and a derivation d of $\mathfrak{g}$, we can form $\mathbb{F}d \ltimes \mathfrak{g}$. This procedure is called *adjoining the derivation* d to $\mathfrak{g}$.

Definition 2.4.8. Let S be a set. A vector space V is said to be S-*graded* if it is the direct sum

$$V = \bigoplus_{\alpha \in S} V_\alpha,$$

of subspaces V_α. In this case, the elements of V_α are said to be *homogeneous of degree* α, and V_α is called the *homogeneous subspace of degree* α or the α-*graded component* of V. For $v \in V_\alpha$ (including $v = 0$) we write

$$\deg v = \alpha.$$

Given another S-graded vector space W, a linear map $f : V \to W$ is called *grading preserving* if

$$f : V_\alpha \to W_\alpha, \quad \text{for all } \alpha \in S.$$

If such a map is a linear isomorphism, V and W are said to be *graded-isomorphic.* If S is an abelian group, a linear map $f : V \to W$ is said to be *homogeneous of degree* β if

$$f : V_\alpha \to W_{\alpha+\beta}, \quad \text{for all } \alpha \in S. \tag{2.8}$$

In that case, we write $\deg f = \beta$. Note that f is grading-preserving if and only if $\deg f = 0$. If S is an abelian group which is a subgroup of the additive group of $\mathbb{F}$, we can define the *degree operator* $d : V \to W$ by the condition

$$d(v) = \alpha v, \text{ for } v \in V_\alpha, \alpha \in S. \tag{2.9}$$

Note that a linear map $f : V \to W$ is grading-preserving if and only if $[d, f] = 0$, and is homogeneous of degree β if and only if $[d, f] = \beta f$.

A subspace W of a S-graded vector space V is *graded* if $W = \bigoplus_{\alpha \in S} W_\alpha$, where $W_\alpha = W \cap V_\alpha$, for $\alpha \in S$. In this case, the quotient V/W is graded in a natural way:

$$V/W = \bigoplus_{\alpha \in S} (V/W)_\alpha = \bigoplus_{\alpha \in S} V_\alpha / W_\alpha.$$

Given a family of S-graded vector spaces $\{V^i\}_i$, the direct sum $X = \bigoplus_i V^i$ is naturally S-graded if we take

$$X_\alpha = \bigoplus_i V^i_\alpha, \text{ for } \alpha \in S.$$

2.5 Tensor products

Definition 2.5.1. Let V be a vector space over a field $\mathbb{F}$ with a basis $\{e_i\}_i$ and let W be a vector space over $\mathbb{F}$ with a basis $\{f_j\}_j$. The *tensor product* $V \otimes W$ is a vector space over $\mathbb{F}$, with a basis $\{e_i \otimes f_j\}_{i,j}$.
If $x = \sum_i x_i e_i \in V$ and $y = \sum_j y_j f_j \in W$, we have a *tensor multiplication* defined by

$$x \otimes y = \sum_{i,j} x_i y_j \, (e_i \otimes f_j) \in V \otimes W.$$

The *tensor map* $\tau : V \times W \to V \otimes W$ defined by $(x, y) \mapsto x \otimes y$ satisfies bilinearity:
(i) $(x + y) \otimes z = (x \otimes z) + (y \otimes z)$;
(ii) $x \otimes (y + z) = (x \otimes y) + (x \otimes z)$;
(iii) $(\lambda x) \otimes y = \lambda(x \otimes y) = x \otimes (\lambda y)$.

Conversely, if β is a bilinear mapping of $V \times W$ into another vector space X, then $\beta(x, y) = \beta\big(\sum_i x_i e_i, \sum_j y_j f_j\big) = \sum_{i,j} x_i y_j \beta(e_i, f_j)$ by bilinearity, and the linear transformation $T : V \otimes W \to X$ that sends $e_i \otimes f_j$ to $\beta(e_i, f_j)$ also sends $x \otimes y$ to $\beta(x, y)$, for all $x \in V, y \in W$. In fact, T is the unique linear transformation such that $\beta = T \circ \tau$. In this sense, every bilinear mapping of $V \times W$ factors uniquely through β (universal property):

$$\begin{tikzcd}
V \times W \arrow[r, "T"] \arrow[d, "\tau"'] & X \\
V \otimes W \arrow[ur, "\beta"'] &
\end{tikzcd}$$

Tensor products of modules over a commutative ring R can be defined by the same universal property (which does not require bases). More explicitly, we have [88, p.434]

Proposition 2.5.2 (Universal property for tensor products)**.** *Let R be a commutative ring and let A, B, C be R-modules. For a mapping $\beta : A \times B \to C$ the following conditions are equivalent:*

1. *β is bilinear;*
2. *$a \mapsto \beta(a, \cdot)$ is a module homomorphism of A into* $\mathrm{Hom}_R(B, C)$*;*
3. *$b \mapsto \beta(\cdot, b)$ is a module homomorphism of B into* $\mathrm{Hom}_R(A, C)$*.*

Bilinear mappings can be defined in the same way for left R-modules over an arbitrary ring R, but then lose properties (2) and (3) above, if only because $\mathrm{Hom}_R(A, C)$ and $\mathrm{Hom}_R(B, C)$ are only abelian groups. It is more fruitful to keep properties (2) and (3), and to forgot bilinearity unless R is commutative. In the simplest form of (2) and (3), C is an abelian group; if B is a left R-module, then $\mathrm{Hom}_{\mathbb{Z}}(B, C)$ is a right R-module and A needs to be a right R-module; so B and $\mathrm{Hom}_{\mathbb{Z}}(A, C)$ are left R-modules. Also, [88, p.435]

Proposition 2.5.3. *Let R be a ring, let A be a right R-module, let B be a left R-module, and let C be an abelian group. For a mapping $\beta : A \times B \to C$ the following conditions are equivalent:*

1. *For all $a, a' \in A$, $b, b' \in B$ and $r \in R$*
   (i) [biadditive] $\beta(a + a', b) = \beta(a, b) + \beta(a', b)$, $\beta(a, b + b') = \beta(a, b) + \beta(a, b')$,
   (ii) [balanced] $\beta(ar, b) = \beta(a, rb)$;

2. *$a \mapsto \beta(a, \cdot)$ is a module homomorphism of A into $\mathrm{Hom}_{\mathbb{Z}}(B, C)$;*

3. *$b \mapsto \beta(\cdot, b)$ is a module homomorphism of B into $\mathrm{Hom}_{\mathbb{Z}}(A, C)$.*

We say that a *bihomomorphism* of modules is a mapping that satisfies the equivalent conditions in Proposition 2.5.3. For example, the left action $(r, x) \mapsto rx$ of R on any left R-module M is a bihomomorphism of $R \times M$ into the underlying abelian group M. If $\beta : A \times B \to C$ is a bihomomorphism and $\varphi : C \to D$ is an homomorphism of abelian groups, then $\varphi \circ \beta : A \times B \to D$ is a bihomomorphism. The tensor product of A and B is an abelian group $A \otimes_R B$ with a bihomomorphism τ of $A \times B$, from which every bihomomorphism of $A \otimes B$ can be recovered uniquely in this fashion:

$$\begin{tikzcd}
A \times B \arrow[d, "\tau"'] \arrow[dr, "\beta"] & \\
A \otimes_R B \arrow[r, "\bar{\beta}"'] & C.
\end{tikzcd}$$

Definition 2.5.4. Let A be a right R-module and let B be a left R-module. A *tensor product* of A and B is an abelian group $A \otimes_R B$ together with a bihomomorphism $\tau : A \times B \to A \otimes_R B$ given by $(a, b) \mapsto a \otimes b$, the *tensor map*, such that, for every abelian group C and bihomomorphism $\beta : A \times B \to C$ there exists a unique homomorphism $\bar{\beta} : A \otimes_R B$ of abelian groups such that $\beta = \bar{\beta} \circ \tau$.

If S is an abelian group and if V and W are S-graded vector spaces, then $V \otimes W$ acquires a unique S-grading by the condition

$$V_\alpha \otimes W_\beta \subseteq (V \otimes W)_{\alpha+\beta}, \text{ for } \alpha, \beta \in S.$$

Using the symbol d_U for the degree operator in the space U, we have

$$d_{V \otimes W} = d_V \otimes 1 + 1 \otimes d_W.$$

This *tensor product grading* extends to an arbitrary finite number of tensor factors.
Now, let G be an abelian group and let A be a nonassociative algebra. Then A is a *G-graded* algebra if it is G-graded as a vector space, so that

$$A = \bigoplus_{\alpha \in G} A_\alpha,$$

and if $A_\alpha A_\beta \subseteq A_{\alpha+\beta}$, for $\alpha, \beta \in U$.

2.6 Module construction

Fix an associative Lie algebra $\mathfrak{g}$. Let V a $\mathfrak{g}$-module and $U \subseteq V$ a submodule. Then the quotient vector space V/U becomes a $\mathfrak{g}$-module, called the *quotient module*, by means of the well defined action

$$x \cdot (v + U) = x \cdot v + U, \quad \text{for } x \in \mathfrak{g}, v \in V.$$

We have an exact sequence of $\mathfrak{g}$-modules

$$0 \longrightarrow U \longrightarrow V \longrightarrow V/U \longrightarrow 0.$$

Given two $\mathfrak{g}$-modules V_1 and V_2, their *direct sum* $V_1 \oplus V_2$ is the $\mathfrak{g}$-module which is $V_1 \oplus V_2$ as a vector space, with V_1 and V_2 retaining their original module structures. In particular, V_1 and V_2 are submodules of $V_1 \oplus V_2$. The direct sum of any collection $\{V_i\}_i$ of $\mathfrak{g}$-modules is defined analogously and is denoted by $\bigoplus_i V_i$.

A $\mathfrak{g}$-module is called *completely reducible* or *semisimple* if it is a direct sum of irreducible submodules (here the null sum is allowed, so that the zero-dimensional module is considered completely reducible). Let G an abelian group and suppose that $\mathfrak{g}$ is G-graded. A $\mathfrak{g}$-module V is *G-graded* if it is G-graded as a vector space, so that $V = \bigoplus_{\alpha \in G} V_\alpha$, and if

$$\mathfrak{g}_\alpha \cdot V_\beta \subseteq V_{\alpha+\beta}, \text{ for } \alpha, \beta \in G,$$

i. e., $\mathfrak{g}_\alpha$ acts as operators of degree α (see (2.8)). Quotients and direct sums of G-graded modules are graded (as modules). In case $\mathfrak{g}$ is a G-graded Lie algebra, with G a subgroup of the additive group of $\mathbb{F}$, let d be the degree derivation of $\mathfrak{g}$. Then a G-graded $\mathfrak{g}$-module V becomes an $\mathbb{F}d \ltimes \mathfrak{g}$-module when d is required to act as the degree operator (2.9) on V (observe that d plays two different compatible roles).
The grading of a graded module can be shifted in the following sense. Suppose that G is a subgroup of an abelian group B and that V is a G-graded A-module, A a G-graded nonassociative algebra. Let $\beta \in B$. Then for each $\alpha \in G$, V_α can be renamed $V_{\alpha+\beta}$, giving V the structure of a B-graded module with $A_\gamma = 0$ for $\gamma \in B - G$ and $V_\gamma = 0$ for $\gamma \in B - (G + \beta)$.

Now let $\mathfrak{g}$ be a Lie algebra. In preparation for constructing the tensor product of $\mathfrak{g}$-modules, we first note that if π_1 and π_2 are two representations of $\mathfrak{g}$ on V which commute in the sense that

$$[\pi_1(x), \pi_2(y)] = 0, \quad \text{for all } x, y \in \mathfrak{g},$$

then $\pi_1 + \pi_2$ is a representation of $\mathfrak{g}$ on V.

Definition 2.6.1. Given two $\mathfrak{g}$-modules V and W, we define the *tensor product module* $V \otimes W$ to be the vector space $V \otimes W$ with the action of $x \in \mathfrak{g}$ determined by the condition

$$x \cdot (v \otimes w) = (x \cdot v) \otimes w + v \otimes (x \cdot w), \text{ for } v \in V, w \in W.$$

This is a $\mathfrak{g}$-module action because the equations

$$x \cdot (v \times w) = (x \cdot v) \otimes w \quad \text{and} \quad x \cdot (v \times w) = v \otimes (x \cdot w)$$

clearly define two commuting $\mathfrak{g}$-module structures on the vector space $V \otimes W$. The tensor product of finitely many $\mathfrak{g}$-modules is defined analogously. If the tensor factors are G-graded modules (G an abelian group), then so is the tensor product.

2.7 Induced modules

Let B a subalgebra of an associative algebra A and let V be a B-module. We denote by $A \otimes_B V$ the quotient of the vector space $A \otimes_{\mathbb{F}} V$ by the subspace spanned by the elements $ab \otimes v - a \otimes b \cdot v$, for $a \in A, b \in B, v \in V$, and we again write $a \otimes b$ for the image of $a \otimes v \in A \otimes_{\mathbb{F}} V$ in $A \otimes_B V$. The space $A \otimes_B V$ carries a natural A-module structure determined by the condition

$$c \cdot (a \otimes v) = ca \otimes v \quad \text{for} \quad a, c \in A, v \in V,$$

and $A \otimes_B V$ is called the A-module *induced* by the B-module V. It is sometimes denoted as $\operatorname{Ind}_B^A V$.

There is a canonical B-module map $i : V \to A \otimes_B V$, given by $v \mapsto 1 \otimes v$, and $\operatorname{Ind}_B^A V$ has the following universal property: Given any A-module W and B-module map $j : V \to W$, there is a unique A-module map $f : \operatorname{Ind}_B^A V \to W$ making the following diagram commute

$$\begin{tikzcd}
\operatorname{Ind}_B^A V \arrow[r, "f"] & W \\
V \arrow[u, "i"] \arrow[ur, "j"'] &
\end{tikzcd}$$

This property characterizes the A-module $\operatorname{Ind}_B^A V$ and the map i up to canonical isomorphism. In fact, if I' is another A-module with a B-module map $i' : V \to I$ satisfying the same condition, the we obtain A-module maps $f : \operatorname{Ind}_B^A V \to I$, $g : I' \to \operatorname{Ind}_B^A V$. But $g \circ f$ and the identity map both make the diagram

$$\begin{tikzcd}
\operatorname{Ind}_B^A V \arrow[r] & \operatorname{Ind}_B^A V \\
V \arrow[u, "i"] \arrow[ur, "i"'] &
\end{tikzcd}$$

commute, so that $g \circ f$ is the identity map by uniqueness. Similarly, $f \circ g$ is the identity on I'. If the algebra A and subalgebra B are G-graded and if V is a G-graded B-module (G an abelian group), then the induced module $\operatorname{Ind}_B^A V$ is a G-graded A-module in a natural way.

Given a group G we define its *group algebra* to be the associative algebra $\mathbb{F}[G]$, which is formally the set of finite linear combinations of elements of G. That is, $\mathbb{F}[G]$ has the set G as a linear basis, and multiplication in $\mathbb{F}[G]$ is simply defined by linear extension of multiplication in G. The identity element of $\mathbb{F}[G]$ is just the identity element of G.

Definition 2.7.1. A *representation* of the group G on a vector space V is a group homomorphism $\pi : G \to \operatorname{Aut} V$. The space V is called a *G-module* or *representation* of G, and just as for associative and Lie algebras, we often use the dot notation

$$g \cdot v = \pi(g)v \quad \text{for} \quad g \in G, v \in V.$$

We have $1 \cdot v = v$ and $(gh) \cdot v = g \cdot (h \cdot v)$, for all $g, h \in G$, and all $v \in V$. We have the usual module-theoretic concepts such as irreducibility and equivalence. If $\pi(G) = 1$, π is called a *trivial* representation. Given G-modules $V_1, \ldots, V_n$, their *tensor product* is the vector space $V_1 \otimes \ldots \otimes V_n$, with G-action determined by

$$g \cdot (v_1 \otimes \ldots \otimes v_n) = (g \cdot v_1) \otimes \ldots \otimes (g \cdot v_n) \quad \text{for} \quad g \in G, v_i \in V_i.$$

The group G has a natural representation on its own group algebra, given by the left multiplication action. This is called the *left regular representation* of G.

Any G-module V becomes a $\mathbb{F}[G]$-module in a canonical way —by extending the map $\pi : G \to \operatorname{Aut} V$ by linearity to an algebra homomorphism of $\mathbb{F}[G]$ to $\operatorname{End} V$—. In fact, the G-modules are essentially the same as the $\mathbb{F}[G]$-modules. For example, the left regular representation of G corresponds to the left multiplication representation of $\mathbb{F}[G]$. If the group G is an abelian group written additively, such as the group $\mathbb{Z}$, there can be confusion as to whether the symbol $a + b$ means the sum in G or the sum in $\mathbb{F}[G]$, for $a, b \in G$. For this reason we use the exponential notation for the elements of G viewed as elements of $\mathbb{F}[G]$ when G is such a group: we write e^a for the element of $\mathbb{F}[G]$ corresponding to $a \in G$. In particular, $e^0 = 1$ and $e^a e^b = e^{a+b}$, for $a, b \in G$.

Given a subgroup H of a group G and a H-module V, we define the G-module *induced* by V to be the G-module associated with the induced $\mathbb{F}[G]$-module $\mathbb{F}[G] \otimes_{\mathbb{F}[H]} V$. We write

$$\operatorname{Ind}_H^G V = \mathbb{F}[G] \otimes_{\mathbb{F}[H]} V.$$

There is a canonical H-module map $i : V \to \mathbb{F}[G] \otimes_{\mathbb{F}[H]} V$ given by $v \mapsto 1 \otimes v$, and the induced module is characterized by the following universal property: Given any G-module W and H-module map $j : V \to W$, there is a unique G-module map $f : \operatorname{Ind}_H^G V \to W$ such that the diagram

$$\begin{tikzcd}
\operatorname{Ind}_H^G V \arrow[r, "f"] & W \\
V \arrow[u, "i"] \arrow[ur, "j"'] &
\end{tikzcd}$$

commutes. It is clear that if $X \subseteq G$ contains exactly one element from each of the left cosets gH of H in G, the we have a linear isomorphism

$$\mathrm{Ind}_H^G V \cong \mathbb{F}[X] \otimes_{\mathbb{F}} V.$$

Here we denote by $\mathbb{F}[X]$ the linear span of X in $\mathbb{F}[G]$ (even if X is not a subgroup). We shall construct the analogue for a Lie algebra of the group algebra of a group —the universal enveloping algebra—. First, we construct the *tensor algebra* $T(V)$ of a vector space V. For $n \geq 0$ define $T^n(V)$ to be the n-th tensor power of V, *i. e.*, the vector space

$$T^n(V) = \underbrace{V \otimes \ldots \otimes V}_{n \text{ times}}.$$

Here it is understood that $T^0(V) = \mathbb{F}$ and $T^1(V) = V$.

Definition 2.7.2. For a vector space V, we define the *tensor algebra* $T(V)$ by

$$T(V) = \bigoplus_{n \geq 0} T^n(V), \tag{2.10}$$

with the associative algebra structure given by the condition

$$(v_1 \otimes \ldots \otimes v_m)(w_1 \otimes \ldots \otimes w_n) = v_1 \otimes \ldots \otimes v_m \otimes w_1 \otimes \ldots \otimes w_n \in T^{m+n}(V).$$

Then $T(V)$ becomes a $\mathbb{Z}$-graded associative algebra with $T(V)_n = T^n(V)$, for $n \geq 0$, and $T(V)_n = 0$, for $n < 0$. Such algebra is characterized by the following universal property: given any associative algebra A and linear map $j : V \to A$, there exists a unique algebra map $f : T(V) \to A$ for which the diagram

$$\begin{tikzcd}
T(V) \arrow[r, "f"] & A \\
V \arrow[u, "i"] \arrow[ru, "j"'] &
\end{tikzcd}$$

commutes, where i is the inclusion map of V into $T(V)$. In the same sense that the tensor algebra is the 'universal associative algebra over V', the *symmetric algebra* $S(V)$ is the universal commutative associative algebra over V. To construct it, let I be the ideal of $T(V)$ generated by all the elements $v \otimes w - w \otimes v$, for $v, w \in V$, so that I is the linear span of the products $a(v \otimes w - w \otimes v)b$, for $a, b \in T(V)$, $v, w \in V$. Form the algebra $S(V) = T(V)/I$. Since I is spanned by homogeneous elements, then it is clear that $S(V)$ is a $\mathbb{Z}$-graded commutative algebra of the form

$$S(V) = \bigoplus_{n \geq 0} S^n(V), \tag{2.11}$$

where $S^n(V) = S(V)_n$, called the n-th *symmetric power* of V, is the image in $S(V)$ of $T^n(V)$. We have $S^0(V) = \mathbb{F}$ and $S^1(V) = V$. The algebra $S(V)$ is characterized by a universal property analogous to the one above, but for linear maps of V into commutative associative algebras. Given a basis $\{v_j\}_{j\in J}$ (J a totally ordered index set) of V, $S(V)$ has a basis consisting of the products $V_{j_1} \cdots V_{j_n}$, for $n \geq 0$, $j_\ell \in J$, $j_1 \leq \ldots \leq j_n$. The space $S^n(V)$ has an obvious basis. If V is G-graded (G an abelian group) then $T(V)$ and $S(V)$ acquire unique algebra G-gradings (different from (2.10) and (2.11)) extending the grading of V. We now turn to universal enveloping algebras.

Definition 2.7.3. Given a Lie algebra $\mathfrak{g}$, the *universal enveloping algebra* $U(\mathfrak{g})$ is constructed as the quotient associative algebra of $T(\mathfrak{g})$ by the ideal generated by the elements $x \otimes y - y \otimes x - [x, y]$, for $x, y \in \mathfrak{g}$. That is

$$U(\mathfrak{g}) = T(\mathfrak{g})/I.$$

Clearly, $\mathbb{F}$ embeds in $U(\mathfrak{g})$. There is a canonical linear map $i : \mathfrak{g} \to U(\mathfrak{g})$ which is an homomorphism of Lie algebras, and $U(\mathfrak{g})$ is characterized by the following universal property: Given any associative algebra A and Lie algebra map $j : \mathfrak{g} \to A$, there is a unique associative algebra map $f : U(\mathfrak{g}) \to A$ making the diagram

$$\begin{tikzcd}
U(\mathfrak{g}) \arrow[r, "f"] & A \\
\mathfrak{g} \arrow[u, "i"] \arrow[ur, "j"'] &
\end{tikzcd}$$

commute. In particular, every $\mathfrak{g}$-module is a $U(\mathfrak{g})$-module in a natural way and conversely.

If the Lie algebra $\mathfrak{g}$ is G-graded (G an abelian group), then $U(\mathfrak{g})$ becomes a G-graded algebra in a canonical way via the G-grading of $T(\mathfrak{g})$. If the Lie algebra $\mathfrak{g}$ is abelian, then $U(\mathfrak{g})$ is just the symmetric algebra $S(\mathfrak{g})$, and in particular, the map $i : \mathfrak{g} \to U(\mathfrak{g})$ is an inclusion and we know a basis of $U(\mathfrak{g})$. For a general Lie algebra $\mathfrak{g}$, the corresponding result is not trivial, and we need the following result [116, p.168]:

Theorem 2.7.4 (Poincaré-Birkhoff-Witt). *The canonical map $i : \mathfrak{g} \to U(\mathfrak{g})$ is injective. Furthermore, let $\{x_j\}_{j\in J}$ (J a totally ordered index set) be a basis of $\mathfrak{g}$. Then the universal enveloping algebra $U(\mathfrak{g})$ has a basis consisting of the ordered products $x_{j_1} \cdots x_{j_n}$, for $n \geq 0$, $j_\ell \in J$, $j_1 \leq \ldots \leq j_n$.*

Now we turn to induced Lie algebra modules.

Definition 2.7.5. Given a subalgebra $\mathfrak{h}$ of a Lie algebra $\mathfrak{g}$, and a $\mathfrak{h}$-module V, the $\mathfrak{g}$-module *induced* by V is by definition the $\mathfrak{g}$-module corresponding to the $U(\mathfrak{g})$-module

$$\operatorname{Ind}_{\mathfrak{h}}^{\mathfrak{g}} V = U(\mathfrak{g}) \otimes_{U(\mathfrak{h})} V.$$

There is a canonical $\mathfrak{h}$-module map $i : V \to \operatorname{Ind}_{\mathfrak{h}}^{\mathfrak{g}} V$, given by $v \mapsto 1 \otimes v$, and the induced module is characterized by the following universal property: For any $\mathfrak{g}$-module W and $\mathfrak{h}$-module map $j : V \to W$, there is a unique $\mathfrak{g}$-module map $f : \operatorname{Ind}_{\mathfrak{h}}^{\mathfrak{g}} V \to W$ making the diagram

$$\begin{tikzcd}
\operatorname{Ind}_{\mathfrak{h}}^{\mathfrak{g}} V \arrow[r, "f"] & W \\
V \arrow[u, "i"] \arrow[ur, "j"'] &
\end{tikzcd}$$

commute. If $\mathfrak{g}, \mathfrak{h}$ and V are G-graded (G an abelian group), then so is $\operatorname{Ind}_{\mathfrak{h}}^{\mathfrak{g}} V$, in a canonical way.

Suppose that $\mathfrak{k}$ and $\mathfrak{h}$ are subalgebras of $\mathfrak{g}$, such that $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{h}$ as vector spaces. Then the Poincaré-Birkhoff-Witt theorem (Theorem 2.7.4) implies that the linear map

$$U(\mathfrak{k}) \otimes_{\mathbb{F}} U(\mathfrak{h}) \to U(\mathfrak{g})$$

defined by $x \otimes y \mapsto xy$ is a linear isomorphism (using a basis of $\mathfrak{g}$ made up by bases of $\mathfrak{k}$ and $\mathfrak{h}$). It follows that the linear map

$$U(\mathfrak{k}) \otimes_{\mathbb{F}} V \to U(\mathfrak{g}) \otimes_{U(\mathfrak{h})} V$$

defined by $x \otimes v \to x \otimes v$ is a linear isomorphism. The action of $\mathfrak{k}$ on $\operatorname{Ind}_{\mathfrak{h}}^{\mathfrak{g}} V$ carries over to the left multiplication action of $\mathfrak{k}$ on $U(\mathfrak{k}) \otimes V$.

We mention an important special construction. Suppose that V is a finite-dimensional vector space, with a non-singular symmetric bilinear form $\langle \cdot, \cdot \rangle$. Let $\{v_1, \ldots, v_n\}$ be a basis of V and let $\{v_1', \ldots, v_n'\}$ be the corresponding dual basis of V^*, defined by

$$\langle v_i, v_j' \rangle = \delta_{ij}, \quad \text{for} \quad i, j = 1, \ldots, n.$$

Thus, the element

$$\omega_0 = \sum_{j=1}^{n} v_j' \otimes v_j \in V \otimes V,$$

is independent of the choice of basis. In fact, consider the linear isomorphism $i : V^* \to V$ from the dual V^* to V determined by $\langle \cdot, \cdot \rangle$, and the canonical linear isomorphism $j : \operatorname{End} V \to V^* \otimes V$ given by

$$\sum_{i=1}^{n} a_i \langle v_i, v \rangle \mapsto \sum_{i=1}^{n} a_i (v_i' \otimes v).$$

Then, $\omega_0 = ((i \otimes 1) \circ j)(1_V)$, where 1_V is denoting the identity in $\operatorname{End} V$. The canonical image

$$\omega_1 = \sum_{j=1}^{n} v_j' v_j \in S^2(V)$$

of ω_0 in the symmetric square of V is also independent of the basis. In particular, if V admits an orthonormal basis $\{e_1, \dots, e_n\}$ (for instance, if $\mathbb{F}$ is algebraically closed) then $\omega_0 = \sum_i e_i \otimes e_i$ and $\omega_1 = \sum_i e_i^2$.

2.8 Affine Lie algebras

Let $\mathfrak{g}$ be a Lie algebra and $\langle \cdot, \cdot \rangle$ a bilinear form on $\mathfrak{g}$ —a bilinear map from $\mathfrak{g} \times \mathfrak{g}$ to $\mathbb{F}$—. Then $\langle \cdot, \cdot \rangle$ is said to be *invariant* or $\mathfrak{g}$-*invariant* if
(i) (associativity) $\langle [x,y], z \rangle = \langle x, [y,z] \rangle$, for $x, y, z \in \mathfrak{g}$.

Suppose that $\langle \cdot, \cdot \rangle$ is an invariant symmetric bilinear form on $\mathfrak{g}$. To the pair $(\mathfrak{g}, \langle \cdot, \cdot \rangle)$ we shall associate two infinite-dimensional graded Lie algebras $\hat{\mathfrak{g}}$ and $\tilde{\mathfrak{g}}$, called the 'affine Lie algebras'.

Let $\mathbb{F}[t, t^{-1}]$ be the commutative associative algebra of Laurent polynomials in an indeterminate t. For a Laurent polynomial

$$f = \sum_{n \in \mathbb{Z}} a_n t^n, \quad a_n \in \mathbb{F}$$

the sum being finite, set $f_0 = a_0$. Let d be the derivation

$$d = t\partial_t \tag{2.12}$$

on $\mathbb{F}[t, t^{-1}]$. Note that $(df)_0 = 0$. Consider the vector space

$$\hat{\mathfrak{g}} = \mathfrak{g} \otimes_{\mathbb{F}} \mathbb{F}[t, t^{-1}] \oplus \mathbb{F}c,$$

where $\mathbb{F}c$ is a one-dimensional space. There is an (alternating) bilinear map $[\cdot, \cdot] : \hat{\mathfrak{g}} \times \hat{\mathfrak{g}} \to \hat{\mathfrak{g}}$ determined by the conditions

$$[c, \hat{\mathfrak{g}}] = [\hat{\mathfrak{g}}, c] = 0;$$

$$[x \otimes f, y \otimes g] = [x,y] \otimes fg + \langle x, y \rangle (df \cdot g)_0 \, c; \tag{2.13}$$

for $x, y \in \hat{\mathfrak{g}}$, $f, g \in \mathbb{F}[t, t^{-1}]$, or equivalently,

$$[c, \hat{\mathfrak{g}}] = [\hat{\mathfrak{g}}, c] = 0;$$

$$[x \otimes t^m, y \otimes t^n] = [x,y] \otimes t^{m+n} + \langle x, y \rangle m \, \delta_{m+n,0} \, c; \tag{2.14}$$

for all $x, y \in \hat{\mathfrak{g}}$, $m, n \in \mathbb{Z}$. It follows directly from the symmetry and invariance of $\langle \cdot, \cdot \rangle$ that $\hat{\mathfrak{g}}$ is a Lie algebra.

Definition 2.8.1. We will call this Lie algebra $\hat{\mathfrak{g}}$ the *affine algebra* or the *untwisted affine algebra* associated with $\mathfrak{g}$ and $\langle \cdot, \cdot \rangle$.

Give the space $\mathfrak{g} \otimes \mathbb{F}[t, t^{-1}]$ the Lie algebra structure by:

$$[x \otimes t^m, y \otimes t^n] = [x, y] \otimes t^{m+n}, \quad \text{for } x, y \in \mathfrak{g},\ m, n \in \mathbb{Z}.$$

Then, there is an exact sequence of Lie algebras via the canonical maps

$$0 \longrightarrow \mathbb{F}c \longrightarrow \hat{\mathfrak{g}} \longrightarrow \mathfrak{g} \otimes \mathbb{F}[t, t^{-1}] \longrightarrow 0,$$

so that $\hat{\mathfrak{g}}$ is a central extension of $\mathfrak{g} \otimes \mathbb{F}[t, t^{-1}]$.
For $x \in \mathfrak{g}$, we shall sometimes write x for the element $x \otimes t^0$ of $\mathfrak{g} \otimes \mathbb{F}[t, t^{-1}]$.

Suppose that $\mathfrak{g}$ is not assumed to be a Lie algebra, but only a nonassociative algebra under $[\cdot, \cdot]$. Suppose also that the form $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$ is not assumed symmetric or invariant, but only bilinear. We can repeat the construction of the vector space $\hat{\mathfrak{g}}$ and of the nonassociative algebra structure $[\cdot, \cdot]$ given by (2.14). In this case, $\hat{\mathfrak{g}}$ is a Lie algebra if and only if $\mathfrak{g}$ is a Lie algebra and the form $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$ is symmetric and $\mathfrak{g}$-invariant.

Led d also denote the derivation of $\hat{\mathfrak{g}}$ determined by

$$\begin{aligned} d(c) &= 0, \\ d(x \otimes f) &= x \otimes df, \end{aligned} \tag{2.15}$$

for $x \in \mathfrak{g}$, $f \in \mathbb{F}[t, t^{-1}]$. Form the semidirect product Lie algebra

$$\tilde{\mathfrak{g}} = \hat{\mathfrak{g}} \rtimes \mathbb{F}d,$$

called the *extended affine algebra* associated with $\mathfrak{g}$ and $\langle \cdot, \cdot \rangle$, or just the *affine algebra*, if no confusion is possible. We obtain a natural gradation

$$\tilde{\mathfrak{g}} = \bigoplus_{n \in \mathbb{Z}} \tilde{\mathfrak{g}}_n$$

by considering the eigenspaces

$$\tilde{\mathfrak{g}}_n = \{x \in \tilde{\mathfrak{g}} : [d, x] = nx\}, \quad n \in \mathbb{Z},$$

of $\operatorname{ad} d$. Then, d is the degree derivation with respect to this grading, and

$$\tilde{\mathfrak{g}} = \begin{cases} \mathfrak{g} \otimes t^n & \text{for } n \neq 0 \\ \mathfrak{g} \oplus \mathbb{F}c \oplus \mathbb{F}d & \text{for } n = 0 \end{cases}$$

where we write $\mathfrak{g} \otimes t^0$ as $\mathfrak{g}$. We also have a gradation of $\hat{\mathfrak{g}}$,

$$\hat{\mathfrak{g}} = \bigoplus_{n \in \mathbb{Z}} \hat{\mathfrak{g}}_n$$

via $\hat{\mathfrak{g}}_n = \tilde{\mathfrak{g}}_n \cap \hat{\mathfrak{g}}$.

When $\mathfrak{h}$ is a subalgebra of $\mathfrak{g}$, we shall consider $\hat{\mathfrak{h}}$ and $\tilde{\mathfrak{h}}$ as subalgebras of $\hat{\mathfrak{g}}$ and $\tilde{\mathfrak{g}}$ in the obvious way. We shall also consider the analogue of affinization by 'twisting' by an involution of $\mathfrak{g}$.

Definition 2.8.2. An automorphism θ of a Lie algebra (or another algebraic structure) is called an *involution* if

$$\theta^2 = 1.$$

Let θ be an involution of a Lie algebra $\mathfrak{g}$ which is also an isometry with respect to the form $\langle \cdot, \cdot \rangle$, *i. e.*,

$$\langle \theta x, \theta y \rangle = \langle x, y \rangle \quad \text{for } x, y \in \mathfrak{g}.$$

For $i \in \mathbb{Z}_2$, set

$$\mathfrak{g}_{(i)} = \{x \in \mathfrak{g} : \theta x = (-1)^i x\}.$$

Then,

- $\mathfrak{g} = \mathfrak{g}_{(0)} \oplus \mathfrak{g}_{(1)}$;
- $[\mathfrak{g}_{(0)}, \mathfrak{g}_{(0)}] \subseteq \mathfrak{g}_{(0)}, \ [\mathfrak{g}_{(0)}, \mathfrak{g}_{(1)}] \subseteq \mathfrak{g}_{(1)}, \ [\mathfrak{g}_{(1)}, \mathfrak{g}_{(1)}] \subseteq \mathfrak{g}_{(0)}$;
- $\langle \mathfrak{g}_{(0)}, \mathfrak{g}_{(1)} \rangle = 0$.

Consider the algebra $\mathbb{F}[t^{1/2}, t^{-1/2}]$ of Laurent polynomials in an indeterminate $t^{1/2}$ (whose square is t), and extend d to a derivation of $\mathbb{F}[t^{1/2}, t^{-1/2}]$ via

$$d : t^{n/2} \mapsto \tfrac{n}{2}\, t^{n/2}, \quad n \in \mathbb{Z}, \tag{2.16}$$

and form

$$\mathfrak{l} = \mathfrak{g} \otimes_{\mathbb{F}} \mathbb{F}[t^{1/2}, t^{-1/2}] \oplus \mathbb{F}c.$$

The formulas (2.13) and (2.14) again make $\mathfrak{l}$ into a Lie algebra. Let ν the automorphism of $\mathbb{F}[t^{1/2}, t^{-1/2}]$ such that

$$\nu : t^{1/2} \mapsto -t^{1/2},$$

and let θ be the automorphism of $\mathfrak{l}$ determined by

$$\theta : \begin{cases} c \mapsto c \\ x \otimes f \mapsto \theta x \otimes \nu f \end{cases},$$

for $x \in \mathfrak{g}$, $f \in \mathbb{F}[t^{1/2}, t^{-1/2}]$. The formula $\theta(x \otimes f) = \theta x \otimes f$ will define another automorphism of $\mathfrak{l}$.

Definition 2.8.3. The *twisted affine algebra* $\hat{\mathfrak{g}}[\theta]$ is then the subalgebra

$$\hat{\mathfrak{g}}[\theta] = \{x \in \mathfrak{l} : \theta x = x\}$$

of fixed points of θ in $\mathfrak{l}$.

We have

$$\hat{\mathfrak{g}}[\theta] = \mathfrak{g}_{(0)} \otimes \mathbb{F}[t, t^{-1}] \oplus \mathfrak{g}_{(1)} \otimes t^{1/2}\mathbb{F}[t, t^{-1}] \oplus \mathbb{F}c.$$

We can again adjoin the derivation d determined by (2.16) as in (2.15) and set

$$\tilde{\mathfrak{g}}[\theta] = \hat{\mathfrak{g}}[\theta] \rtimes \mathbb{F}d,$$

the *extended twisted affine algebra* associated with $\mathfrak{g}$, $\langle \cdot, \cdot \rangle$ and θ. The eigenspaces of $\operatorname{ad} d$ make $\hat{\mathfrak{g}}[\theta]$ and $\tilde{\mathfrak{g}}[\theta]$ into $\frac{1}{2}\mathbb{Z}$-graded Lie algebras. Note that if $\theta = 1$, then $\tilde{\mathfrak{g}}[\theta]$ degenerates to the untwisted affine algebra $\tilde{\mathfrak{g}}$.

Finally, we remark that the process of twisted affinization can be extended to any automorphism of finite order of $\mathfrak{g}$, which is an isometry with respect to $\langle \cdot, \cdot \rangle$.

Chapter 3

The Root space decomposition

Now we will consider a class of Lie algebras (the complex semisimple ones), that their representations can be described, similarly to $\mathfrak{sl}_3(\mathbb{C})$, by a 'theorem of the highest weight'. We develop the structures needed to state the theorem of the highest weight. Although this chapter could be understood simply as a description of the structure of semisimple Lie algebras, without any mention of representation theory, it is helpful to have the representations in mind. The representation theory, especially in light of $\mathfrak{sl}_n(\mathbb{C})$, motivates the notions of Cartan subalgebras, roots, and the Weyl group. See [179] or [91] a more detailed treatment of these topics.

3.1 Representations

Definition 3.1.1. Let G be a matrix Lie group. Then, a (finite-dimensional) *complex representation* of G is a Lie group homomorphism $\Pi : G \to GL_n(\mathbb{C})$ (with $n \geq 1$) or, in other words, a Lie group homomorphism $\Pi : G \to GL(V)$, where V is a finite-dimensional complex vector space (with $\dim V \geq 1$). A finite-dimensional *real representation* of G is a Lie group homomorphism Π of G into $GL_n(\mathbb{R})$ or into $GL(V)$, where V is a finite-dimensional real vector space.

Definition 3.1.2. If $\mathfrak{g}$ is a real or complex Lie algebra, then a finite-dimensional *complex representation* of $\mathfrak{g}$ is a Lie algebra homomorphism $\pi : \mathfrak{g} \to \mathfrak{gl}_n(\mathbb{C})$ (or into $\mathfrak{gl}(V)$), where V is a finite-dimensional complex vector space. We can define a real representation of $\mathfrak{g}$ in a similar way. If Π or π is a one-to-one homomorphism, then the representation is called *faithful*.

One should think of a representation as a linear action of a group or Lie algebra on a vector space (since to every $g \in G$, there is associated an operator $\Pi(g)$, which acts on the vector space V). If $\mathfrak{g}$ is a real Lie algebra, we will consider mainly complex representations of $\mathfrak{g}$.

Definition 3.1.3. Let Π be a finite-dimensional real or complex representation of a matrix Lie group G, acting on a space V. A subspace W of V is called *invariant* if $\Pi(A)w \in W$, for all $w \in W$ and all $A \in G$. An invariant subspace W is called *nontrivial* if $W \neq \{0\}$ and $W \neq V$. A representation with no nontrivial invariant subspaces is called *irreducible*. The terms invariant, nontrivial, and irreducible are defined analogously for representations of Lie algebras.

Definition 3.1.4. Let G be a matrix Lie group, let Π be a representation of G acting on the space V, and let Σ be a representation of G acting on the space W. A linear map $\varphi : V \to W$ is called an *intertwining* map of representations if

$$\varphi\big(\Pi(A)v\big) = \Sigma(A)\varphi(v),$$

for all $A \in G$ and all $v \in V$. The analogous property defines intertwining maps of representations of a Lie algebra. If φ is an intertwining map of representations and, in addition, φ is invertible, then φ is said to be an *equivalence* of representations. If there exists an isomorphism between V and W, then the representations are said to be *equivalent*.

If G is a matrix Lie group with Lie algebra $\mathfrak{g}$ and Π is a (finite-dimensional real or complex) representation of G, acting on the space V, then there is a unique representation π of $\mathfrak{g}$ acting on the same space, such that

$$\Pi(e^x) = e^{\pi(x)}, \quad \text{for all } x \in \mathfrak{g}.$$

This representation π can be computed as

$$\pi(x) = \tfrac{d}{dt}\Pi(e^{tx})\big|_{t=0},$$

and satisfies $\pi(AxA^{-1}) = \Pi(A)\Pi(x)\Pi(A)^{-1}$, for all $x \in \mathfrak{g}$ and all $A \in G$. We state the following two propositions. For a proof of them, see [91, p.93].

Proposition 3.1.5. 1. *Let G be a connected matrix Lie group with Lie algebra $\mathfrak{g}$. Let Π be a representation of G and π the associated representation of $\mathfrak{g}$. Then, Π is irreducible if and only if π is irreducible.*

2. *Let G be a connected matrix Lie group, let Π_1 and Π_2 be representations of G, and let π_1 and π_2 be the associated Lie algebra representations. Then, π_1 and π_2 are equivalent if and only Π_1 and Π_2 are equivalent.*

Proposition 3.1.6. *Let $\mathfrak{g}$ be a real Lie algebra and $\mathfrak{g}_{\mathbb{C}}$ its complexification. Then, every finite-dimensional complex representation π of $\mathfrak{g}$ has a unique extension to a complex-linear representation of $\mathfrak{g}_{\mathbb{C}}$, also denoted π and given by*

$$\pi(x + iy) = \pi(x) + i\pi(y), \quad \textit{for all } x, y \in \mathfrak{g}.$$

Furthermore, π is irreducible as a representation of $\mathfrak{g}_{\mathbb{C}}$ if and only if it π irreducible as a representation of $\mathfrak{g}$.

We give some examples of representations:

Example 3.1.7. (The standard representation)
A matrix Lie group G is, by definition, a subset of some $GL_n(\mathbb{C})$. The inclusion map of G into $GL_n(\mathbb{C})$ (*i. e.*, $\Pi(A) = A$) is a representation of G, called the *standard representation*

of G. If G happens to be contained in $GL_n(\mathbb{R})$ or $GL_n(\mathbb{C})$, then we can think of the standard representation as a real representation if we prefer. Thus, for example, the standard representation of $SO_3(\mathbb{C})$ is the one in which $SO_3(\mathbb{C})$ acts in the usual way on $\mathbb{R}^3$ and the standard representation of $SU(2)$ is the one in which $SU(2)$ acts on $\mathbb{C}^2$ in the usual way. If G is a subgroup of $GL_n(\mathbb{R})$ or $GL_n(\mathbb{C})$, then its Lie algebra $\mathfrak{g}$ will be a subalgebra of $\mathfrak{gl}_n(\mathbb{R})$ or $\mathfrak{gl}_n(\mathbb{C})$. The inclusion of $\mathfrak{g}$ into $\mathfrak{gl}_n(\mathbb{R})$ or $\mathfrak{gl}_n(\mathbb{C})$ is a representation of $\mathfrak{g}$, called the *standard representation.*

Example 3.1.8. (The trivial representation)
Consider the one-dimensional complex vector space $\mathbb{C}$. Given any matrix Lie group G, we can define the trivial representation of G, $\Pi : G \to GL_1(\mathbb{C})$, by the formula

$$\Pi(A) = I, \quad \text{for all } A \in G.$$

Of course, this is an irreducible representation, since $\mathbb{C}$ has no nontrivial subspaces, and thus no nontrivial invariant subspaces. If $\mathfrak{g}$ is a Lie algebra, we can also define the trivial representation of $\mathfrak{g}$, $\pi : \mathfrak{g} \to \mathfrak{gl}_1(\mathbb{C})$, by

$$\pi(x) = 0, \quad \text{for all } x \in \mathfrak{g}.$$

This is an irreducible representation.

Example 3.1.9 (The adjoint representation)**.** Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. We define the *adjoint mapping* $\mathrm{Ad} : G \to GL(\mathfrak{g})$ by the formula

$$\mathrm{Ad}_A(X) = AXA^{-1}.$$

Since Ad is a Lie group homomorphism into a group of invertible operators, we see that, in fact, Ad is a representation of G, acting on the space $\mathfrak{g}$. We call Ad the *adjoint representation* of G. The adjoint representation is a real representation of G (if $\mathfrak{g}$ is a complex subspace of $\mathbb{C}^{n\times n}$, then we can think of the adjoint representation as a complex representation). Similarly, if $\mathfrak{g}$ is a Lie algebra, we have the adjoint map $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$ defined by the formula (see Chapter 2)

$$\mathrm{ad}_x(y) = [x, y].$$

We know that ad is a Lie algebra homomorphism and is, therefore, a representation of $\mathfrak{g}$, called the *adjoint representation.* In the case that $\mathfrak{g}$ is the Lie algebra of some matrix Lie group G, Ad and ad are related by $e^{\mathrm{ad}_x} = \mathrm{Ad}_{e^x}$.

Now we will discuss a classical example, namely the irreducible complex representations of the Lie algebra $\mathfrak{su}(2)$. This computation is important for several reasons. In the first place, $\mathfrak{su}(2) \cong \mathfrak{so}(3)$ and the representations of $\mathfrak{so}(3)$ are of physical significance, particularly in quantum mechanics [91]. In the second place, the representation theory of $\mathfrak{su}(2)$ is an illuminating example of how one uses relations to determine the representations of a Lie algebra. Also, in determining the representations of semisimple Lie algebras (Section 3.2), it usually uses the representation of $\mathfrak{su}(2)$.

Example 3.1.10. (The Irreducible Representations of $\mathfrak{sl}_2(\mathbb{C})$)
Every finite-dimensional complex representation π of $\mathfrak{su}(2)$ extends to a complex-linear representation (also called π) of the complexification of $\mathfrak{su}(2)$, namely $\mathfrak{sl}_2(\mathbb{C})$. The extension of π to $\mathfrak{sl}_2(\mathbb{C})$ is irreducible if and only if the original representation is irreducible. We will use the following basis for $\mathfrak{sl}_2(\mathbb{C})$:

$$h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix};$$

which have the commutation relations

$$\begin{aligned} [h, x] &= 2x, \\ [h, y] &= -2y, \\ [x, y] &= h. \end{aligned}$$

If V is a (finite-dimensional complex) vector space and A, B and C are operators on V satisfying

$$\begin{aligned} [A, B] &= 2B, \\ [A, C] &= -2C, \\ [B, C] &= A, \end{aligned}$$

then because of the skew symmetry and bilinearity of brackets, the linear map $\pi : \mathfrak{sl}_2(\mathbb{C}) \to \mathfrak{gl}(V)$ satisfying

$$\pi(h) = A, \quad \pi(x) = B, \quad \pi(y) = C,$$

will be a representation of $\mathfrak{sl}_2(\mathbb{C})$.

We state an important result we will use later. For a proof see [179, p.268].

Theorem 3.1.11. *Suppose π is any finite-dimensional, complex-linear representation of $\mathfrak{sl}_2(\mathbb{C})$ acting on a space V. Then, we have the following results:*

1. *Every eigenvalue of $\pi(h)$ must be an integer.*
2. *If v is a nonzero element of V such that $\pi(x)v = 0$ and $\pi(h)v = \lambda v$, then λ is a non-negative integer. Furthermore, the vectors $v, \pi(y)v, \ldots, \pi(y)^{\lambda}v$ are linearly independent and their span is an irreducible invariant subspace of dimension $\lambda + 1$.*

Analogously to the previous example, we can think of study the representations of $\mathfrak{su}(3)$ by studying the representations of its complexification $\mathfrak{sl}_3(\mathbb{C})$. We will discuss this particular case in Section 3.9. In general, studying the irreducible representations of $\mathfrak{su}(n)$ is equivalent to studying the irreducible (complex-linear) representations of $\mathfrak{sl}_n(\mathbb{C})$. Passing to the complexified Lie algebra makes our computations easier, and we can find a nice basis for $\mathfrak{sl}_n(\mathbb{C})$ that has no counterpart among the bases of $\mathfrak{sl}(n)$.

3.2 Complete Reducibility and Semisimple algebras

Definition 3.2.1. A finite-dimensional representation of a group or Lie algebra is said to be *completely reducible* if it is isomorphic to a direct sum of a finite number of irreducible representations. A group or Lie algebra is said to have the *complete reducibility property* if every finite-dimensional representation of it is completely reducible.

The complete reducibility property is a very special one that most groups and Lie algebras do not have. If a group or Lie algebra does have the complete reducibility property, then the study of its representations reduces to the study of its irreducible representations, which simplifies the analysis considerably. The following results are useful [91, p.119]

Proposition 3.2.2. *If V is a completely reducible representation of a group or Lie algebra, then the following properties hold:*

1. *Every invariant subspace of V is completely reducible.*
2. *Given any invariant subspace U of V, there is another invariant subspace $\tilde{U}$ such that $V = U \oplus \tilde{U}$.*

Proposition 3.2.3. *Let G be a matrix Lie group. Let Π be a finite-dimensional unitary representation of G, acting on a finite-dimensional real or complex Hilbert space V. Then, Π is completely reducible.*

If Π is a representation of a finite group G, acting on a space V, we can choose an arbitrary inner product $\langle \cdot, \cdot \rangle$ on V. Then, we can define a new inner product $\langle \cdot, \cdot \rangle_G$ on V by

$$\langle v_1, v_2 \rangle_G = \sum_{g \in G} \langle \Pi(g)v_1, \Pi(g)v_2 \rangle. \tag{3.1}$$

Furthermore, if $h \in G$, then

$$\begin{aligned} \langle \Pi(h)v_1, \Pi(h)v_2 \rangle_G &= \sum_{g \in G} \langle \Pi(g)\Pi(h)v_1, \Pi(g)\Pi(h)v_2 \rangle \\ &= \sum_{g \in G} \langle \Pi(gh)v_1, \Pi(gh)v_2 \rangle \end{aligned}$$

However, as g ranges over G, so does gh. Thus, in fact, $\langle \Pi(h)v_1, \Pi(h)v_2 \rangle_G = \langle v_1, v_2 \rangle_G$; that is, Π is a unitary representation with respect to the inner product $\langle \cdot, \cdot \rangle$. Thus, Π is isomorphic to a direct sum of irreducibles, by Proposition 3.2.3 and we have

Proposition 3.2.4. *Every finite group has the complete reducibility property.*

There is a variant of the above argument which can be used to prove the following result [116]:

Proposition 3.2.5. *If G is a compact matrix Lie group, G has the complete reducibility property.*

The argument below is sometimes called 'Weyl's Unitarian trick'. Its proof requires the notion of Haar measure (see, for example, [116]). A *left Haar* measure on a matrix Lie group G is a nonzero measure μ on the Borel σ-algebra in G with the following two properties:

- It is locally finite (*i. e.*, every point in G has a neighborhood with finite measure),
- It is left-translation invariant: $\mu(gE) = \mu(E)$, for all $g \in G$ and all Borel sets $E \subseteq G$.

It is a fact, which we cannot prove here, that every matrix Lie group has a left Haar measure and that this measure is unique up to multiplication by a constant. One can analogously define right Haar measure, and a similar theorem holds for it. Left Haar measure and right Haar measure may or may not coincide (a group for which they do is called *unimodular*). Now, the key fact for our purpose is that left Haar measure is finite if and only if the group G is compact. Suppose, then, that Π is a finite-dimensional representation of a compact group G acting on a space V. Let $\langle \cdot, \cdot \rangle$ be an arbitrary inner product on V and define a new inner product $\langle \cdot, \cdot \rangle_G$ on V (analogous to (3.1)) by

$$\langle v_1, v_2 \rangle_G = \int_G \langle \Pi(g)v_1, \Pi(g)v_2 \rangle \, d\mu(g), \tag{3.2}$$

where μ is a left Haar measure. Again, it is possible to check that $\langle \cdot, \cdot \rangle_G$ is an inner product. Furthermore, if $h \in G$, then by the left-invariance of μ,

$$\begin{aligned} \langle \Pi(h)v_1, \Pi(h)v_2 \rangle_G &= \int_G \langle \Pi(g)\Pi(h)v_1, \Pi(g)\Pi(h)v_2 \rangle \, d\mu(g) \\ &= \int_G \langle \Pi(gh)v_1, \Pi(gh)v_2 \rangle \, d\mu(g) = \langle v_1, v_2 \rangle_G. \end{aligned}$$

So, Π is a unitary representation with respect to $\langle \cdot, \cdot \rangle_G$ and thus completely reducible (note that the integral in (3.2) is convergent because μ is finite).

Recall from Chapter 2 that if $\mathfrak{g}$ is a complex Lie algebra, then an ideal in $\mathfrak{g}$ is a complex subalgebra $\mathfrak{h}$ of $\mathfrak{g}$ with the property that for all $x \in \mathfrak{g}$ and all $h \in \mathfrak{h}$, we have $[x, h] \in \mathfrak{h}$. Recall also that a complex Lie algebra $\mathfrak{g}$ is called indecomposable if the only ideals in $\mathfrak{g}$ are $\mathfrak{g}$ and (0) . A complex Lie algebra $\mathfrak{g}$ is called simple if $\mathfrak{g}$ is indecomposable and $\dim \mathfrak{g} \geq 2$.

There is an analogy between finite-dimensional Lie algebras and finite groups. Subalgebras in the Lie algebra setting are the analogs of subgroups in the finite group setting, and ideals in the Lie algebra setting are the analogs of normal subgroups in the finite group setting. In this analogy, the one-dimensional Lie algebras (which are precisely the Lie algebras having no nontrivial subalgebras) are the analogs of the cyclic groups of prime order (which are precisely the groups having no nontrivial subgroups). However, there is a discrepancy in terminology: cyclic groups of prime order are called simple but one-dimensional Lie algebras are not called simple. This terminological convention is important to bear in mind in the following definition.

Definition 3.2.6. A complex Lie algebra is called *reductive* if it is isomorphic to a direct sum of indecomposable Lie algebras. A complex Lie algebra is called *semisimple* if it isomorphic to a direct sum of simple Lie algebras.

Note that a reductive Lie algebra is a direct sum of indecomposable algebras, which are either simple or one-dimensional commutative. Thus, a reductive Lie algebra is one that decomposes as a direct sum of a semisimple algebra (coming from the simple terms in the direct sum) and a commutative algebra (coming from the one-dimensional terms in the direct sum). We will assume that the complex semisimple Lie algebras we study are given to us as subalgebras of some $\mathfrak{gl}_n(\mathbb{C})$, since by Ado's Theorem 2.3.7 every finite-dimensional Lie algebra has a faithful finite-dimensional representation. In fact, for semisimple Lie algebras, the adjoint representation is always faithful, as is shown in [91, p.158]

Proposition 3.2.7. *A complex Lie algebra* $\mathfrak{g}$ *is reductive precisely if the adjoint representation is completely reducible.*

In fact, the complexification of the Lie algebra of a connected compact matrix Lie group is reductive. This follows from the above proposition and the property that connected compact groups have the complete reducibility property (Proposition 3.2.5). Note that the Lie algebra of a compact Lie group may be only reductive and not semisimple. For example, the Lie algebra of S^1 is one dimensional and, thus, not semisimple. We have the following characterization result of semisimple Lie algebras (see [179, p.348])

Theorem 3.2.8. *A complex Lie algebra is semisimple if, and only if, it is isomorphic to the complexification of the Lie algebra of a simply-connected compact matrix Lie group.*

We have already seen that if $\mathfrak{g}$ is the complexification of the Lie algebra of a compact simply-connected group K, then $\mathfrak{g}$ is reductive, even if K is not simply connected. Thus $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$, with $\mathfrak{g}_1$ semisimple and $\mathfrak{g}_2$ commutative. It can be shown that the Lie algebra $\mathfrak{k}$ of K decomposes as $\mathfrak{k} = \mathfrak{k}_1 \oplus \mathfrak{k}_2$, where $\mathfrak{g}_1 = \mathfrak{k}_1 + i\mathfrak{k}_1$ and $\mathfrak{g}_2 = \mathfrak{k}_2 + i\mathfrak{k}_2$. Then, K decomposes as $K_1 \times K_2$, where K_1 and K_2 are simply connected and where K_2 is commutative. However, a simply-connected commutative Lie group is isomorphic to $\mathbb{R}^n$, which is noncompact for $n \geq 1$. Thus, the compactness of K means that $\mathfrak{k}_2 = \{0\}$, in which case $\mathfrak{g}_2 = \{0\}$ and $\mathfrak{g} = \mathfrak{g}_1$ is semisimple.

For the other direction, given a complex semisimple Lie algebra, we must find the correct real form whose corresponding simply-connected group is compact (*c.f.* [179]).

Definition 3.2.9. If $\mathfrak{g}$ is a complex semisimple Lie algebra, then a *compact real form* of $\mathfrak{g}$ is a real subalgebra $\mathfrak{k}$ of $\mathfrak{g}$ with the property that every $x \in \mathfrak{g}$ can be written uniquely as $x = x_1 + ix_2$, with x_1 and x_2 in $\mathfrak{k}$ and such that there is a compact simply-connected matrix Lie group K' such that the Lie algebra $\mathfrak{k}'$ of K' is isomorphic to $\mathfrak{k}$.

We have the following important fact [91, p.159].

Proposition 3.2.10. *Let $\mathfrak{g}$ be a complex semisimple Lie algebra. If $\mathfrak{g}$ is a subalgebra of $\mathfrak{gl}_n(\mathbb{C})$ and $\mathfrak{k}$ is a compact real form of $\mathfrak{g}$, then the connected Lie subgroup K of $GL_n(\mathbb{C})$ whose Lie algebra is $\mathfrak{k}$ is compact.*
In particular, every complex semisimple Lie algebra has the complete reducibility property.

This last statement holds because the representations of $\mathfrak{g}$ are in one-to-one correspondence with the representations of K, and compact groups have the complete reducibility property (Proposition 3.2.5). Actually, only the semisimple ones have the complete reducibility property, and thus, complete reducibility is sometimes taken as the definition of semisimplicity for Lie algebras. For an algebraic proof of complete reducibility of semisimple Lie algebras, see [100]). Up to now, we have considered only complex semisimple Lie algebras, since these are the ones whose representations we will consider. Nevertheless, we can define the terms ideal, indecomposable, simple, reductive, and semisimple for real Lie algebras in precisely the same way as for the complex case.

Let us consider some examples of Lie algebras that are reductive or semisimple. The following table lists some of the complex Lie algebras that we have encountered already that are either reductive or semisimple (see [22]). Here, 'reductive' means actually 'reductive but not semisimple'.

Group	Reductive/Semisimple
$\mathfrak{sl}_n(\mathbb{C})$ $(n \geq 2)$	semisimple
$\mathfrak{so}_n(\mathbb{C})$ $(n \geq 3)$	semisimple
$\mathfrak{so}_2(\mathbb{C})$	reductive
$\mathfrak{gl}_n(\mathbb{C})$ $(n \geq 1)$	reductive
$\mathfrak{sp}_n(\mathbb{C})$ $(n \geq 1)$	semisimple

Table 3.1: Semisimple properties of some classical complex Lie algebras.

The other Lie algebras we have examined, such as the Lie algebras of the Heisenberg group, are neither reductive nor semisimple.

Group	Reductive/Semisimple
$\mathfrak{su}(n)$ $(n \geq 2)$	semisimple
$\mathfrak{so}(n)$ $(n \geq 3)$	semisimple
$\mathfrak{so}(2)$	reductive
$\mathfrak{sp}(n)$ $(n \geq 1)$	semisimple
$\mathfrak{sp}_n(\mathbb{R})$ $(n \geq 1)$	semisimple
$\mathfrak{sl}_n(\mathbb{R})$ $(n \geq 2)$	semisimple
$\mathfrak{gl}_n(\mathbb{R})$ $(n \geq 1)$	reductive

Table 3.2: Semisimple properties of some classical real Lie algebras.

3.3 Cartan subalgebras

Definition 3.3.1. If $\mathfrak{g}$ is a complex semisimple Lie algebra, then a *Cartan subalgebra* of $\mathfrak{g}$ is a complex subspace $\mathfrak{h}$ of $\mathfrak{g}$ with the following properties:
(i) For all h_1 and h_2 in $\mathfrak{h}$, $[h_1, h_2] = 0$.
(ii) For all $x \in \mathfrak{g}$, if $[h, x] = 0$ for all $h \in \mathfrak{h}$, then x is in $\mathfrak{h}$.
(iii) For all $h \in \mathfrak{h}$, $\operatorname{ad} h$ is diagonalizable.

Condition (i) says that $\mathfrak{h}$ is a commutative subalgebra of $\mathfrak{g}$. Condition (ii) says that $\mathfrak{h}$ is a maximal commutative subalgebra, *i. e.*, not contained in any larger commutative subalgebra. Condition (iii) says that each $\operatorname{ad} h$ $(h \in \mathfrak{h})$ is diagonalizable. Since the h's in $\mathfrak{h}$ commute, the $\operatorname{ad} h$'s also commute, and thus they are simultaneously diagonalizable. (It is a standard result in linear algebra that any commuting family of diagonalizable matrices is simultaneously diagonalizable; see [97]). Of course, the definition of a Cartan subalgebra makes sense in any Lie algebra, semisimple or not. However, if $\mathfrak{g}$ is not semisimple, then $\mathfrak{g}$ may not have any Cartan subalgebras. Even in the semisimple case we must prove that a Cartan subalgebra exists (see [91, p.163]).

Proposition 3.3.2. *Let $\mathfrak{g}$ be a complex semisimple Lie algebra, let $\mathfrak{k}$ be a compact real form of $\mathfrak{g}$, and let $\mathfrak{t}$ be any maximal commutative subalgebra of $\mathfrak{k}$. Define $\mathfrak{h} \subseteq \mathfrak{g}$ to be $\mathfrak{h} = \mathfrak{t} + i\mathfrak{t}$. Then, $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}$.*

Note that $\mathfrak{k}$ (or any other Lie algebra) contains a maximal commutative subalgebra. After all, let $\mathfrak{t}_1$ be any one-dimensional subspace of $\mathfrak{k}$. Then, $\mathfrak{t}_1$ is a commutative subalgebra of $\mathfrak{k}$. If $\mathfrak{t}_1$ is maximal, then we are done; if not, then we choose some commutative subalgebra $\mathfrak{t}_2$ properly containing $\mathfrak{t}_1$. Then, if $\mathfrak{t}_2$ is maximal, we are done, and if not, we choose a commutative subalgebra $\mathfrak{t}_3$ properly containing $\mathfrak{t}_2$. Since $\mathfrak{k}$ is finite dimensional, this process cannot go on forever and we will eventually get a maximal commutative subalgebra.

It is possible to prove that every Cartan subalgebra of $\mathfrak{g}$ arises as in Proposition 3.3.2 (for some compact real form $\mathfrak{k}$ and some maximal commutative subalgebra $\mathfrak{t}$ of $\mathfrak{k}$) and also that Cartan subalgebras are unique up to conjugation. In particular, all Cartan subalgebras of a given complex semisimple Lie algebra have the same dimension. In light of this result, the following definition makes sense.

Definition 3.3.3. If $\mathfrak{g}$ is a complex semisimple Lie algebra, then the *rank* of $\mathfrak{g}$ is the dimension of any Cartan subalgebra.

3.4 Roots and Root Spaces

From now on we assume that we have chosen a compact real form $\mathfrak{k}$ of $\mathfrak{g}$ and a maximal commutative subalgebra $\mathfrak{t}$ of $\mathfrak{k}$, and we consider the Cartan subalgebra $\mathfrak{h} = \mathfrak{t} + i\mathfrak{t}$. We assume also that we have chosen an inner product on $\mathfrak{g}$ that is invariant under the adjoint action of K and that takes real values on $\mathfrak{k}$.

Definition 3.4.1. A *root* of $\mathfrak{g}$ (relative to the Cartan subalgebra $\mathfrak{h}$) is a nonzero linear functional α on $\mathfrak{h}$ such that there exists a nonzero element x of $\mathfrak{g}$ with

$$[h, x] = \alpha(h)x,$$

for all $h \in \mathfrak{h}$.

The set of all roots is denoted by Φ. The condition on x says that x is an eigenvector for each $\operatorname{ad} h$, with eigenvalue $\alpha(h)$. Note that if x is actually an eigenvector for each $\operatorname{ad} h$ with $h \in \mathfrak{h}$, then the eigenvalues must depend linearly on h. That is why we insist that α be a linear functional on $\mathfrak{h}$. So, a root is just a (nonzero) collection of simultaneous eigenvalues for the $\operatorname{ad} h$'s. Note that any element of $\mathfrak{h}$ is a simultaneous eigenvector for all the $\operatorname{ad} h$'s, with all eigenvalues equal to zero, but we only call α a root if α is nonzero. Of course, for any root α, some of the $\alpha(h)$'s may be equal to zero; we just require that not all of them be zero. Note that the set of linear functionals on $\mathfrak{h}$ that are imaginary on $\mathfrak{t}$ forms a real vector space whose real dimension equals the complex dimension of $\mathfrak{h}$. If $\mathfrak{t}^*$ denotes the space of real-valued linear functionals on $\mathfrak{t}$, then the roots are contained in $i\mathfrak{t}^* \subseteq \mathfrak{h}^*$.

Definition 3.4.2. If α is a root of the Lie algebra $\mathfrak{g}$ (relatively to the subalgebra $\mathfrak{h}$), then the *root space* $\mathfrak{g}_\alpha$ is the space of all $x \in \mathfrak{g}$ for which $[h, x] = \alpha(h)x$, for all $h \in \mathfrak{h}$. An element of $\mathfrak{g}_\alpha$ is called a *root vector* (for the root α).

More generally, if α is any element of $\mathfrak{h}^*$, we define $\mathfrak{g}_\alpha$ to be the space of all $x \in \mathfrak{g}$ for which $[h, x] = \alpha(h)x$, for all $h \in \mathfrak{h}$ (but we do not call $\mathfrak{g}_\alpha$ a root space unless α is actually a root). Taking $\alpha = 0$, we see that $\mathfrak{g}_0$ is the set of all elements of $\mathfrak{g}$ that commute with every element of $\mathfrak{h}$. Since $\mathfrak{h}$ is a maximal commutative subalgebra, we conclude that $\mathfrak{g}_0 = \mathfrak{h}$. If α is not zero and not a root, then $\mathfrak{g}_\alpha = \{0\}$. Now, since $\mathfrak{h}$ is commutative, the operators

$\operatorname{ad} h$, $h \in \mathfrak{h}$, all commute. Furthermore, by the definition of Cartan subalgebra, each $\operatorname{ad} h$, $h \in \mathfrak{h}$, is diagonalizable. It follows that the $\operatorname{ad} h$'s, are simultaneously diagonalizable. As a result, $\mathfrak{g}$ can be decomposed as the direct sum of $\mathfrak{h}$ and the root spaces $\mathfrak{g}_\alpha$:

$$\mathfrak{g} = \mathfrak{h} \oplus \bigoplus_{\alpha \in \Phi} \mathfrak{g}_\alpha.$$

This means that every element of $\mathfrak{g}$ can be written uniquely as a sum of an element of $\mathfrak{h}$ and one element from each root space $\mathfrak{g}_\alpha$.

We resume some elementary properties of roots. You usually can find a proof of theses properties in almost books about representation theory. See for example [91] or [59] for the proofs.

Proposition 3.4.3. *(i) For any α and β in $\mathfrak{h}^*$, $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subseteq \mathfrak{g}_{\alpha+\beta}$.*
(ii) If $\alpha \in \mathfrak{h}^*$ is a root, then so is $-\alpha$.
(iii) If α is a root, then the only multiples of α that are roots are α and $-\alpha$.
(iv) The roots span $\mathfrak{h}^*$.
(v) If α is a root, then the root space $\mathfrak{g}_\alpha$ is one dimensional.
(vi) For each root α, we can find nonzero elements $x_\alpha \in \mathfrak{g}_\alpha$, $y_\alpha \in \mathfrak{g}_{-\alpha}$, and $h_\alpha \in \mathfrak{h}$ such that

$$[h_\alpha, x_\alpha] = 2x_\alpha, \quad [h_\alpha, y_\alpha] = -2y_\alpha, \quad [x_\alpha, y_\alpha] = h_\alpha.$$

The element h_α is unique, i. e., independent of the choice of x_α and y_α.

Last point of the proposition above tells us that x_α, y_α, and h_α span a subalgebra of $\mathfrak{g}$ isomorphic to $\mathfrak{sl}_2(\mathbb{C})$. The elements h_α of $\mathfrak{h}$ are called the *co-roots*. Their properties are closely related to the properties of the roots themselves.
Given any linear functional $\alpha \in \mathfrak{h}^*$ (not necessarily a root), there exists a unique element $h^\alpha \in \mathfrak{h}$ such that

$$\alpha(h) = \langle h^\alpha, h \rangle,$$

for all $h \in \mathfrak{h}$, where we take the inner product to be linear in the second factor. The map $\alpha \mapsto h^\alpha$ is a one-to-one and onto correspondence between $\mathfrak{h}^*$ and $\mathfrak{h}$. However, this correspondence is not linear but rather conjugate-linear, since the inner product is conjugate-linear in the first factor (where h^α is). It is convenient to permanently identify each root $\alpha \in \mathfrak{h}^*$ with the corresponding element $h^\alpha \in \mathfrak{h}$. Having done this, we then omit the h^α notation and denote that element of $\mathfrak{h}$ simply as α.
The reader can find some more properties and relations of roots and co-roots in [91]. We only mention that if $\alpha \in \mathfrak{h}$ is a root in the sense of last paragraph and h_α is the corresponding co-root, then α and h_α are related by the formulas

$$h_\alpha = \frac{2\alpha}{\langle \alpha, \alpha \rangle}, \quad \alpha = \frac{2h_\alpha}{\langle h_\alpha, h_\alpha \rangle}. \tag{3.3}$$

The real content of this proposition is that once we use the inner product to identify $\mathfrak{h}^*$ with $\mathfrak{h}$ (so that the roots and co-roots now live in the same space), α and h_α are multiples of one another. Once this is known, the normalization is determined by the condition that $\langle \alpha, h_\alpha \rangle = 2$, which reflects that $[h_\alpha, x_\alpha] = 2x_\alpha$. Observe that both formulas (3.3) are consistent with the relation $\langle \alpha, h_\alpha \rangle = 2$. We conclude with the following [91, p.173]

Theorem 3.4.4. *For all roots $\alpha, \beta \in \mathfrak{h}$ (in the notation above), the quantities*

$$2\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} \quad \textit{and} \quad 2\frac{\langle h_\alpha, h_\beta \rangle}{\langle h_\alpha, h_\alpha \rangle}$$

are integers and, furthermore,

$$2\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle} = 2\frac{\langle h_\alpha, h_\beta \rangle}{\langle h_\alpha, h_\alpha \rangle}.$$

3.5 The Weyl Group

We use here the compact-group approach to defining the Weyl group, as opposed to the Lie algebra approach. The compact-group approach makes certain aspects of the Weyl group more transparent. Nevertheless, the two approaches are equivalent. We continue with the setting of the previous section. Thus, $\mathfrak{g}$ is a complex semisimple Lie algebra given to us as a subalgebra of some $\mathfrak{gl}_n(\mathbb{C})$. We have chosen a compact real form $\mathfrak{k}$ of $\mathfrak{g}$ and we let K be the compact subgroup of $GL_n(\mathbb{C})$ whose Lie algebra is $\mathfrak{k}$. We have chosen a maximal commutative subalgebra $\mathfrak{t}$ of $\mathfrak{k}$, and we work with the associated Cartan subalgebra $\mathfrak{h} = \mathfrak{t} + i\mathfrak{t}$. We have chosen an inner product on $\mathfrak{g}$ that is invariant under the adjoint action of K and that takes real values on $\mathfrak{k}$. Consider the following two subgroups of K:

$$\begin{aligned} Z(\mathfrak{t}) &= \{A \in K : \operatorname{ad} A(h) = h, \ \forall h \in \mathfrak{t}\}, \\ N(\mathfrak{t}) &= \{A \in K : \operatorname{ad} A(h) \subseteq \mathfrak{t}, \ \forall h \in \mathfrak{t}\}. \end{aligned}$$

Clearly, $Z(\mathfrak{t})$ is a normal subgroup of $N(\mathfrak{t})$. If T is the connected Lie subgroup of K with Lie algebra $\mathfrak{t}$, then $T \subseteq Z(\mathfrak{t})$, since T is generated by elements of the form e^h with $h \in \mathfrak{t}$. It turns out that, in fact, $Z(\mathfrak{t}) = T$. See [22].

Definition 3.5.1. The *Weyl group* for $\mathfrak{g}$ is the quotient group $W = N(\mathfrak{t})/Z(\mathfrak{t})$.

We can define an action of W on $\mathfrak{t}$ as follows. For each element $w \in W$, choose an element A of the corresponding equivalence class in $N(\mathfrak{t})$. Then for $h \in \mathfrak{t}$ we define the action of w on h by

$$w \cdot h = \operatorname{ad} A(h).$$

In fact, this action is well defined (*i. e.*, independent of the choice of A in a given equivalence class). Since $\mathfrak{h} = \mathfrak{t} + i\mathfrak{t}$, each linear transformation of $\mathfrak{t}$ extends uniquely to a complex-linear transformation of $\mathfrak{h}$. Thus, we also think of W as acting on $\mathfrak{h}$. If w is an element of the

Weyl group, then we write $w \cdot h$ for the action of w on an element h of $\mathfrak{h}$. It can be seen that W is isomorphic to the group of linear transformations of $\mathfrak{h}$ that can be expressed as $\operatorname{ad} A$ for some $A \in N(\mathfrak{t})$. The following states basic properties of the Weyl group [91, p.174]

Proposition 3.5.2. *1. The inner product $\langle \cdot, \cdot \rangle$ on $\mathfrak{h}$ is invariant under the action of W.*

2. The set $\Phi \subseteq \mathfrak{h}$ of roots is invariant under the action of W.

3. The set of co-roots is invariant under the action of W, and $w \cdot h_\alpha = h_{w \cdot \alpha}$, for all $w \in W$, $\alpha \in \Phi$.

4. The Weyl group is a finite group.

We state some an important property, leading to a 'dual' nature of roots [91, p.178]:

Proposition 3.5.3. *For each root α, there exists an element w_α of W such that*

$$w_\alpha \cdot \alpha = -\alpha$$

and such that

$$w_\alpha \cdot h = h,$$

for all $h \in \mathfrak{h}$ with $\langle \alpha, h \rangle = 0$.

Note that since h_α is a multiple of α, saying $w_\alpha \cdot \alpha = -\alpha$ is equivalent to saying that $w_\alpha \cdot h_\alpha = -h_\alpha$. The linear operator corresponding to the action of w_α on $\mathfrak{h}$ is 'the reflection about the hyperplane perpendicular to α'. This means that w_α acts as the identity on the hyperplane (of codimension one) perpendicular to α and as minus the identity on the span of α. We can work out a formula for w_α as follows. Any vector β can be decomposed uniquely as a multiple of α plus a vector orthogonal to α. This decomposition is given explicitly by

$$\beta = \frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}\alpha + \left(\beta - \frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}\alpha\right), \tag{3.4}$$

where the second term is indeed orthogonal to α. Now, to obtain $w_\alpha \cdot \beta$, we should change the sign of the part of β parallel to α and leave alone the part of β that is orthogonal to α. This means that we change the sign of the first term on the right-hand side of (3.4), giving

$$w_\alpha \cdot \beta = \beta - 2\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}\alpha. \tag{3.5}$$

We now have another way of thinking about the quantity $2\frac{\langle \alpha, \beta \rangle}{\langle \alpha, \alpha \rangle}$ in Theorem 3.4.4: it is the coefficient of α in the expression for $w_\alpha \cdot \beta$. So, we can re-express Theorem 3.4.4 as follows.

Corollary 3.5.4. *If α and β are roots, then $\beta - w_\alpha \cdot \beta$ is an integer multiple of α.*

Finally, we state a useful characterization of the Weyl group [116, p.208]:

Theorem 3.5.5. *The Weyl group W is generated by the elements w_α as α ranges over all roots.*

That is to say, the smallest subgroup of W that contains all of the w_α's is W itself. This is somewhat involved to prove and we will not do so here; see [22] or [116]. In the Lie algebra approach to the Weyl group, the Weyl group is defined as the set of linear transformations of $\mathfrak{h}$ generated by the reflections w_α. Theorem 3.5.5 shows that the Lie algebra definition of the Weyl group gives the same group as the compact-group approach.

3.6 Root Systems

In the previous section we have established several properties of roots. For example, we know that the roots are imaginary on $\mathfrak{t}$, which, after transferring the roots from $\mathfrak{h}^*$ to $\mathfrak{h}$, means that the roots live in $i\mathfrak{t} \subseteq \mathfrak{h}$. The inner product $\langle \cdot, \cdot \rangle$ was constructed to take real values on $\mathfrak{k}$, and hence on $\mathfrak{t}$. The inner product then also takes real values on $i\mathfrak{t}$, since $\langle ix, iy \rangle = (-i)i\langle w, y \rangle = \langle x, y \rangle$. So, the roots live in the real inner-product space $E = i\mathfrak{t}$. From Proposition 3.4.3 we know that the roots span $i\mathfrak{t}$ and that if α is a root, then $-\alpha$ is the only other multiple of α also root. Furthermore, Theorem 3.4.4 tell us that for any roots α and β, the number $2\frac{\langle \alpha,\beta \rangle}{\langle \alpha,\alpha \rangle}$ is an integer. Finally, we have established that the roots are invariant under the action of the Weyl group, and Theorem 3.5.5 tells us that the Weyl group contains the reflection about the hyperplane orthogonal to each root α. We summarize these results in the following theorem.

Theorem 3.6.1. *The roots of $\mathfrak{g}$ form a finite set of nonzero elements of a real inner-product space E and have the following properties:*

1. *The roots span E.*

2. *If α is a root, then $-\alpha$ is a root and the only multiples of α that are roots are α and $-\alpha$.*

3. *If α is a root, let w_α denote the linear transformation of E given by $w_\alpha \cdot \beta = \beta - 2\frac{\langle \alpha,\beta \rangle}{\langle \alpha,\alpha \rangle}\alpha$. Then, for all roots α and β, $w_\alpha \cdot \beta$ is also a root.*

4. *If α and β are roots, then the quantity $2\frac{\langle \alpha,\beta \rangle}{\langle \alpha,\alpha \rangle}$ is an integer.*

Definition 3.6.2. Any collection R of vectors in a finite-dimensional real inner-product space having these properties of Theorem 3.6.1 is called a *root system*.

The Weyl group for a root system R is the group of linear transformations of E generated by the w_α's. Note that item 4 is equivalent to saying that $\beta - w_\alpha \cdot \beta$ must be an integer multiple of α for all roots α and β. We have also established certain important properties of the root spaces that are not properties of the roots themselves, namely that each root space $\mathfrak{g}_\alpha$ is one dimensional and that out of $\mathfrak{g}_\alpha$, $\mathfrak{g}_{-\alpha}$, and $[\mathfrak{g}_\alpha, \mathfrak{g}_{-\alpha}]$, we can form a subalgebra isomorphic to $\mathfrak{sl}_2(\mathbb{C})$. Finally, we claim that the co-roots h_α themselves form a root system. Theorem 3.4.4 tells us that the co-roots satisfy property 4 and Proposition 3.5.2 tells us that the set of co-roots is invariant under the Weyl group and hence, in particular, under the reflections w_α. However, note that since h_α is a multiple of α, the reflection generated by h_α is the same as the reflection generated by α. Thus, the set of co-roots satisfies property 3. Properties 1 and 2 for the co-roots follow from the corresponding properties for the roots, since each h_α is a multiple of α. The set of co-roots is called the *dual root system* to the set of roots. See Section 3.9 for more information on root systems, including some pictures.

In the next section, we will present the irreducible representations of $\mathfrak{g}$ in terms of a 'highest weight'. What we need is simply some consistent notion of higher and lower that will allow us to divide the root vectors x_α into 'raising operators' and 'lowering operators'. This should be done in such a way that the commutator of two raising operators is, again, a raising operator and not a lowering operator. This means that we want to divide the roots into two groups, one of which will be called 'positive' and the other 'negative'. This should be done is such a way that if the sum of positive roots is again a root, that root should be positive. There is no unique way to make the division into positive and negative; any consistent division will do. The uniqueness theorems of the next section show that it does not really matter which choice we make. The following definition and theorem shows that it is possible to make a good choice.

Definition 3.6.3. Suppose that E is a finite-dimensional real inner-product space and that $R \subseteq E$ is a root system. Then, a *base* for R is a subset $\Delta = \{\alpha_1, \ldots, \alpha_r\}$ of R such that Δ forms a basis for E as a vector space and such that for each $\alpha \in R$, we have

$$\alpha = n_1\alpha_1 + \ldots + n_r\alpha_r,$$

where $n_j \in \mathbb{Z}$ and either all $n_j \geq 0$ or all $n_j \leq 0$.

Once a base Δ has been chosen, the α's for which $n_j \geq 0$, $\forall j$, are called the *positive* roots (with respect to the given choice of Δ) and the α's with $n_j \leq 0$, $\forall j$, are called the *negative* roots. The elements of Δ are called the *positive simple roots*. We will denote R^+ the set of all positive roots, and R^- the set of all negative roots, so then R is the disjoint union $R = R^+ \cup R^-$. To be a base (in the sense of root systems), $\Delta \subseteq R$ must in particular be a basis for E in the vector space sense. In addition, the expansion of any $\alpha \in R$ in terms of the elements of Δ must have integer coefficients and all of the nonzero coefficients must be of the same sign.

3.7 Integral and Dominant integral elements

Definition 3.7.1. An element ω of $\mathfrak{h}$ is called an *integral* element if $\langle \omega, h_\alpha \rangle$ is an integer, for each root α.

As explained in next, the integral elements are precisely the elements of $\mathfrak{h}$ that arise as weights of finite-dimensional representations of $\mathfrak{g}$. In fact, we can prove that the set of integral elements is invariant under the action of the Weyl group. Checking that (ω, h_α) is an integer for every root α is a rather tiresome process. Fortunately, it suffices to check just for the positive simple roots: if ω is an element of $\mathfrak{h}$ for which $\langle\omega, h_\alpha\rangle$ is an integer for all positive simple roots α, then $\langle\omega, h_\alpha\rangle$ is an integer for all roots α, and thus ω is an integral element.

Recalling the expression (3.3) for h_α in terms of α, we may restate a characterization of integral elements as follows: an element $\omega \in \mathfrak{h}$ is integral if and only if

$$2\frac{\langle\omega, \alpha\rangle}{\langle\alpha, \alpha\rangle}$$

is an integer for each positive simple root α. In particular, every root is an integral element. Recall now from elementary linear algebra that if ω and α are any two elements of an inner-product space, then the orthogonal projection of ω onto α is given by $\frac{\langle\omega,\alpha\rangle}{\langle\alpha,\alpha\rangle}\alpha$. Thus, we may reformulate the notion of an integral element yet again as: ω is integral if and only if the orthogonal projection of ω onto each positive simple root α is an integer or half-integer multiple of α. This characterization of the integral elements will help us visualize graphically what the set of integral elements looks like in example (see Section 3.9).

Definition 3.7.2. An element ω of $\mathfrak{h}$ is called a *dominant integral* element if $\langle\omega, h_\alpha\rangle$ is a non-negative integer for each positive simple root α.

Equivalently, ω is a dominant integral element if $2\frac{\langle\omega,\alpha\rangle}{\langle\alpha,\alpha\rangle}$ is a non-negative integer for each positive simple root α. If ω is dominant integral, then $\langle\omega, h_\alpha\rangle$ will automatically be a non-negative integer for each positive root α, not just the positive simple ones.

Definition 3.7.3. Suppose π is a finite-dimensional representation of $\mathfrak{g}$ on a vector space V. Then, $\omega \in \mathfrak{h}$ is called a *weight* for π if there exists a nonzero vector $v \in V$ such that

$$\pi(h)v = \langle\omega, h\rangle v,$$

for all $h \in \mathfrak{h}$. A nonzero vector v satisfying condition above is called a *weight vector* for the weight ω, and the set of all vectors (zero or nonzero) satisfying this condition is called the *weight space* with weight ω. The dimension of the weight space is called the *multiplicity* of the weight.

To understand this definition, suppose that $v \in V$ is a simultaneous eigenvector for each $\pi(h)$, $h \in \mathfrak{h}$. This means that for each $h \in \mathfrak{h}$, there is a number λ_h such that $\pi(h)v = \lambda_h v$. Since the representation $\pi(h)$ is linear in h, the λ_h's must depend linearly on h as well; that is, the map $h \mapsto \lambda_h$ is a linear functional on $\mathfrak{h}$. Then, there is a unique element ω of $\mathfrak{h}$ such that $\lambda_h = \langle\omega, h\rangle$. Thus, a weight vector is nothing but a simultaneous eigenvector for all the λ_h's and the vector ω is simply a convenient way of encoding the eigenvalues. Note that the roots (in the dual notation) are precisely the nonzero weights of the adjoint representation of $\mathfrak{g}$. It can be shown that two equivalent representations have the same weights and multiplicities. It is true, although by no means obvious, that every integral element actually arises as a weight of some finite-dimensional representation of $\mathfrak{g}$, see [91]. Then, for any finite-dimensional representation π of $\mathfrak{g}$, the weights of π and their multiplicity are invariant under the action of the Weyl group.

Definition 3.7.4. Let ω_1 and ω_2 be two elements of $\mathfrak{h}$. We say that ω_1 is *higher* than ω_2 (or, equivalently, ω_2 is lower than ω_1) if there exist non-negative real numbers $a_1, \ldots, a_r$ such that

$$\omega_1 - \omega_2 = a_1\alpha_1 + \ldots + a_r\alpha_r,$$

where $\Delta = \{\alpha_1, \ldots, \alpha_r\}$ is the set of positive simple roots. This relationship is often written as $\omega_1 \succeq \omega_2$ or $\omega_2 \preceq \omega_1$.

If π is a representation of $\mathfrak{g}$, then a weight ω_0 for π is said to be a *highest weight* if for all weights ω of π, $\omega \preceq \omega_0$. We now state an important result of this chapter [91, p.197]

Theorem 3.7.5 (Theorem of the Highest Weight)**.** *We have:*
1. Every irreducible representation has a highest weight.
2. Two irreducible representations with the same highest weight are equivalent.
3. The highest weight of every irreducible representation is a dominant integral element.
4. Every dominant integral element occurs as the highest weight of an irreducible representation.

3.8 The Weyl character formula

Let Σ be a finite-dimensional irreducible representation of K acting on a vector space V, then we consider the space of matrix entries of Σ. Suppose we choose a basis $\{u_k\}$ for V. Then, for each $x \in K$, the linear operator $\Sigma(x)$ can be expressed as a matrix with respect to this basis; we denote the entries of this matrix as $\Sigma_{k\ell}$. Then, a matrix entry for Σ is a function on K that can be expressed in the form

$$f(x) = \sum_{k,\ell=1}^{\dim V} \alpha_{k\ell}\Sigma(x)_{k\ell}, \tag{3.6}$$

for some set of constants $a_{k\ell}$. We can describe the space of matrix entries in a basis-independent way as the space of functions that can be expressed in the form

$$f(x) = \mathrm{tr}(\Sigma(x)A) = \mathrm{trace}(\Sigma(x)A), \tag{3.7}$$

for some linear operator A on V. To see the equivalence of these two forms, let $A_{k\ell}$ be the matrix for the operator A in the basis $\{u_k\}$. Then, the matrix for $\Sigma(x)A$ is given by the matrix product $(\Sigma(x)A)_{k\ell} = \sum_m \Sigma(x)_{km}A_{m\ell}$, so

$$\operatorname{tr}(\Sigma(x)A) = \sum_{k,m=1}^{\dim V} \Sigma(x)_{km}A_{m\ell}.$$

Thus, every function of the form (3.7) can be expressed in form (3.6) with $\alpha_{k\ell} = A_{\ell k}$, and conversely.

Definition 3.8.1. Let K be a simply-connected compact Lie group and let Σ be a finite-dimensional irreducible representation of K. Then, the *character* of Σ is the function on K given by

$$\operatorname{char}\Sigma(x) = \operatorname{tr}(\Sigma(x)) = \operatorname{trace}(\Sigma(x)).$$

This function is a matrix entry, obtained by taking $A = I$ in (3.7) or taking $\alpha_{k\ell} = \delta_{k\ell}$ in (3.6). The character is special because it satisfies

$$\operatorname{tr}(\Sigma(xyx^{-1})) = \operatorname{tr}(\Sigma(x)\Sigma(y)\Sigma(x)^{-1}) = \operatorname{tr}(\Sigma(y)),$$

for all $x, y \in K$. Recall from Section 2.1 that any function f satisfying $f(xyx^{-1}) = f(y)$, $\forall x, y \in K$, is called a class function (constant on each conjugacy class of K). Only as a note, the Peter-Weyl Theorem states that the family of class functions $\operatorname{tr}\Sigma$ forms an orthonormal basis for $L^2(K, \mu)$, where Σ ranges over the equivalence classes of irreducible finite-dimensional representations of K.

We now assume that K is a simply-connected compact Lie group (there is also a version of the result for connected compact Lie groups that are not simply connected). We choose, as usual, a maximal commutative subalgebra $\mathfrak{t}$ of $\mathfrak{k}$ and we let T be the connected Lie subgroup of K whose Lie algebra is $\mathfrak{t}$. It can be shown that T is a closed subgroup of K (called a 'maximal torus'). It can further be shown that every element of K is conjugate to an element of T. This means that the values of a class function on K are, in principle, determined by its values on T. The Weyl character formula is a formula for the restriction to T of the character of an irreducible representation of K.
We let $\mathfrak{g}$ denote the complexification of the Lie algebra $\mathfrak{t}$ of K, so that $\mathfrak{g}$ is a complex semisimple Lie algebra. Then, $\mathfrak{h} = \mathfrak{t} + i\mathfrak{t}$ is a Cartan subalgebra in $\mathfrak{g}$. If we follow dual notation and regard the roots as elements of $\mathfrak{h}$ (not in $\mathfrak{h}^*$), we know that if $\alpha \in \mathfrak{h}$ is a root, then $\langle \alpha, h\rangle$ is imaginary for all $h \in \mathfrak{t}$, which means that α itself is in $i\mathfrak{t}$. It is then convenient to introduce the *real roots*, which are simply $\frac{1}{i}$ times the ordinary roots. This means that a real root is a nonzero element α of $\mathfrak{t}$ with the property that there exists a nonzero $x \in \mathfrak{g}$ with

$$[h, x] = i\langle \alpha, h\rangle x,$$

for all $h \in \mathfrak{t}$ (or, equivalently, for all $h \in \mathfrak{h}$). We can also introduce the *real co-roots* as the elements of $\mathfrak{t}$ of the form $h_\alpha = \frac{2\alpha}{\langle \alpha, \alpha \rangle}$, where α is a real root.

In the same way, we will consider the *real weights*, which we think of as elements of $\mathfrak{t}$ in the same way as for the roots. So, if (Σ, V) is an irreducible representation of $\mathfrak{g}$, then an element $\omega \in \mathfrak{t}$ is called a real weight for Σ if there exists a nonzero vector $v \in V$ such that

$$\sigma(h)v = i\langle \omega, h \rangle v,$$

for all $h \in \mathfrak{t}$. Here, σ is the Lie algebra representation associated to the group representation Σ. An element ω of $\mathfrak{t}$ is said to be *integral* if $\langle \omega, h_\alpha \rangle$ is an integer for each real co-root h_α. (All of the 'real' objects are simply $\frac{1}{i}$ times the corresponding objects without the qualifier 'real'). The real weights of any finite-dimensional representation of $\mathfrak{g}$ must be integral. For the rest of this section, all of roots and weights will be assumed real, even if this is not explicitly stated.

If α is an integral element, then it can be shown that there is a function f on T satisfying

$$f(e^h) = e^{i\langle \alpha, h \rangle}, \tag{3.8}$$

for all $h \in \mathfrak{h}$. Note that because T is connected and commutative, every element $t \in T$ can be expressed as $t = e^h$. However, a given t can be expressed as $t = e^h$ in many different ways; the content of the above assertion is that the right-hand side of (3.8) is independent of the choice of h for a given t. This means that we want to say that the right-hand side of (3.8) defines a function on T, not just on t.

Next, we introduce the element δ of $\mathfrak{t}$ defined to be half the sum of the positive roots:

$$\delta = \frac{1}{2} \sum_{\alpha \in \Phi^+} \alpha.$$

It can be shown that δ is an integral element. (Clearly, $\rho = 2\delta$ is integral, but it is not obvious that δ itself is integral). Finally, if w is any element of the Weyl group W, we think of w as an orthogonal linear transformation of $\mathfrak{t}$ in which case, $\epsilon(w) = \det(w) = \pm 1$. We are now ready to state the Weyl character formula [91, p.213].

Theorem 3.8.2 (Weyl Character Formula). *If Σ is an irreducible representation of K with highest real weight ω, then we have*

$$\operatorname{char} \Sigma(e^h) = \frac{\sum_{w \in W} \epsilon(w) e^{i\langle w\cdot(\omega+\rho), h\rangle}}{\sum_{w \in W} \epsilon(w) e^{i\langle w\cdot\rho, h\rangle}},$$

for all $h \in \mathfrak{t}$, for which the denominator of right-hand side above is nonzero. Or equivalently,

$$\left(\sum_{w \in W} \epsilon(w) e^{i\langle w\cdot\rho, h\rangle} \right) \operatorname{char} \Sigma(e^h) = \sum_{w \in W} \epsilon(w) e^{i\langle w\cdot(\omega+\rho), h\rangle},$$

for all $h \in \mathfrak{t}$. Here, ρ denotes the sum of the positive real roots.

The set of points h for which the denominator of the Weyl character formula, the so-called *Weyl denominator*, is nonzero is dense in $\mathfrak{t}$. At points where the denominator is zero, there is an apparent singularity in the formula for $\operatorname{char}\Sigma$. However, actually at such points the numerator is also zero and the character itself remains finite (as must be the case since, from the definition of the character, it is well defined and finite at every point). Note that the character formula gives a formula for the restriction of $\operatorname{char}\Sigma$ to T. Since $\operatorname{char}\Sigma$ is a class function and since (as we have asserted but not proved) every element of K is conjugate to an element of T, knowing $\operatorname{char}\Sigma$ on T determines, $\operatorname{char}\Sigma$ on all of K. A sketch of the proof of the Weyl character formula can be found in [22] or [91].

In fact, the expression $\sum_w \epsilon(w)e^{i\langle w\cdot\omega,h\rangle}$ appearing on denominator of the Weyl character formula can be written in an alternate way. This is established in the next result [116, p.264]

Theorem 3.8.3 (Weyl Denominator Identity). *On the same assumption of Weyl character formula, we have*

$$\sum_{w\in W}\epsilon(w)e^{i\langle w\cdot\rho,h\rangle} = e^{\rho}\prod_{\alpha\in\Phi^+}(1-e^{\alpha}).$$

3.9 Representations of $SU(3)$

As an illustration of the concepts introduced in this chapter, we will discuss the special case of the representation theory of $SU(3)$. The main result we have done in this chapter is Theorem 3.7.5, which states that an irreducible finite-dimensional representation of a semisimple Lie algebra can be classified in terms of its highest weight.

The group $SU(3)$ is simply connected, and so the finite-dimensional representations of $SU(3)$ are in one-to-one correspondence with the finite-dimensional representations of the Lie algebra $\mathfrak{su}(3)$. Meanwhile, the complex representations of $\mathfrak{su}(3)$ are in one-to-one correspondence with the complex-linear representations of the complexified Lie algebra $(\mathfrak{su}(3))_{\mathbb{C}} = \mathfrak{sl}_3(\mathbb{C})$ (Table 2.4). Moreover, a representation of $SU(3)$ is irreducible if and only if the associated representation of $\mathfrak{su}(3)$ is irreducible, and this holds if and only if the associated complex-linear representation of $\mathfrak{sl}_2(\mathbb{C})$ is irreducible. (This follows from Proposition 3.1.5, Proposition 3.1.6, and the connectedness of $SU(3)$). This correspondence is determined by the property that

$$\Pi(e^x) = e^{\pi(x)}, \quad \text{for all } x \in \mathfrak{su}(3) \subseteq \mathfrak{sl}(3).$$

Since $SU(3)$ is compact, Proposition 3.2.5 tells us that all of the finite-dimensional representations of $SU(3)$ are direct sums of irreducible representations. The above paragraph then implies that the same holds for $\mathfrak{sl}_3(\mathbb{C})$, that is, $\mathfrak{sl}_3(\mathbb{C})$ has the complete reducibility property. We can apply the same reasoning to the simply-connected group $SU(2)$, its Lie algebra $\mathfrak{su}(2)$, and its complexified Lie algebra $\mathfrak{sl}_2(\mathbb{C})$. Thus, every finite-dimensional representation of $\mathfrak{sl}_2(\mathbb{C})$ or $\mathfrak{sl}_3(\mathbb{C})$ decomposes as a direct sum of irreducible invariant subspaces.

We will use the following basis for $\mathfrak{sl}_3(\mathbb{C})$:

$$h_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad h_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix};$$
$$x_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad x_3 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix};$$
$$y_1 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad y_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad y_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Note that the span of $\{h_1, x_1, y_1\}$ is a subalgebra of $\mathfrak{sl}_3(\mathbb{C})$ which is isomorphic to $\mathfrak{sl}_2(\mathbb{C})$ (Example 3.1.10) by ignoring the third row and column in each matrix. Similarly, the span of $\{h_2, x_2, y_2\}$ is a subalgebra isomorphic to $\mathfrak{sl}_2(\mathbb{C})$. Thus, we have the following commutation relations:

$$\begin{aligned} [h_1, x_1] &= 2x_1, & [h_2, x_2] &= 2x_2, \\ [h_1, y_1] &= -2y_1, & [h_2, y_2] &= -2y_2, \\ [x_1, y_1] &= h_1, & [x_2, y_2] &= h_2. \end{aligned} \tag{3.9}$$

We now list all of the commutation relations among the basis elements which involve at least one of h_1 and h_2 (this includes some repetitions of the above commutation relations).

$$[h_1, h_2] = 0;$$

$$\begin{aligned} &[h_1, x_1] = 2x_1, \quad [h_1, y_1] = -2y_1, \\ &[h_2, x_1] = -x_1, \quad [h_2, y_1] = y_1; \end{aligned}$$

$$\begin{aligned} &[h_1, x_2] = -x_2, \quad [h_1, y_2] = y_2, \\ &[h_2, x_2] = 2x_2, \quad [h_2, y_2] = -2y_2; \end{aligned}$$

$$\begin{aligned} &[h_1, x_3] = x_3, \quad [h_1, y_3] = -y_3, \\ &[h_2, x_3] = x_3, \quad [h_2, y_3] = -y_3. \end{aligned}$$

Finally, we list all of the remaining commutation relations.

$$\begin{aligned}
&[x_1, y_1] = h_1, \quad [x_2, y_2] = h_2\\
&[x_3, y_3] = h_1 + h_2;\\
\\
&[x_1, x_2] = x_3, \quad [y_1, y_2] = -y_3,\\
&[x_1, y_2] = 0, \quad [x_2, y_1] = 0;\\
\\
&[x_1, x_3] = 0, \quad [y_1, y_3] = 0,\\
&[x_2, x_3] = 0, \quad [y_2, y_3] = 0;\\
\\
&[x_2, y_3] = y_1, \quad [x_3, y_2] = x_1,\\
&[x_1, y_3] = -y_2, \quad [x_3, y_1] = -x_2.
\end{aligned}$$

All of the analysis we will do for the representations of $\mathfrak{sl}_3(\mathbb{C})$ will be in terms of the above basis. From now on, all representations of $\mathfrak{sl}_3(\mathbb{C})$ will be assumed to be finite dimensional and complex linear.

Now, denote by $\mathfrak{h}$ the complex subalgebra generated by the elements $\{h_1, h_2\}$. Observe from the relations listed above that:
(i) $[h_1, h_2] = 0$;
(ii) $[x_i, h_j] \neq 0$ and $[y_i, h_j] \neq 0$, for $i = 1, 2, 3$ and $j = 1, 2$;
(iii) The operators $\operatorname{ad} h_1$ and $\operatorname{ad} h_2$ are diagonalizable.
Thus, the algebra $\mathfrak{h}$ satisfies the conditions (i), (ii) and (iii) of Definition 3.3.1, so $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{sl}_3(\mathbb{C})$.

Recall from Section 3.7 that a weight for a representation π of the algebra $\mathfrak{sl}_3(\mathbb{C})$ (respect to the Cartan subalgebra $\mathfrak{h}$) is an element $\omega \in \mathfrak{sl}_3(\mathbb{C})$ such that there exists a vector v satisfying

$$\pi(h)v = \langle \omega, h \rangle v, \quad \text{for all } h \in \mathfrak{h}.$$

If we denote $m_1 = \langle \omega, h_1 \rangle$ and $m_2 = \langle \omega, h_2 \rangle$, then we can see ω as the ordered pair $\omega = (m_1, m_2)$, and the above condition says that ω is a weight if

$$\pi(h_1)v = m_1 v, \quad \pi(h_2)v = m_2 v. \tag{3.10}$$

A nonzero vector v satisfying relations (3.10) is called a weight vector corresponding to the weight $\omega = (m_1, m_2)$. Recall also that the multiplicity of ω a weight is the dimension of the corresponding weight space, *i. e.* the space of all vectors v satisfying (3.10). Thus, a weight is simply a pair of simultaneous eigenvalues for $\pi(h_1)$ and $\pi(h_2)$. It can be shown that equivalent representations have the same weights and multiplicities.

Here is the advantage of work with the complexification of Lie algebras: since we are working over the complex numbers, $\pi(h_1)$ has at least one eigenvalue $m_1 \in \mathbb{C}$. If $W \subseteq V$ is the eigenspace for $\pi(h_1)$ with eigenvalue m_1, since $[h_1, h_2] = 0$, $\pi(h_2)$ commutes with $\pi(h_1)$, and, so, $\pi(h_2)$ must map W into itself. Thus, $\pi(h_2)$ can be viewed as an operator on W, and its restriction of to W must have at least one eigenvector w with eigenvalue $m_2 \in \mathbb{C}$, giving w a simultaneous eigenvector for $\pi(h_1)$ and $\pi(h_2)$ with eigenvalues m_1 and m_2, respectively. Hence, we have

Proposition 3.9.1. *Every representation π of $\mathfrak{sl}_3(\mathbb{C})$ has at least one weight.*

Now, every representation π of $\mathfrak{sl}_3(\mathbb{C})$ can be viewed, by restriction, as a representation of the subalgebra $\{h_1, x_1, y_1\} \cong \mathfrak{sl}_2(\mathbb{C})$. Note that even if π is irreducible as a representation of $\mathfrak{sl}_3(\mathbb{C})$, there is no reason to expect that it will still be irreducible as a representation of the subalgebra $\{h_1, x_1, y_1\}$. Nevertheless, π restricted to $\{h_1, x_1, y_1\}$ must be some finite-dimensional representation of $\mathfrak{sl}_2(\mathbb{C})$. The same reasoning applies to the restriction of π to the subalgebra $\{h_2, x_2, y_2\}$, which is also isomorphic to $\mathfrak{sl}_2(\mathbb{C})$. Now, recall Theorem 3.1.11, which tells us that in any finite-dimensional representation of $\mathfrak{sl}_2(\mathbb{C})$, irreducible or not, all of the eigenvalues of $\pi(h)$ must be integers. Applying this result to the restriction of π to $\{h_1, x_1, y_1\}$ and to the restriction of π to $\{h_2, x_2, y_2\}$, we can state the following corollary

Corollary 3.9.2. *If π is a representation of $\mathfrak{sl}_3(\mathbb{C})$, then all of the weights of π are of the form $\omega = (m_1, m_2)$, with m_1 and m_2 being integers.*

Recall now that a root of $\mathfrak{sl}_3(\mathbb{C})$ (relative to the Cartan subalgebra $\mathfrak{h}$) is a nonzero linear functional α on $\mathfrak{h}$ such that there exists a nonzero element z of $\mathfrak{sl}_3(\mathbb{C})$ with

$$[h, z] = \alpha(h)z, \quad \text{for all } h \in \mathfrak{h}.$$

We can see α as an ordered pair $\alpha = (a_1, a_2) \in \mathbb{C}^2$, and the above condition says that a nonzero α is a root of $\mathfrak{sl}_3(\mathbb{C})$ if

$$[h_1, z] = a_1 z, \quad [h_2, z] = a_2 z. \tag{3.11}$$

The element z satisfying relations (3.11) is called a root vector corresponding to the root $\alpha = (a_1, a_2)$. Thus, z is a simultaneous eigenvector for $\operatorname{ad} h_1$ and $\operatorname{ad} h_2$. This means that z is a weight vector for the adjoint representation with weight (a_1, a_2). Taking into account the nonzero condition for (a_1, a_2), we may say that the roots are precisely the nonzero weights of the adjoint representation. Corollary 3.9.2 then tells us that for any root, both a_1 and a_2 must be integers, which we can also see directly in Table 3.3. The commutation relations (3.9) tell us what the roots for $\mathfrak{sl}_3(\mathbb{C})$ are. There are six roots:
Note that h_1 and h_2 are also simultaneous eigenvectors for $\operatorname{ad} h_1$ and $\operatorname{ad} h_2$, but they are not root vectors because the simultaneous eigenvalues are both zero. Since the vectors in Table 3.3 together with h_1 and h_2 form a basis for $\mathfrak{sl}_3(\mathbb{C})$, it is not hard to show that the roots listed in Table 3.3 are the only roots. These six roots form a root system, conventionally

root α	root vector z
$(2,-1)$	x_1
$(-1,2)$	x_2
$(1,1)$	x_3
$(-2,1)$	y_1
$(1,-2)$	y_2
$(-1,-1)$	y_3

Table 3.3: Roots for the Lie algebra $\mathfrak{sl}_3(\mathbb{C})$.

called A_2. It is convenient to single out a basis Δ consisting on the two roots corresponding to x_1 and x_2 and give them special names:

$$\begin{aligned} \alpha_1 &= (2,-1), \\ \alpha_2 &= (-1,2). \end{aligned} \tag{3.12}$$

The roots α_1 and α_2 are called the positive simple roots, because they have the property that all of the roots can be expressed as linear combinations of α_1 and α_2 with integer coefficients, and these coefficients are (for each root) either all nonnegative or nonpositive. This is verified by direct computation:

root α	linear combination z
$(2,-1)$	α_1
$(-1,2)$	α_2
$(1,1)$	$\alpha_1+\alpha_2$
$(-2,1)$	$-\alpha_1$
$(1,-2)$	$-\alpha_2$
$(-1,-1)$	$-\alpha_1-\alpha_2$

Table 3.4: Roots for $\mathfrak{sl}_3(\mathbb{C})$ in terms of the basis $\Delta=\{\alpha_1,\alpha_2\}$.

The decision to designate α_1 and α_2 as the positive simple roots is arbitrary; any other pair of roots with similar properties would do just as well. For example, if we set $\alpha_3=\alpha_1+\alpha_2$, then we can define $\Delta'=\{\alpha_1,\alpha_3\}$ as another basis. Figure 3.1 shows the root system for $\mathfrak{sl}_3(\mathbb{C})$.

The significance of the roots for the representation theory of $\mathfrak{sl}_3(\mathbb{C})$ is contained in the following lemma.

root α	linear combination z
$(2,-1)$	α_1
$(-1,2)$	$-\alpha_1-\alpha_3$
$(1,1)$	α_3
$(-2,1)$	$-\alpha_1$
$(1,-2)$	$\alpha_1+\alpha_3$
$(-1,-1)$	$-\alpha_3$

Table 3.5: Roots for $\mathfrak{sl}_3(\mathbb{C})$ in terms of the basis $\Delta' = \{\alpha_1, \alpha_3\}$.

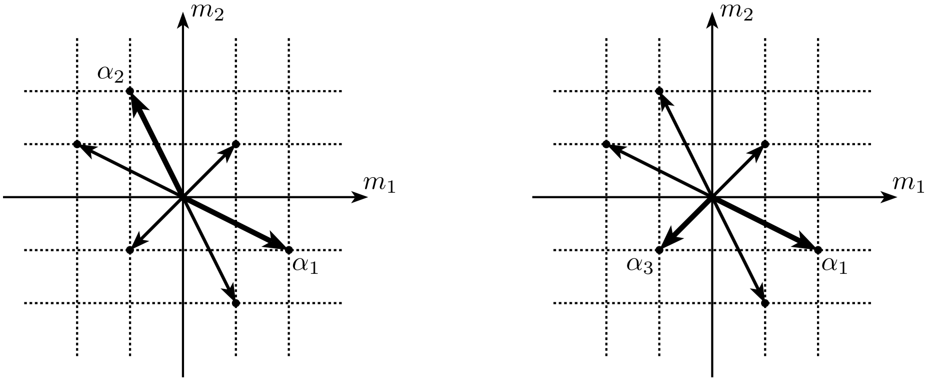

Figure 3.1: The root system A_2 for the Lie algebra $\mathfrak{sl}_3(\mathbb{C})$.

Lemma 3.9.3. *Let $\alpha = (a_1, a_2)$ be a root and z_α a corresponding root vector in $\mathfrak{sl}_3(\mathbb{C})$. Let π be a representation of $\mathfrak{sl}_3(\mathbb{C})$, $\omega = (m_1, m_2)$ a weight for π, and $v \neq 0$ a corresponding weight vector. Then,*

$$\begin{aligned}\pi(h_1)\pi(z_\alpha)v &= (m_1+a_1)\pi(z_\alpha)v,\\ \pi(h_2)\pi(z_\alpha)v &= (m_2+a_2)\pi(z_\alpha)v.\end{aligned}$$

Thus, either $\pi(z_\alpha)v = 0$ or $\pi(z_\alpha)v$ is a new weight vector with weight $\omega + \alpha = (m_1 + a_1, m_2 + a_2)$.

Proof. In fact, the definition of a root tells us that we have the commutation relation $[h_1, z_\alpha] = a_1 z_\alpha$. Thus,

$$\begin{aligned}\pi(h_1)\pi(z_\alpha)v &= \big(\pi(z_\alpha)\pi(h_1) + a_1\pi(z_\alpha)\big)v = \pi(z_\alpha)(m_1 v) + a_1\pi(z_\alpha)v\\ &= (m_1+a_1)\pi(z_\alpha)v,\end{aligned}$$

and a similar argument allows us to compute $\pi(h_2)\pi(z_\alpha)v$. □

We see then that if we have a representation with a weight $\omega = (m_1, m_2)$, then by applying the root vectors x_1, x_2, x_3, y_1, y_2 and y_3, we can get some new weights of the form $\omega + \alpha$, where α is the root. Of course, some of the time $\pi(z_\alpha)v$ will be zero, in which case $\omega + \alpha$ is not necessarily a weight. In fact, since our representation is finite dimensional, there can be only finitely many weights, so we must get zero quite often. Now we would like to single out in each representation a highest weight. Recall from a previous section that if ω_1 and ω_2 are two weights, then ω_1 is higher than ω_2 ($\omega_1 \succ \omega_2$) if $\omega_1 - \omega_2$ can be written in the form

$$\omega_1 - \omega_2 = a_1\alpha_1 + a_2\alpha_2, \quad \text{with } a_1, a_2 \geq 0,$$

and recall that maximal elements of this partial relation are called highest weights. Note that the relation of 'higher' is only a partial ordering; that is, one can easily have ω_1 and ω_2 such that $\omega_1 \not\succ \omega_2$ neither $\omega_1 \not\prec \omega_2$. For example, $\alpha_1 - \alpha_2$ is neither higher nor lower than 0. This, in particular, means that a finite set of weights need not have a highest element (*e. g.*, the set $\{0, \alpha_1 - \alpha_2\}$ has no highest element). Note also that the coefficients a_1 and a_2 do not have to be integers, even if both ω_1 and ω_2 have integer entries. For example, $(1, 0)$ is higher than $(0, 0)$ since $(1, 0) - (0, 0) = (1, 0) = \frac{2}{3}\alpha_1 + \frac{1}{3}\alpha_2$. Recall also that an ordered pair (m_1, m_2) with m_1 and m_2 being non-negative integers is called a dominant integral element.

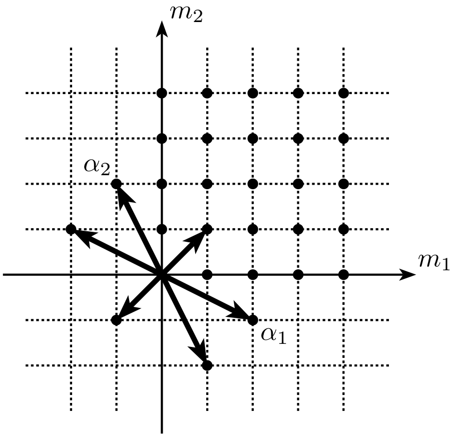

Figure 3.2: Roots and dominant integral elements for $\mathfrak{sl}_3(\mathbb{C})$.

Theorem 3.7.5 tell us that the highest weight of each irreducible representation of $\mathfrak{sl}_3(\mathbb{C})$ is a dominant integral element and, conversely, that every dominant integral element occurs as the highest weight of some irreducible representation. Since $(1, 0) = \frac{2}{3}\alpha_1 + \frac{1}{3}\alpha_2$ and $(0, 1) = \frac{1}{3}\alpha_1 + \frac{2}{3}\alpha_2$, we see that every dominant integral element is higher than zero. However, if ω has integer coefficients and is higher than zero, this does not necessarily mean that ω is dominant integral (for example, $\alpha_2 = (2, -1)$ is higher than zero, but is not dominant integral). Figure 3.2 shows the roots and dominant integral elements for $\mathfrak{sl}_3(\mathbb{C})$. This picture is made using the obvious basis for the space of weights; that is, the

x-coordinate is the eigenvalue of h_1 and the y-coordinate is the eigenvalue of h_2.

It is possible to obtain much more information about the irreducible representations besides the highest weight. For example, we have the following formula for the dimension of the representation with highest weight (m_1, m_2). It is a consequence of the Weyl character formula (see [100]):

Theorem 3.9.4. *The irreducible representation with highest weight* (m_1, m_2) *has dimension*

$$\tfrac{1}{2}(m_1 + l)(m_2 + 1)(m_1 + m_2 + 2).$$

There is an important symmetry to the representations of $\mathfrak{sl}_3(\mathbb{C})$ involving the Weyl group. To understand the idea behind the Weyl group symmetry, let us observe that the representations of $\mathfrak{sl}_3(\mathbb{C})$ are, in a certain sense, invariant under the adjoint action of $SU(3)$. This means the following: let π be a finite-dimensional representation of $\mathfrak{sl}_3(\mathbb{C})$ acting on a vector space V and let Π be the associated representation of $SU(3)$ acting on the same space. For any $A \in SU(3)$, we can define a new representation π_A of $\mathfrak{sl}_3(\mathbb{C})$, acting on the same vector space V, by setting

$$\pi_A(x) = \pi(AxA^{-1}).$$

Since the adjoint action of A on $\mathfrak{sl}_3(\mathbb{C})$ is a Lie algebra automorphism, this is, again, a representation of $\mathfrak{sl}_3(\mathbb{C})$. This new representation is to be equivalent to the original representation; and direct calculation shows that $\Pi(A)$ is an intertwining map between (π, V) and (π_A, V). We may say, then, that the adjoint action of $SU(3)$ is a symmetry of the set of equivalence classes of representations of $\mathfrak{sl}_3(\mathbb{C})$.

Now, we have analyzed the representations of $\mathfrak{sl}_3(\mathbb{C})$ by simultaneously diagonalizing the operators $\pi(h_1)$ and $\pi(h_2)$. Of course, this means that any linear combination of $\pi(h_1)$ and $\pi(h_2)$ is also simultaneously diagonalized. So, what really counts is the two-dimensional subspace $\mathfrak{h}$ of $\mathfrak{sl}_3(\mathbb{C})$ spanned by h_1 and h_2, the Cartan subalgebra of $\mathfrak{sl}_3(\mathbb{C})$. In general, the adjoint action of $A \in SU(3)$ does not preserve the space $\mathfrak{h}$ and so the equivalence of π and π_A does not (in general) tell us anything about the weights of π. However, there are elements $A \in SU(3)$ for which $\operatorname{ad} A$ does preserve $\mathfrak{h}$. We have already seen that these elements make up the Weyl group for $SU(3)$ and give rise to a symmetry of the set of weights of any representation π. So, we may say that the Weyl group is the 'residue' of the adjoint symmetry of the representations (discussed in the previous paragraph) that is left after we focus our attention on the Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{sl}_3(\mathbb{C})$.

In the case of the two-dimensional subspace of $\mathfrak{sl}_3(\mathbb{C})$ spanned by h_1 and h_2, let Z be the subgroup of $SU(3)$ consisting of those $A \in SU(3)$ such that $\operatorname{Ad} A(h) = h$, for all $h \in \mathfrak{h}$. Let N be the subgroup of $SU(3)$ consisting of those $A \in SU(3)$ such that $\operatorname{Ad} A(h)$ is an element of $\mathfrak{h}$, for all $h \in \mathfrak{h}$. In fact, Z and N are actually subgroups of $SU(3)$ and Z is a normal subgroup of N. This leads us to the Weyl group $W = N/Z$ of $SU(3)$. We can define an action of W on $\mathfrak{h}$ as follows. For each element w of W, choose an element A of the corresponding equivalence class in N. Then for $h \in \mathfrak{h}$ we define the action $w \cdot h$ by

$$w \cdot h = \operatorname{Ad} A(h).$$

To see that this action is well defined, suppose B is another element of the same equivalence class as A. Then $B = AC$, with $C \in Z$ and thus, $\operatorname{Ad} B(h) = \operatorname{Ad} AC(h) = \operatorname{Ad} A \operatorname{Ad} C(h) = \operatorname{Ad} A(h)$, by the definition of Z. It can be proved that W is isomorphic to the group of linear transformations of $\mathfrak{h}$ that can be expressed as Ad_A for some $A \in N$. In fact, the group Z consists precisely of the diagonal matrices inside $SU(3)$, namely the matrices of the form

$$A = \begin{pmatrix} e^{i\theta} & 0 & 0 \\ 0 & e^{i\phi} & 0 \\ 0 & 0 & e^{-i(\theta+\phi)} \end{pmatrix}, \quad \text{for } \theta, \phi \in \mathbb{R}.$$

The group N consists of precisely those matrices $A \in SU(3)$ such that for each $k = 1, 2, 3$ there exist $\ell \in \{1, 2, 3\}$ and $\phi \in \mathbb{R}$ such that $Ae_k = e^{i\theta} e_\ell$. Here, $\{e_1, e_2, e_3\}$ is the standard basis for $\mathbb{C}^3$. Hence, the Weyl group $W = N/Z$ is isomorphic to the permutation group on three elements (see [91] for a proof).

In the case of $SU(3)$, it is possible to identify the Weyl group with a certain subgroup of N, instead of as the quotient group N/Z. We want to show that the Weyl group is a symmetry of the weights of any finite-dimensional representation of $\mathfrak{sl}_3(\mathbb{C})$. To understand this, we adopt a less basis-dependent view of the weights. We have defined a weight as a pair (m_1, m_2) of simultaneous eigenvalues for $\pi(h_1)$ and $\pi(h_2)$. However, if a vector v is an eigenvector for $\pi(h_1)$ and $\pi(h_2)$ then it is also an eigenvector for $\pi(h)$ for any element h of the space $\mathfrak{h}$ spanned by h_1 and h_2. Furthermore, the eigenvalues must depend linearly on h since if h and j are any two elements of $\mathfrak{h}$ and $\pi(h)v = \lambda_1 v$ and $\pi(j) = \lambda_2 v$, then

$$\pi(ah + bj)v = (a\pi(h) + b\pi(j))v = (a\lambda_1 + b\lambda_2)v.$$

So, we may make the following basis-independent notion of a weight.

Definition 3.9.5. Let $\mathfrak{h}$ be the subspace of $\mathfrak{sl}_3(\mathbb{C})$ spanned by h_1 and h_2 and let π be a finite-dimensional representation of $\mathfrak{sl}_3(\mathbb{C})$ acting on a vector space V. A linear functional $\mu \in \mathfrak{h}$ is called a *weight* for π if there exists a nonzero vector $v \in V$ such that $\pi(h)v = \mu(h)v$, for all $h \in \mathfrak{h}$. Such a vector v is called a *weight vector* with weight μ.

So, a weight is just a collection of simultaneous eigenvalues of all the elements h of $\mathfrak{h}$, which must depend linearly on h and, therefore, define a linear functional on $\mathfrak{h}$. Since h_1 and h_2 span $\mathfrak{h}$, the linear functional μ is determined by the value of $\mu(h_1)$ and $\mu(h_2)$, and thus our new notion of weight is equivalent to our old notion of a weight as just a pair of simultaneous eigenvalues of $\pi(h_1)$ and $\pi(h_2)$. The reason for adopting this basis-independent approach is that the action of the Weyl group does not preserve the basis $\{h_1, h_2\}$ for $\mathfrak{h}$. The Weyl group is (or may be thought of as) a group of linear transformations of $\mathfrak{h}$. This means that W acts linearly on $\mathfrak{h}$, and we denote this action as $w \cdot h$. We can define an associated action on the dual space $\mathfrak{h}^*$ in the following way: For $\mu \in \mathfrak{h}^*$ and $w \in W$, we define $w \cdot \mu$ to be the element of $\mathfrak{h}^*$ given by

$$(w \cdot \mu)(h) = \mu(w^{-1} \cdot h).$$

The main point of the Weyl group from the point of view of representation theory, namely that the weights of any representation are invariant under the action of the Weyl group. More explicitly, suppose that π is any finite-dimensional representation of $\mathfrak{sl}_3(\mathbb{C})$ and that $\mu \in \mathfrak{h}^*$ is a weight for π. Then, for any $w \in W$, $w \cdot \mu$ is also a weight of $\mathfrak{h}^*$, and the multiplicity of $w \cdot \mu$ is the same as the multiplicity of μ. In other words, since the roots are nothing but the nonzero weights of the adjoint representation, this result tells us that the roots are invariant under the action of the Weyl group. In order to visualize the action of the Weyl group, it is convenient to identify $\mathfrak{h}^*$ with $\mathfrak{h}$ by means of an inner product on $\mathfrak{h}$ that is invariant under the action of the Weyl group. Recall that $\mathfrak{h}$ is a subspace of the space of diagonal matrices, and we can use the Hilbert-Schmidt inner product $\langle A, B\rangle = \operatorname{tr}(A^*B)$. Since the Weyl group acts by permuting the diagonal entries, this inner product (restricted to the subspace $\mathfrak{h}$) is preserved by the action of W.

We now use this inner product on $\mathfrak{h}^*$ to identify $\mathfrak{h}$. Given any element α of $\mathfrak{h}$, the map $h \mapsto \langle \alpha, h\rangle$ is a linear functional on $\mathfrak{h}$ (*i. e.*, an element of $\mathfrak{h}^*$). Every linear functional on $\mathfrak{h}$ can be represented in this way for a unique α in $\mathfrak{h}$. Identifying each linear functional with the corresponding element of $\mathfrak{h}$, we will now regard a weight for (π, V) as a nonzero element of $\mathfrak{h}$ with the property that there exists a nonzero $v \in V$ such that

$$\pi(h)v = \langle \alpha, h\rangle v,$$

for all $h \in \mathfrak{h}$. This is the same as Definition 3.9.5 except that now, α lives in $\mathfrak{h}$ and we write $\langle \alpha, h\rangle$ instead of $\alpha(h)$ on the right. The roots, being weights for the adjoint representation, are viewed in a similar way. Now that the roots and weights live in $\mathfrak{h}$ instead of $\mathfrak{h}^*$, we can use the above inner product on $\mathfrak{h}$, and with this new point of view the roots α_1 and α_2 are identified with the following elements of $\mathfrak{h}$:

$$\alpha_1 = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 0 & & \\ & 1 & \\ & & -1 \end{pmatrix}.$$

To check this, we note that these matrices are indeed in $\mathfrak{h}$ since the diagonal entries sum to zero. Then, direct calculation shows that $\langle \alpha_1, h_1\rangle = 2$, $\langle \alpha_1, h_2\rangle = -1$, $\langle \alpha_2, h_1\rangle = -1$ and $\langle \alpha_2, h_2\rangle = 2$, in agreement with our earlier definition of α_1 and α_2 in (3.12). So, then, we can compute the lengths and angles as $||\alpha_1||^2 = \langle \alpha_1, \alpha_1\rangle = 2$, $||\alpha_2||^2 = \langle \alpha_2, \alpha_2\rangle = 2$, and $\langle \alpha_1, \alpha_2\rangle = -1$. This means that (with respect to this inner product) α_1 and α_2 both have length $\sqrt{2}$ and the angle θ between them satisfies $\cos\theta = -\frac{1}{2}$, so that $\theta = 120°$. We now consider the dominant integral elements, which are the possible highest weights of irreducible representations of $\mathfrak{sl}_3(\mathbb{C})$. With our new point of view, these are the elements μ of $\mathfrak{h}$ such that $\langle \mu, h_1 \rangle$ and $\langle \mu, h_2 \rangle$ are non-negative integers. We begin by considering the *fundamental weights* μ_1 and μ_2 defined by

$$\langle \mu_1, h_1 \rangle = 1, \qquad \langle \mu_1, h_2 \rangle = 0,$$
$$\langle \mu_2, h_1 \rangle = 0, \qquad \langle \mu_2, h_2 \rangle = 1.$$

These can be expressed in terms of α_1 and α_2 as follows:

$$\begin{aligned} \mu_1 &= \tfrac{2}{3}\alpha_1 + \tfrac{1}{3}\alpha_2, \\ \mu_2 &= \tfrac{1}{3}\alpha_1 + \tfrac{2}{3}\alpha_2, \end{aligned}$$

obtaining

$$\mu_1 = \begin{pmatrix} \frac{2}{3} & & \\ & \frac{-1}{3} & \\ & & \frac{-1}{3} \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} \frac{1}{3} & & \\ & \frac{1}{3} & \\ & & \frac{-2}{3} \end{pmatrix}.$$

A calculation then shows that μ_1 and μ_2 each have length $\frac{\sqrt{6}}{3}$ and that the angle between them is 60°. The set of dominant integral elements is then precisely the set of linear combinations of μ_1 and μ_2 with non-negative integer coefficients. Note that $\mu_1 + \mu_2 = \alpha_1 + \alpha_2$, an observation that helps in drawing Figure 3.3 below. Figure 3.3 shows the same information as Figure 3.2, namely, the roots and the dominant integral elements, but now drawn relative to a Weyl-invariant inner product. We draw only the two-dimensional real subspace of $\mathfrak{h}$ consisting of those elements μ such that $\langle \mu, h_1 \rangle$ and $\langle \mu, h_2 \rangle$ are real, since all the roots and weights have this property. In this figure, the arrows indicate the roots, the black dots indicate dominant integral elements (*i. e.*, points μ such that $\langle \mu, h_1 \rangle$ and $\langle \mu, h_2 \rangle$ are non-negative integers), and the triangular grid indicates integral elements (*i. e.*, points μ such that $\langle \mu, h_1 \rangle$ and $\langle \mu, h_2 \rangle$ are integers).

Let us see how the Weyl group acts on Figure 3.3. Let (1 2 3) denote the cyclic permutation of 1,2 and 3, and let $w_{(1\ 2\ 3)}$ denote the corresponding Weyl group element. Then, $w_{(1\ 2\ 3)}$ acts by cyclically permuting the diagonal entries of each element of h. Thus, $w_{(1\ 2\ 3)}$ takes α_1 to α_2 and takes α_2 to $-(\alpha_1 + \alpha_2)$. This action is a 120° rotation, counterclockwise in Figure 3.3. Next, let (1 2) be the permutation that interchanges 1 and 2 and let $w_{(1\ 2)}$ be the corresponding Weyl group element. Then, $w_{(1\ 2)}$ acts by interchanging the first two diagonal entries of each element of h and thus takes α_1 to $-\alpha_1$ and takes α_2 to $\alpha_1 + \alpha_2$. This corresponds to a reflection about the line perpendicular to α_1. The reader is invited to calculate the action of the remaining Weyl group elements, and observe that the Weyl group consists of six elements: the symmetry group of an equilateral triangle centered at the origin, as indicated in Figure 3.4:

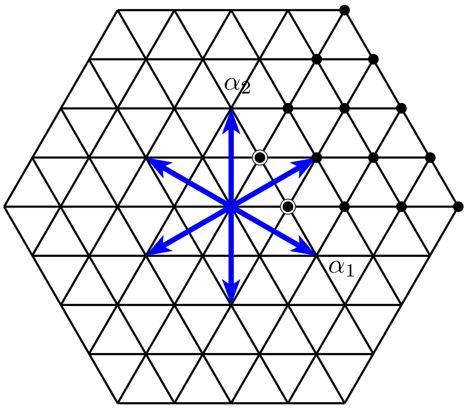

Figure 3.3: Roots and dominant integral elements for $\mathfrak{sl}_3(\mathbb{C})$ in the root basis.

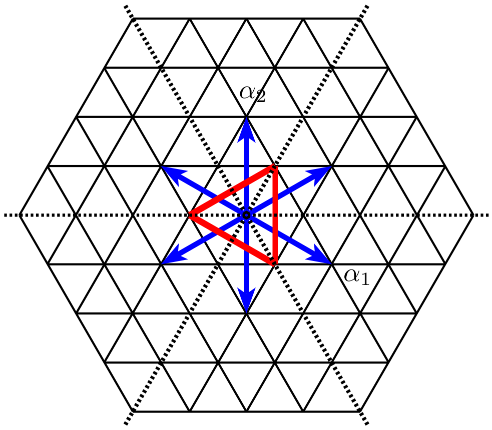

Figure 3.4: The Weyl group for $\mathfrak{sl}_3(\mathbb{C})$.

Chapter 4

Vertex algebras and Kac-Moody algebras

The notion of a vertex algebras was introduced by Borcherds in [8]. This is a rigorous mathematical definition of the chiral part of a 2-dimensional quantum field theory studied by physicist since the landmark paper of Belavin, Polyakov and Zamolodchikov [4]. Basically, vertex algebras are the rigorous formalization of the bosonic theory in mathematical physics. The main objective of this chapter is to give a quickly understanding of what vertex algebras and vertex operator algebras are. For a more detailed study, the reader may refer to [113], [67] and [7]. In Chapter 9 we shall discuss some other interesting topics —for our purposes— relating conformal field theory and the Moonshine phenomenon.

Subsequently, we will describe how vertex algebras arise on the mathematical scene, particularly in the developing of the representation theory of affine Kac-Moody algebras. In this part it will be useful to have in mind some of the results obtained in the previous chapter. The reader may refer to [112], [182] or [75] for a complete presentation of these topics.

4.1 Vertex operator algebras

Definition 4.1.1. Let $\mathbb{F}$ be a field. A *vertex algebra* over $\mathbb{F}$ is a vector space V over $\mathbb{F}$ with a collection of bilinear maps $V \times V \to V$

$$(u, v) \mapsto u_n v,$$

for all $n \in \mathbb{Z}$, and satisfying the following axioms
(i) $u_n v = 0$, for n sufficiently large, *i. e.*, there exists $n_0 \in \mathbb{N}$ (depending on u and v) such that $u_n v = 0$, for all $n \geq n_0$;
(ii) There exists an element $1 \in V$ such that[1]

$$\begin{aligned} 1_{-1}v = v,\ 1_n v = 0 \quad &\text{for all} \quad n \neq -1, \\ v_{-1}1 = v,\ v_n 1 = 0 \quad &\text{for all} \quad n \geq 0; \end{aligned}$$

(iii) (Borcherds' identity) for all $u, v, w \in V$ and $m, n, k \in \mathbb{Z}$

$$\sum_{i\geq 0} \binom{m}{i} (u_{n+i}v)_{m+k-i}w = \sum_{i\geq 0} (-1)^i \binom{n}{i} \big[u_{m+n-i}(v_{k+i}w) - (-1)^n v_{k+n-i}(u_{m+i}w)\big].$$

[1] This distinguished vector 1 usually appears as $|0\rangle$ in physicist notation. For example see [113].

We will work only with the case $\mathbb{F} = \mathbb{R}$. Relation (iii) is called Borcherds' identity. Since it is somewhat reminiscent of the Jacobi identity for Lie algebras (see Definition 2.3.1), it is called sometimes the Jacobi identity for vertex algebras.

Denote by $\operatorname{End} V[[z, z^{-1}]]$ the set of formal series (in the indeterminate z) of the form

$$\sum_{n\in\mathbb{Z}} \varphi_n z^{-n-1},$$

where $\varphi_n \in \operatorname{End} V$, for all $n \in \mathbb{Z}$. For each $u \in V$, we can define the *vertex operator* $Y(u, z) : V \to V$ by

$$Y(u, z) = \sum_{n\in\mathbb{Z}} u_n z^{-n-1}, \tag{4.1}$$

where $u_n \in \operatorname{End} V$ is given by $v \mapsto u_n v$. We denote by $V[[z]]$ the set of all formal series $\sum_{i\geq 0} v_i z^i$, for $v_i \in V$. Also, consider the formal expression

$$\delta(z - w) = z^{-1} \sum_{n\in\mathbb{Z}} \left(\frac{w}{z}\right)^n \in \mathbb{F}[[z, z^{-1}, w, w^{-1}]].$$

Then, the axioms for a vertex algebra can be written in terms of such vertex operators as follows:

(i) $Y(u, z)v$ has coefficient of z^n equal to 0 for all n sufficiently small (*i. e.*, there exists $n_0 \in \mathbb{N}$ depending on u and v such that $u_n v = 0$, $\forall n \geq n_0$);
(ii) There exists an element $1 \in V$ such that

- $Y(1, z) \equiv 1_V$ is the identity on $\operatorname{End} V$,
- $Y(v, z)1 \in V[[z]]$, for all $v \in V$,
- $\lim\limits_{z\to 0} Y(v, z)1 = v$.

(iii) $\delta(z_1 - z_2)Y(u, z_1)Y(v, z_2) - \delta(z_2 - z_1)Y(v, z_2)Y(u, z_1) = \delta(z_1 - z_0)Y\big(Y(u, z_0)v, z_2\big)$, for all $u, v \in V$.

For a series $a(z) = \sum_n a_n z^n \in V[[z, z^{-1}]]$, we will denote by

$$\partial a(z) = \sum_n n a_n z^{n-1}$$

the formal derivative of $a(z)$. Let $T : V \to V$ the linear map defined by

$$T(v) = v_{-2}1.$$

Denote by $[\cdot,\cdot]$ the usual bracket $[a,b]=ab-ba$ defined in $\operatorname{End}V[[z,z^{-1}]]$. Note that if $[T,\sum_n a_nz^n]=\sum_n[T,a_n]z^n$, for an element $\sum_n a_nz^n\in\operatorname{End}V[[z,z^{-1}]]$, then

$$[T,Y(u,z)]=\Big[T,\sum_n u_nz^{-n-1}\Big]=\sum_n[T,u_n]z^{-n-1}=\sum_n(Tu_n-u_nT)z^{-n-1}. \tag{4.2}$$

Also,

$$\partial Y(u,z)=\partial\sum_n u_nz^{-n-1}=\sum_n(-n-1)u_nz^{-n-2}=\sum_n(-n)u_{n-1}z^{-n-1}. \tag{4.3}$$

On the other hand, if we take $w=1$, and $m=0,k=-2$ on Borcherds' identity, we obtain

$$\begin{aligned}\sum_{i\geq 0}\binom{0}{i}(u_{n+i}v)_{-2-i}1 &= \sum_{i\geq 0}(-1)^i\binom{n}{i}\big[u_{n-i}(v_{-2+i}1)-(-1)^nv_{-2+n-i}(u_i1)\big]\\ &= \sum_{i\geq 0}(-1)^i\binom{n}{i}u_{n-i}(v_{-2+i}1).\end{aligned}$$

Note that the sum on the left-hand side above involves only one term (when $i=0$), and the right-hand side actually has only two terms (when $i=0,1$), since $v_n1=0$, for all $n\geq 0$. In fact, we have $-nu_{n-1}v=(u_nv)_{-2}1-u_n(v_{-2}1)$ and it follows that

$$-nu_{n-1}v=(u_nv)_{-2}1-u_n(v_{-2}1)=(Tu_n-u_nT)v,\quad \text{for all } n\in\mathbb{Z}. \tag{4.4}$$

From equation (4.2), (4.3) and (4.4), then we deduce that $-nu_{n-1}=[T,u_n]$, and hence $[T,Y(u,z)]=\partial Y(u,z)$. We summarize this and other similar results in the next statement. See [113, p.117–118] for a complete proof of this fact.

Theorem 4.1.2. *We have the following equivalent axioms for a vertex algebra:*
(i') (translation covariance) $[T,Y(u,v)]=\partial Y(u,z)$, for all $u\in V$;
(ii') (vacuum) $Y(1,z)=1_V$ and $Y(u,z)1|_{z=0}=u$, for all $u\in V$;
(iii') (locality) $(z-w)^nY(u,z)Y(v,w)=(z-w)^nY(v,w)Y(u,z)$ for n sufficiently large (depending on u and v).

Now, we apply the operator T repeatedly to the equation $Tv=v_{-2}1$. Since $Tv_n=[T,v_n]+v_nT$, we have that

$$T(v_n1)=(-n)v_{n-1}1+(v_nT)1=(-n)v_{n-1}1+v_n(1_{-2}1)=(-n)v_{n-1}1, \tag{4.5}$$

and so

$$\begin{aligned}T^2v &= T(v_{-2}1)=2v_{-3}1 \text{ so that } v_{-3}1=\tfrac{1}{2}T^2v,\\ T^3v &= T(2v_{-3}1)=(2\cdot 3)v_{-4}1 \text{ so that } v_{-4}1=\tfrac{1}{3!}T^3v.\end{aligned}$$

Continuing this process on identity (4.5), an induction on n guarantees that $v_{-n}1 = \frac{1}{(n-1)!}T^{n-1}v$, for all $n \geq 1$. Hence

$$\begin{aligned} Y(u,z)1 &= \sum_n (u_n 1)z^{-n-1} = \sum_{n\geq 0}(u_n 1)z^{-n-1} + \sum_{n<0}(u_n 1)z^{-n-1} \\ &= \sum_{n\geq 0}(u_{-(n+1)}1)z^n = \sum_{n\geq 0}\frac{1}{n!}T^n u\, z^n \\ &= \left(e^{zT}\right)u, \end{aligned}$$

for all $u \in V$.

Remark 4.1.3. The bracket operation defined by $[u,v] = u_0 v$ makes V/TV into a Lie algebra (see [12] for a proof of this). Borcherds' identity is described in [67] as being 'very concentrated'. It can be shown that it is equivalent to three simpler identities [113]. For $\theta, \varphi \in \operatorname{End} V$, we define $[\theta,\varphi] = \theta\varphi - \varphi\theta$. Also, if $a(z) = \sum_n a_n z^{-n-1} \in \operatorname{End} V[[z,z^{-1}]]$, we define

$$\begin{aligned} a(z)_+ &= \sum_{n<0} a_n z^{-n-1} = a_{-1} + a_{-2}z + a_{-3}z^2 + \dots \\ a(z)_- &= \sum_{n\geq 0} a_n z^{-n-1} = a_0 z^{-1} + a_1 z^{-2} + a_2 z^{-3} + \dots \end{aligned}$$

Definition 4.1.4. Given $a(z), b(z) \in \operatorname{End} V[[z,z^{-1}]]$, we also define their *normal ordered product* as

$$: a(z)b(z) : = a(z)_+ b(z) + b(z)a(z)_-.$$

In terms of the notation just introduced, Borcherds' identity can be shown to be equivalent to the following three simpler identities:

(a) $[u_m, Y(v,z)] = \sum_{i\geq 0}\binom{m}{i} Y(u_i v, z)z^{m-i}$, for all $u,v \in V$, $m \in \mathbb{Z}$;

(b) $: Y(u,z)Y(v,z) : = Y(u_{-1}v, z)$, for all $u, v \in V$;

(c) $Y(Tu,z) = \partial Y(u,z)$, for all $u \in V$.

Definition 4.1.5. An element ω of a vertex algebra V is a *conformal vector* of central charge c, if is an even vector satisfying:

- $\omega_0 v = Tv$, for all $v \in V$;
- $\omega_1 \omega = 2\omega$;

- $\omega_2\omega = 0$;
- $\omega_3\omega = \frac{c}{2}\mathbf{1}$;
- $\omega_i\omega = 0$, for all $i \geq 4$;
- $V = \bigoplus_{n\in\mathbb{Z}} V_n$, where each V_n is the set of eigenvectors of the linear operator ω_1

  $$V_n = \{v \in V : \omega_1 v = nv\}$$

  corresponding to the eigenvalue $n \in \mathbb{Z}$.

In other words, ω is a conformal vector if the corresponding vertex operator $Y(\omega, z)$ is a Virasoro field with central charge c, *i. e.*, a formal series $L(z)$ satisfying

$$L(z)L(w) = \frac{c/2}{(z-w)^4} + \frac{2L(w)}{(z-w)^2} + \frac{\partial L(w)}{z-w}.$$

In particular, for a conformal vector ω we have, $\mathbf{1} \in V_0$ and $\omega \in V_2$. Note that when a vertex algebra V has a conformal vector, it admits an action of the Lie algebra called Virasoro algebra (see next section).

Definition 4.1.6. A vertex algebra endowed with a conformal vector ω as in Definition 4.1.5 is called a *vertex operator algebra* (or a *conformal vertex algebra*) of rank c.

4.2 The Virasoro algebra

Let $p(t) \in \mathbb{F}[t, t^{-1}]$ and consider the derivation

$$T_{p(t)} = p(t)\partial \tag{4.6}$$

of $\mathbb{F}[t, t^{-1}]$. The linear space of all derivations of $\mathbb{F}[t, t^{-1}]$ of type (4.6) has the structure of a Lie algebra with respect to the natural Lie bracket

$$[T_{p(t)}, T_{q(t)}] = T_{p(t)q'(t)-p'(t)q(t)}$$

for $p(t), q(t) \in \mathbb{F}[t, t^{-1}]$. We denote this algebra by $\mathfrak{d}$ and we choose the following basis of $\mathfrak{d}$:

$$\{d_n = -t^{n+1}\partial = -t^n d : n \in \mathbb{Z}\},$$

where d is the derivation in (2.12). Then, the commutators have the form $[d_m, d_n] = (m-n)d_{m+n}$, for $m, n \in \mathbb{Z}$.

In fact, if $T \in \operatorname{End}\mathbb{F}[t, t^{-1}]$ is a derivation, and we set

$$p(t) = T(t), \tag{4.7}$$

then we have $T(1) = T(1 \cdot 1) = T(1) + T(1)$, so that

$$T(1) = 0, \tag{4.8}$$

and $0 = T(tt^{-1}) = T(t)t^{-1} + tT(t^{-1})$ implies that

$$T(t^{-1}) = -t^{-2}T(t). \tag{4.9}$$

Since formulas (4.7), (4.8) and (4.9) also hold for $T_{p(t)}$, we see that $T_{p(t)}$ and T agree on all powers of t. Thus, we have the following statement explaining the importance of $\mathfrak{d}$.

Proposition 4.2.1. *The derivations of* $\mathbb{F}[t, t^{-1}]$ *form precisely the Lie algebra* $\mathfrak{d}$.

Any three generators of $\mathfrak{d}$ of the form d_n, d_0, d_{-n}, $n \in \mathbb{Z}^+$ span a subalgebra of $\mathfrak{d}$ isomorphic to the Lie algebra $\mathfrak{sl}_2(\mathbb{F})$ of 2×2 matrices over $\mathbb{F}$ of trace 0. We shall single out the subalgebra

$$\mathfrak{p} = \mathbb{F}d_1 + \mathbb{F}d_0 + \mathbb{F}d_{-1}.$$

As in the case of affine Lie algebras, we consider central extensions. We denote by $\mathfrak{v}$ the following one-dimensional central extension of $\mathfrak{d}$ with basis consisting of a central element c and elements L_n, $n \in \mathbb{Z}$, corresponding to the basis elements d_n, of $\mathfrak{d}$. For $m, n \in \mathbb{Z}$

$$\begin{aligned} [L_m, L_n] &= (m-n)L_{m+n} + \tfrac{1}{12}(m^3 - m)\delta_{m+n,0}c & (4.10) \\ &= (m-n)L_{m+n} + \tfrac{1}{2}\binom{m+1}{3}\delta_{m+n,0}c; & (4.11) \end{aligned}$$

and $[L_m, c] = 0$, for all $m \in \mathbb{Z}$.

Definition 4.2.2. The Lie algebra $\mathfrak{d}$ above is called the *Virasoro algebra*.

Note that the central extension (4.10) is trivial when restricted to the subalgebra $\mathfrak{p}$ of $\mathfrak{d}$.

We can form equivalent extensions of $\mathfrak{d}$ by setting

$$L'_n = L_n + \beta_n c, \quad \text{for } \beta_n \in \mathbb{F},\ n \in \mathbb{Z}. \tag{4.12}$$

Then, the extension (4.10) is modified by the subtraction of the term $(m-n)\beta_{m+n}c$. Although, the significance of the extension (4.10) is given by the next result (see [67, p.33] for a proof).

Proposition 4.2.3. *The extension (4.10) of the Lie algebra* $\mathfrak{d}$ *is the unique nontrivial 1-dimensional extension up to isomorphism.*

Now we return to the situation in which V is a vertex operator algebra with conformal vector ω. Define the linear map $L_i : V \to V$ by

$$L_n = \omega_{n+1}, \quad \text{for } n \in \mathbb{Z}.$$

The properties of ω imply that

$$[L_m, L_n] = (m-n)L_{m+n} + \tfrac{1}{2}\binom{m+1}{3}\delta_{m+n,0}c\,1_V, \tag{4.13}$$

in analogy to (4.11) (see [113] for details). Thus, V is a module for the Virasoro algebra in which the central element c is represented by $c\,1_V$, where c is the central charge of ω.

4.3 Kac-Moody algebras and their representation

We shall now describe how vertex operators arose in the representation theory of affine Kac-Moody algebras. In order to explain this we first recall some ideas from the theory of Lie algebras (see Chapter 3).

We suppose first that $\mathfrak{g}$ is a finite dimensional simple Lie algebra over $\mathbb{C}$. Then, $\mathfrak{g}$ has a decomposition

$$\mathfrak{g} = \mathfrak{n}^+ \oplus \mathfrak{h} \oplus \mathfrak{n}^-, \tag{4.14}$$

where

$$\mathfrak{n}^+ = \sum_{\alpha\in\Phi^+} \mathfrak{g}_\alpha, \quad \mathfrak{n}^- = \sum_{\alpha\in\Phi^-} \mathfrak{g}_\alpha,$$

and all $\mathfrak{g}_\alpha$ satisfy

$$\dim \mathfrak{g}_\alpha = 1 \quad \text{and} \quad [\mathfrak{h}, \mathfrak{g}_\alpha] = \mathfrak{g}_\alpha.$$

Here, $\mathfrak{h}$ is a Cartan subalgebra of $\mathfrak{g}$, and the 1-dimensional spaces $\mathfrak{g}_\alpha$ are the root spaces of $\mathfrak{g}$ with respect to $\mathfrak{h}$, each of which gives rise to a 1-dimensional representation α of $\mathfrak{h}$ given by

$$[x, x_\alpha] = \alpha(x)\cdot x_\alpha, \text{ for all } x_\alpha \in \mathfrak{g}_\alpha, x \in \mathfrak{h}.$$

Recall that the set $\Phi = \Phi^+ \cup \Phi^-$ is the set of roots of $\mathfrak{g}$. Φ contains a subset $\Delta = \{\alpha_1, \ldots, \alpha_r\}$ of simple roots, where $r = \dim \mathfrak{h}$. The roots in Φ^+ are linear combinations of elements of Δ with non-negative integer coefficients and roots in Φ^- are linear combinations of elements of Δ with non-positive integer coefficients. The free abelian group

$$Q = \mathbb{Z}\alpha_1 + \mathbb{Z}\alpha_2 + \ldots + \mathbb{Z}\alpha_r$$

is called the root lattice. The real vector space $Q \otimes_{\mathbb{Z}} \mathbb{R}$ can be given the structure of a Euclidean space in a natural way. Let w_i be the reflection in the wall orthogonal to α_i. Then, we have seen that the group of isometries of $Q \otimes \mathbb{R}$ generated by $w_1, \ldots, w_r$ is the Weyl group W of $\mathfrak{g}$. Remember that this is a finite group which permutes the elements of Φ (Proposition 3.5.2). Each root is the image of some simple root under an element of W. We have

$$w_i(\alpha_j) = \alpha_j - A_{ij}\alpha_i, \text{ for } A_{ij} \in \mathbb{Z}. \tag{4.15}$$

Definition 4.3.1. The matrix $A = [A_{ij}]$ is called the *Cartan matrix* of $\mathfrak{g}$ (relatively to the Cartan subalgebra $\mathfrak{h}$).

Any Cartan matrix $[A_{ij}]$ satisfies the following conditions:

- $A_{ii} = 2$, for each i;
- $A_{ij} \in \{0, -1, -2, -3\}$ if $i \neq j$;
- $A_{ij} = 0$ if and only if $A_{ji} = 0$;
- $A_{ij} \in \{-2, -3\} \Rightarrow A_{ji} = -1$.

The Lie algebra $\mathfrak{g}$ can be defined by generators and relations depending only on the Cartan matrix A. The Cartan subalgebra $\mathfrak{h}$ has a basis $h_1, \ldots, h_r$ of elements satisfying

$$\alpha_j(h_i) = A_{ij}.$$

We also recall how to describe the finite dimensional irreducible $\mathfrak{g}$-modules. These are parametrized by dominant integral weights. Remember that a weight is an element of $\mathfrak{h}^*$. A weight ω is said dominant if $\omega(h_i)$ is a non-negative real number for all i. Finally, ω is dominant integral if each $\omega(h_i)$ is a non-negative integer. The fundamental weights $\omega_1, \ldots, \omega_r$ are defined by

$$\omega_i(h_j) = \delta_{ij}, \quad \text{for } i, j)1, 2, \ldots, r.$$

Thus, each dominant integral weight λ is a non-negative integral combination of the fundamental weights.

For each dominant integral weight λ there is a corresponding finite dimensional irreducible $\mathfrak{g}$-module M_λ. This module M_λ is a direct sum of 1-dimensional $\mathfrak{h}$-modules which, when collected together, give rise to a weight space decomposition

$$M_\lambda = \bigoplus_\mu M_\lambda^\mu,$$

where $M_\lambda^\mu = \{v \in M_\lambda : xv = \mu(x)v, \forall x \in \mathfrak{h}\}$ (here the sum is over all $\mu \in \mathfrak{h}^*$). In a similar form to Definition 3.8.1, we define a notion of character for this module M_λ

Definition 4.3.2. The *character* of the $\mathfrak{h}$-module M_λ is the function given by

$$\operatorname{char} M_\lambda = \sum_\mu (\dim M_\lambda^\mu) e^\mu.$$

Note that the character defined above is an element of the group ring $\mathbb{F}[G]$ of the free abelian group G generated by the fundamental weights. As usual, we write this group multiplicatively, replacing a weight μ by e^μ and such that

$$e^\mu e^\nu = e^{\mu+\nu}.$$

The Weyl character formula (Theorem 3.8.2) applied to this new definition of character asserts that

$$\left(\sum_{w\in W}\epsilon(w)w(e^{\rho})\right)\operatorname{char} M_{\lambda}=\sum_{w\in W}\epsilon(w)w(e^{\lambda+\rho}), \tag{4.16}$$

where $\epsilon = \det \in \{\pm 1\}$, is the homomorphism given by $\epsilon(w_i) = -1$, for all i, and $\rho = \sum_{i=1}^{r}\omega_i$. In an analogous way to Theorem 3.8.3, the expression $\sum_w \epsilon(w)w(e^{\rho})$ appearing in the denominator or $\operatorname{char} M_{\lambda}$ can be written in an alternative form. In this case, Weyl's denominator identity asserts that

$$\sum_{w\in W}\epsilon(w)w(e^{\rho}) = e^{\rho}\prod_{\alpha\in\Phi^{+}}(1-e^{-\alpha}). \tag{4.17}$$

The well known theory of finite dimensional simple Lie algebras over $\mathbb{C}$ which we have just outlined was generalized by Kac [110] and Moody [149] to give the theory of Kac-Moody algebras. In order to obtain such Lie algebras, we begin with the notion of generalized Cartan matrix.

Definition 4.3.3. A *generalized Cartan matrix* (sometimes abbreviated GCM) is any matrix $A = [A_{ij}]$ satisfying the conditions

- $A_{ij} \in \mathbb{Z}$;
- $A_{ii} = 2$, for all i;
- $A_{ij} \leq 0$, if $i \neq j$;
- $A_{ij} = 0$ if and only if $A_{ji} = 0$.

A Lie algebra is then defined by generators and relations depending on the generalized Cartan matrix A, just as in the case of finite dimensional simple Lie algebras. This Lie algebra is then extended by outer derivations to ensure that the simple roots are linearly independent, even though the matrix A is singular (may occur).

Definition 4.3.4. The resulting Lie algebra above is called the *Kac-Moody algebra* given by the generalized Cartan matrix A.

The main differences from the finite dimensional case are as follows:

1. The Lie algebra $\mathfrak{g}$ can have infinite dimension.
2. The root spaces $\mathfrak{g}_{\alpha}$ can have dimension greater than 1.
3. The Weyl group W can be infinite.
4. There can be both real roots and imaginary roots.

In this case, a root α is called *real* if $\langle \alpha, \alpha \rangle > 0$ and *imaginary* if $\langle \alpha, \alpha \rangle \leq 0$. All simple roots of a Kac-Moody algebra are real, and any real root can be obtained from a simple root just transforming by some appropriate element of the Weyl group.
A generalized Cartan matrix A is said *symmetrisable* if $A = DB$, where B is symmetric and D is a non-singular diagonal matrix. We shall restrict attention to Kac-Moody algebras with symmetrisable GCM. For each dominant integral weight λ, there is a corresponding irreducible module M_λ for such a Kac-Moody algebra. The character of this module M_λ is given by Kac's character formula

$$\left(\sum_{w \in W} \epsilon(w) w(e^\rho) \right) \operatorname{char} M_\lambda = \sum_{w \in W} \epsilon(w) w(e^{\lambda+\rho}), \tag{4.18}$$

Of course this look very much like Weyl's character formula (4.16) in the finite dimensional case, but here the sums over W are infinite. The denominator of Kac's character formula can be written in an alternative way, giving Kac's denominator identity (analogous to (4.17))

$$\sum_{w \in W} \epsilon(w) w(e^\rho) = e^\rho \prod_{\alpha \in \Phi^+} (1 - e^{-\alpha})^{\operatorname{mult} \alpha}. \tag{4.19}$$

The left-hand side is an infinite sum and the right-hand side is an infinite product. The multiplicity mult α is given by

$$\operatorname{mult} \alpha = \dim \mathfrak{g}_\alpha.$$

In the finite dimensional case, all the roots have multiplicity 1 and so equation (4.19) reduces to Weyl's denominator identity (4.17).

There is a remarkable trichotomy in the theory of Kac-Moody algebras. Let A be an $n \times n$ generalized Cartan matrix. Given a vector $u \in \mathbb{R}^r$ we write $u \succ 0$ if all coordinates u_i of u are positive; and $u \prec 0$ if all u_i are negative. It turns out that if the GCM A is indecomposable, exactly one of the following conditions holds:

- There exist $u \succ 0$ with $Au \succ 0$;
- There exist $u \succ 0$ with $Au = 0$;
- There exist $u \succ 0$ with $Au \prec 0$.

We say that A has *finite type* if the first condition holds; *affine type* if the second condition holds, and *indefinite type* if we have the third condition. The indecomposable Kac-Moody algebras of finite type correspond to the finite dimensional simple Lie algebras. All Kac-Moody algebras of finite or affine type are symmetrisable.

Examples of affine Kac-Moody algebras are the non-twisted affine algebras (Section 2.8). These may be constructed as follows. We start with a finite dimensional simple Lie algebra $\mathfrak{g}$ over $\mathbb{C}$. We then consider the algebra $\mathfrak{g} \otimes \mathbb{C}[t, t^{-1}]$. This algebra has a $\mathbb{C}$-basis

$$h_i \otimes t^m, \; x_\alpha \otimes t^m, \quad \text{for } i = 1, \dots, r, \; m \in \mathbb{Z},$$

where x_α is a non-zero vector or $\mathfrak{g}_\alpha$. Also, this algebra has a non-trivial 1-dimensional central extension. This is the Lie algebra

$$\mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}k,$$

in which multiplication is given by

$$[x \otimes t^m + \lambda k, y \otimes t^n + \lambda' k] = [x, y] \otimes t^{m+n} + \langle x, y\rangle \, m\delta_{m+n,0}\, k,$$

for $x, y \in \mathfrak{g}$. Here $\langle \cdot, \cdot \rangle$ is a non-degenerate invariant symmetric bilinear form on $\mathfrak{g}$ (recall equation (2.14)). We have already called this Lie algebra $\hat{\mathfrak{g}}$ the untwisted affine algebra associated with $\mathfrak{g}$. If we extend this Lie algebra by adjoining a 1-dimensional space of outer derivations, thus we obtain the Lie algebra

$$\tilde{\mathfrak{g}} = \hat{\mathfrak{g}} \rtimes \mathbb{C}d = \mathfrak{g} \otimes_{\mathbb{C}} \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}k \oplus \mathbb{C}d,$$

with multiplication

$$[x \otimes t^m + \lambda k + \mu d, y \otimes t^n + \lambda' k + \mu' d] = [x, y] \otimes t^{m+n} - x \otimes \mu' m t^m + y \otimes \mu n t^n + \langle x, y\rangle \, m\delta_{m+n,0}\, k.$$

We have called this the extended affine algebra associated with $\mathfrak{g}$ (Section 2.8). This Lie algebra $\tilde{\mathfrak{g}}$ is an affine Kac-Moody algebra whose generalized Cartan matrix has degree $n+1$. Its GCM is obtained from the Cartan matrix of $\mathfrak{g}$ by adjoining an additional row and column. Let α the highest root of the Lie algebra $\mathfrak{g}$ and consider $\alpha_0 = -\alpha$. The Cartan matrix of $\mathfrak{g}$ is the $n \times n$ matrix $A = [A_{ij}]$, where

$$A_{ij} = 2\frac{\langle \alpha_i, \alpha_j\rangle}{\langle \alpha_i, \alpha_i\rangle}, \quad \text{for} \quad i, j = 1, 2, \ldots, r;$$

and the generalized Cartan matrix of the affine Lie algebra $\tilde{\mathfrak{g}}$ is the $(n+1) \times (n+1)$ matrix

$$A = [A_{ij}], \quad \text{for} \quad i, j = 0, 1, 2, \ldots, r,$$

given by the same formula. Thus, the GCM of $\tilde{\mathfrak{g}}$ gives rise to a corresponding irreducible $\tilde{\mathfrak{g}}$-module M_λ, as described above. In particular, we obtain such modules associated with the fundamental weights $\omega_0, \omega_1, \ldots, \omega_r$ of $\tilde{\mathfrak{g}}$. The fundamental representation of $\tilde{\mathfrak{g}}$ associated to the weight ω_0 is called the *basic representation* of $\tilde{\mathfrak{g}}$. The basic representation plays a crucial role in the many applications of the representation theory of affine Kac-Moody algebras. We shall describe a module giving the basic representation of $\tilde{\mathfrak{g}}$ in the case when $\mathfrak{g}$ is simply laced, *i. e.*, when $A_{ij} = 0$ or -1 for all $i, j \in \{1, 2, \ldots, r\}$ with $i \neq j$.

Suppose this is so and let Q be the root lattice of $\mathfrak{g}$. There is a symmetric $\mathbb{Z}$-bilinear form $Q \times Q \to \mathbb{Z}$ given by $(\alpha, \beta) \mapsto \langle \alpha, \beta\rangle$ uniquely determined by the conditions

$$\langle \alpha_i, \alpha_j\rangle = A_{ij}, \quad \text{for} \quad \alpha_i, \alpha_j \in \Delta.$$

This lattice Q has a unique central extension $\hat{Q}$ with

$$1 \longrightarrow C_2 \longrightarrow \hat{Q} \longrightarrow Q \longrightarrow 1$$

in which

$$aba^{-1}b^{-1} = (-1)^{\langle \bar{a}, \bar{b} \rangle}, \quad \text{for} \;\; a, b \in \hat{Q},$$

where $a \mapsto \bar{a}$ is the map $\hat{Q} \to Q$. For each $\gamma \in Q$, we choose an element $e_\gamma \in \hat{Q}$ such that $\overline{e_\gamma} = \gamma$ and $e_0 = 1$. Then we have

$$e_\alpha e_\beta = e_{\alpha+\beta} \epsilon(\alpha, \beta), \;\; \text{for all} \;\; \alpha, \beta \in Q.$$

The map $\epsilon : Q \times Q \to \{\pm 1\}$ is a 2-cocycle. It is known from the theory of finite dimensional Lie algebras that the root vectors x_α of $\mathfrak{g}$ can be chosen so that

$$[x_\alpha, x_\beta] = \epsilon(\alpha, \beta) x_{\alpha+\beta},$$

whenever $\alpha, \beta, \alpha + \beta \in \Phi$.

We are now in position to describe a $\tilde{\mathfrak{g}}$-module giving the basic representation. Let $\tilde{\mathfrak{h}}^-$ the subalgebra of $\tilde{\mathfrak{g}}$ given by

$$\tilde{\mathfrak{h}}^- = \bigoplus_{n<0} (\mathfrak{h} \otimes t^n),$$

and let $S(\tilde{\mathfrak{h}}^-)$ be the symmetric algebra on $\tilde{\mathfrak{h}}^-$ (recall Section 2.7). Let $\mathbb{C}[Q]$ the group algebra of Q with basis e^γ for $\gamma \in Q$. We write

$$V_Q = S(\tilde{\mathfrak{h}}^-) \otimes_{\mathbb{C}} \mathbb{C}[Q]. \tag{4.20}$$

It turns out that V_Q can be made into a $\tilde{\mathfrak{g}}$-module affording the basic representation. The central element K of $\tilde{\mathfrak{g}}$ acts as the identity map on V_Q. The elements $h \otimes t^n$ for $h \in \mathfrak{h}$ act as follows:

$$h \otimes t^n \;\; \text{acts as} \;\; \begin{cases} h(n) \otimes 1 & \text{if } n \neq 0, \\ 1 \otimes h(0) & \text{if } n = 0. \end{cases}$$

Here, $h(0) \in \operatorname{End} \mathbb{C}[Q]$ is given by $h(0)e^\gamma = \gamma(h)e^\gamma$, and $h(n) \in \operatorname{End} S(\tilde{\mathfrak{h}}^-)$ is given by the following rule

- If $n < 0$, then $h(n)$ is multiplication by $h \otimes t^n$.
- If $n > 0$, then $h(n)$ is the derivation of $S(\tilde{\mathfrak{h}}^-)$ determined by

$$\begin{aligned} h(n)(x \otimes t^{-n}) &= n(x, h), \;\; \text{for } x \in \mathfrak{h}, \\ h(x)(x \otimes t^{-m}) &= 0, \;\; \text{if } m \neq n. \end{aligned}$$

We now consider the action of the elements $x_\alpha \otimes t^n \in \tilde{\mathfrak{g}}$ on V_Q. Let $x_\alpha(n) \in \operatorname{End} V_Q$ be the endomorphism induced by $x_\alpha \in t^n$. The endomorphisms $x_\alpha(n)$ turn out to be complicated expressions not appealing to the intuition. However, it is possible to make sense of them by combining then into a vertex operator

$$Y(\alpha, z) = \sum_{n \in \mathbb{Z}} x_\alpha(n) z^{-n-1},$$

which is given by an explicit formula. In order to explain this, we introduce the following notation.

There is an isomorphism $\mathfrak{h}^* \to \mathfrak{h}$ given by $\lambda \mapsto h_\lambda$, where $\mu(h_\lambda) = \langle \lambda, \mu \rangle$. In this way $\mathfrak{h}^*$ may be identified with elements of $\mathfrak{h}$ (the corresponding co-roots). Thus, the elements $\alpha(n) \in \operatorname{End} S(\tilde{\mathfrak{h}}^-)$ are defined as above for $n \neq 0$. The vertex operator $Y(\alpha, z)$ is then given by

$$Y(\alpha, z) = \exp\left(\sum_{n<0} -\frac{\alpha(n)}{n} z^{-n} \right) \exp\left(\sum_{n>0} -\frac{\alpha(n)}{n} z^{-n} \right) e_\alpha z^\alpha.$$

Here, $z^\alpha \in \operatorname{End} V_Q$ is given by $1 \otimes z^\alpha$ for $z^\alpha \in \operatorname{End} \mathbb{C}[Q]$; and $z^\alpha e^\gamma = z^{\langle \alpha, \gamma \rangle} e^\gamma$. Also, $e_\alpha \in \operatorname{End} V_Q$ is given by $1 \otimes e_\alpha$ for $e_\alpha \in \operatorname{End} \mathbb{C}[Q]$; and $e_\alpha e^\gamma = \epsilon(\alpha, \gamma) e^{\alpha+\gamma}$ (see for example [112]).

We conclude this section with the definition of automorphism of a vertex algebra. We will see in Chapter 5 that the Monster group $\mathbb{M}$ occur as a group of automorphisms of a vertex algebra.

Definition 4.3.5. Let V be a vertex algebra with conformal vector ω. An *automorphism* of V is an invertible linear map $g : V \to V$ satisfying

$$\begin{aligned} gY(v,z)g^{-1} &= Y(gv, z), \quad \text{for all } v \in V; \\ g(\omega) &= \omega. \end{aligned}$$

This then, is how vertex operators first appeared in the representation theory of affine Kac-Moody algebras. In fact, as we shall see, the vector space V_Q can be made into a vertex algebra.

4.4 Lattices

Definition 4.4.1. By a *lattice* of rank $n \in \mathbb{N}$ we shall mean a rank n free abelian group L provided with a rational-valued symmetric $\mathbb{Z}$-bilinear form

$$\langle \cdot, \cdot \rangle : L \times L \to \mathbb{Q}$$

A lattice isomorphism is sometimes called an *isometry*. A lattice L is said *non-degenerate* if its form $\langle\cdot,\cdot\rangle$ is non-degenerate in the sense that for $\alpha \in L$

$$\langle\alpha, L\rangle \text{ implies } \alpha = 0.$$

Given a lattice L, we see by choosing a $\mathbb{Z}$-base of L that $\langle L, L\rangle \subseteq \frac{1}{r}\mathbb{Z}$ for some $r \in \mathbb{Z}^+$. That is

$$\langle\cdot,\cdot\rangle : L \times L \to \tfrac{1}{r}\mathbb{Z}.$$

We canonically embed L in the $\mathbb{Q}$-vector space

$$L_{\mathbb{Q}} = L \otimes_{\mathbb{Z}} \mathbb{Q}, \tag{4.21}$$

which is n-dimensional since a $\mathbb{Z}$-basis of L is a $\mathbb{Q}$-basis of $L_{\mathbb{Q}}$, and we extend $\langle\cdot,\cdot\rangle$ to a symmetric $\mathbb{Q}$-bilinear form

$$\langle\cdot,\cdot\rangle : L_{\mathbb{Q}} \times L_{\mathbb{Q}} \to \mathbb{Q}. \tag{4.22}$$

Note that every element of $L_{\mathbb{Q}}$ is of the form α/N for some $\alpha \in L$ and $N \in \mathbb{Z}^{\times}$. The lattice L is non-degenerate if and only if the form (4.22) is non-degenerate, and this amounts to the condition

$$\det[\langle\alpha_i, \alpha_j\rangle]_{i,j} \neq 0, \tag{4.23}$$

for a $\mathbb{Z}$-base $\{\alpha_1, \ldots, \alpha_n\}$ of L. A lattice may be equivalently defined as the $\mathbb{Z}$-span of a basis of a finite-dimensional rational vector space equipped with a symmetric bilinear form.

Let L be a lattice. For $m \in \mathbb{Q}$, we set

$$L_m = \{\alpha \in L : \langle\alpha, \alpha\rangle = m\}. \tag{4.24}$$

The lattice L is said to be *even* if $\langle\alpha, \alpha\rangle \in 2\mathbb{Z}$, for all $\alpha \in L$. It is said *integral* if $\langle\alpha, \beta\rangle \in \mathbb{Z}$, for all $\alpha, \beta \in L$; and is said *positive definite* if $\langle\alpha, \alpha\rangle > 0$, for all $\alpha \in L - \{0\}$, or equivalently, for $\alpha \in L_{\mathbb{Q}} - \{0\}$. The polarization formula

$$\langle\alpha, \beta\rangle = \tfrac{1}{2}\big(\langle\alpha + \beta, \alpha + \beta\rangle - \langle\alpha, \alpha\rangle - \langle\beta, \beta\rangle\big), \tag{4.25}$$

shows that an even lattice is integral. The *dual* of L is the set

$$L^{\circ} = \{\alpha \in L_{\mathbb{Q}} : \langle\alpha, L\rangle \subseteq \mathbb{Z}\}. \tag{4.26}$$

This set is again a lattice if and only if L is non-degenerate, and in this case, L° has as a base the dual base $\{\alpha_1^*, \ldots, \alpha_n^*\}$ of a given base $\{\alpha_1, \ldots, \alpha_n\}$ of L, defined by

$$\langle\alpha_i^*, \alpha_j\rangle = \delta_{ij} \text{ for } i, j = 1, \ldots, n.$$

Note that L is integral if and only if $L \subseteq L^{\circ}$. The lattice L is said to be *self-dual* if $L = L^{\circ}$. This is equivalent to L being integral and *unimodular*, which means that

$$\big|\det[\langle\alpha_i, \alpha_j\rangle]_{i,j}\big| = 1.$$

In fact, if L is integral and non-degenerate, the $[\langle\alpha_i,\alpha_j\rangle]_{i,j}$ is the matrix of the embedding map $L\to L^\circ$ with respect to the given base and its dual base, and the unimodularity amounts to the condition that this embedding be an isomorphism of abelian groups.

Generalizing (4.21) and (4.22), we embed L in the $\mathbb{E}$-vector space

$$L_{\mathbb{E}} = L\otimes_{\mathbb{Z}}\mathbb{E},$$

for any field of characteristic zero, and we extend $\langle\cdot,\cdot\rangle$ to the symmetric $\mathbb{E}$-bilinear form

$$\langle\cdot,\cdot\rangle : L_{\mathbb{E}}\times L_{\mathbb{E}}\to\mathbb{E}.$$

Then, L is positive definite if and only if the real vector space $L_{\mathbb{R}}$ is an Euclidean space. In this case,

$$|L_m|<\infty,\quad\text{for } m\in\mathbb{Q}, \tag{4.27}$$

since L_m is the intersection of the discrete set L with a compact set (a sphere) in $L_{\mathbb{R}}$.

Using the Schwarz inequality, we observe that if the lattice L is integral and positive definite, and if $\alpha,\beta\in L_2$, then

$$\langle\alpha,\beta\rangle\in\{0,\pm1,\pm2\}$$

and

$$\begin{aligned}\langle\alpha,\beta\rangle &= -2 \text{ if and only if } \alpha+\beta=0,\\ \langle\alpha,\beta\rangle &= -1 \text{ if and only if } \alpha+\beta\in L_2,\\ \langle\alpha,\beta\rangle &\geq 0 \text{ if and only if } \alpha+\beta\notin L_2\cup\{0\}.\end{aligned}$$

Let L be an even lattice. Set

$$\breve{L}=L/2L,$$

and view the elementary abelian 2-group $\breve{L}$ as a vector space over the field $\mathbb{F}_2$. Denote by $L\to\breve{L}$ the canonical map $\alpha\mapsto\breve{\alpha}=\alpha+2L$. Since a $\mathbb{Z}$-base of L reduces to an $\mathbb{F}_2$-basis of $\breve{L}$, then

$$\dim\breve{L}=\operatorname{rank}L.$$

There is a canonical $\mathbb{Z}$-bilinear form $c_0 : L\times L\to\mathbb{Z}/2\mathbb{Z}$ given by

$$(\alpha,\beta)\mapsto\langle\alpha,\beta\rangle+2\mathbb{Z}$$

on L, and c_0 is alternating because L is even. the form c_0 induces a (well-defined) alternating $\mathbb{F}_2$-bilinear map $c_1 : \breve{L}\times\breve{L}\to\mathbb{Z}/2\mathbb{Z}$ given by

$$(\breve{\alpha},\breve{\beta})\mapsto\langle\alpha,\beta\rangle+2\mathbb{Z},$$

for $\alpha,\beta\in L$. There is a canonical quadratic form q_1 on $\breve{L}$ with associated bilinear form c_1:

$$q_1 : \breve{L}\to\mathbb{Z}/2\mathbb{Z},\quad\text{with } \breve{\alpha}\mapsto\tfrac{1}{2}\langle\alpha,\alpha\rangle+2\mathbb{Z}, \tag{4.28}$$

for $\alpha \in L$. this form is well-defined: if $\breve{\alpha} = \breve{\beta}$, then $\beta - \alpha = \gamma \in 2L$, and

$$\tfrac{1}{2}\langle \beta, \beta \rangle = \tfrac{1}{2}\langle \alpha, \alpha \rangle + \langle \alpha, \gamma \rangle + \tfrac{1}{2}\langle \gamma, \gamma \rangle \equiv \tfrac{1}{2}\langle \alpha, \alpha \rangle \pmod 2.$$

It is clear that c_1 is the associated form. Denote by $q_0 : L \to \mathbb{Z}/2\mathbb{Z}$, where

$$\alpha \mapsto \tfrac{1}{2}\langle \alpha, \alpha \rangle + 2\mathbb{Z},$$

the pullback of q_1 to L. There exist $\mathbb{Z}$-bilinear forms $\epsilon_0 : L \times L \to \mathbb{Z}/2\mathbb{Z}$ such that

$$\epsilon_0(\alpha, \alpha) = q_0(\alpha) = \tfrac{1}{2}\langle \alpha, \alpha \rangle + 2\mathbb{Z}, \text{ for} \alpha \in L,$$

and consequently

$$\epsilon_0(\alpha, \beta) - \epsilon_0(\beta, \alpha) = c_0(\alpha, \beta) = \langle \alpha, \beta \rangle + 2\mathbb{Z},$$

for $\alpha \in L$. Note that q_1 (or equivalently, c_1) is nonsingular if, and only if, the determinant (4.23) is odd —in particular, if L is unimodular—. These considerations are important in the construction of central extensions.

Let L be a positive definite lattice. The *theta function* θ_L associated to L is defined to be the formal series in the variable $q = e^{2\pi i z}$ (see Section 6.7) given by

$$\theta_L(q) = \sum_{\alpha \in L} q^{\langle \alpha, \alpha \rangle / 2} = \sum_{m \in \mathbb{Q}} |L_m| q^{m/2}, \tag{4.29}$$

(recall (4.24) and (4.27)). If L is even and unimodular, the theta function θ_L has important modular transformation properties under the modular group $SL_2(\mathbb{Z})$, when z is a complex variable in the upper half plane $\mathbb{H}$ (see Chapter 6).

4.5 The vertex algebra of an even lattice

Let L be an even lattice, *i. e.*, a free abelian group of finite rank with a symmetric nondegenerate bilinear form $L \times L \to \mathbb{Z}$

$$(\alpha, \beta) \mapsto \langle \alpha, \beta \rangle$$

such that $\langle \alpha, \alpha \rangle \in 2\mathbb{Z}$ for all $\alpha \in L$.

Example 4.5.1. The root lattice Q of a simply laced simple Lie algebra considered in Section 4.3 is even.

Let $\mathfrak{h} = L \otimes_{\mathbb{Z}} \mathbb{C}$ and consider

$$\tilde{\mathfrak{h}}^- = \bigoplus_{n<0} (\mathfrak{h} \otimes t^n).$$

Let $\mathbb{C}[L]$ be the group algebra of L, with basis e^{γ} for $\gamma \in L$. Define V_L by equation (4.20)

$$V_L = S(\tilde{\mathfrak{h}^-}) \otimes_{\mathbb{C}} \mathbb{C}[L].$$

then, it is possible to make V_L into a vertex algebra with the property that for all $\alpha \in L$ we have

$$Y(1 \otimes e^{\alpha}, z) = \exp\left(\sum_{n<0} -\frac{\alpha(n)}{n} z^{-n}\right) \exp\left(\sum_{n>0} -\frac{\alpha(n)}{n} z^{-n}\right) e_{\alpha} z^{\alpha},$$

where the elements $\alpha(n), e_{\alpha}$ and z^{α} of $\operatorname{End} V_L$ are defined as in Section 4.3 (*c.f.* [67] or [113]).

More generally, for

$$v = (h_1 \otimes t^{-n_1}) \cdots (h_k \otimes t^{-n_k}) \otimes e^{\alpha} \in V_L$$

we have

$$Y(v, z) = :\frac{1}{(n_1 - 1)!}(\frac{d}{dz})^{n_1 - 1} h_1(z) \cdots \frac{1}{(n_k - 1)!}(\frac{d}{dz})^{n_k - 1} h_k(z)\, Y(1 \otimes e^{\alpha}, z):$$

where the normal ordered product of more than two factors is defined inductively from the right in terms of Definition 4.1.5, that is, for example

$$: a_1(z) a_2(z) a_3(z) : = : a_1(z)\big(: a_2(z) a_3(z) : \big) : \ .$$

Thus, for $h \in \mathfrak{h}$ and $m > 0$ we have

$$Y\big((h \otimes t^{-m}) \otimes 1, z\big) = \frac{1}{(m-1)!}\Big(\frac{d}{dz}\Big)^{m-1} h(z),$$

where $h(z) = \sum_{n \in \mathbb{Z}} h(n) z^{-n-1}$ and $h(n) \in \operatorname{End} V_L$ is defined as in Section 4.3. The element $1 \in \operatorname{End} V_L$ is $1 \otimes 1$. Also, V_L has a conformal vector ω given by

$$\omega = \left(\frac{1}{2} \sum_{i=1}^{r} (h_i' \otimes t^{-1})(h_i \otimes t^{-1})\right) \otimes 1 \in V_L,$$

where $h_1, \ldots, h_r$ is a basis for $\mathfrak{h}$, and $h_1', \ldots, h_r'$ is the dual basis with respect to $\langle \cdot, \cdot \rangle$. The element ω is independent of this choice of basis (*c.f.* [67]).

The maps $L_i : V_L \to V_L$ defined by $L_i = \omega_{i+1}$ satisfy

$$[L_i, L_j] = (i - j) L_{i+j} + \frac{1}{2}\binom{i+1}{3} \dim \mathfrak{h}\, \delta_{i+j,0}\, 1_{V_L}.$$

Thus, V_L is a module for the Virasoro algebra under which the central element c acts as multiplication by $\dim \mathfrak{h} = \operatorname{rank} L$ (see [67]). Since the subspace of the Virasoro algebra spanned by L_{-1}, L_0 and L_1 is a 3-dimensional subalgebra isomorphic to $\mathfrak{sl}_2(\mathbb{C})$, we may regard V_L as a $\mathfrak{sl}_2(\mathbb{C})$-module. Then, elements of the 1-dimensional subspace spanned by $1 \in V_L$ are annihilated by $\mathfrak{sl}_2(\mathbb{C})$ and are, in fact, the only elements of V_L with this property.

4.6 Dynkin Diagrams of classic Lie algebras

In this section, we discuss informally (without proof) the classification, up to equivalence, of root systems. This leads to a classification, up to equivalence, of semisimple Lie algebras. The classification of root systems is given in terms of an object called the Dynkin diagram.

Suppose $A = \{\alpha_1, \ldots, \alpha_r\}$ is a base for a root system R. Then, the *Dynkin diagram* for R (relative to the base A) is a graph having vertices $v_1, \ldots, v_r$. Between any two vertices, we place either no edge, one edge, two edges, or three edges as follows. Consider distinct indices i and j. If the corresponding roots α_i and α_j are orthogonal, then we put no edge between v_i and v_j. In the cases where α_i and α_j are not orthogonal, we put

- one edge between v_i and v_j if α_i and α_j have the same length,
- two edges if the longer of α_i and α_j is $\sqrt{2}$ longer than the shorter, and
- three edges if the longer of α_i and α_j is $\sqrt{3}$ longer than the shorter.

In addition, if α_i and α_j are not orthogonal and not of the same length, then we decorate the edges between v_i and v_j with an arrow pointing from the vertex associated to the longer root toward the vertex associated to the shorter root (thinking of the arrow as a 'greater than' sign). These are all the possible values, since we have [91, p.246–249]

Proposition 4.6.1. *Suppose that α, β are roots (of a root system R), α is not multiple of β, and $\langle \alpha, \alpha \rangle \geq \langle \beta, \beta \rangle$. Then, one of the following holds:*

1. $\langle \alpha, \beta \rangle = 0$*;*
2. $\langle \alpha, \alpha \rangle = \langle \beta, \beta \rangle$*, and the angle between α and β is 60^o or 120^o;*
3. $\langle \alpha, \alpha \rangle = 2\langle \beta, \beta \rangle$*, and the angle between α and β is 45^o or 135^o;*
4. $\langle \alpha, \alpha \rangle = 3\langle \beta, \beta \rangle$*, and the angle between α and β is 30^o or 150^o.*

Proposition 4.6.2. *If α and β are distinct elements of a base Δ for R, then $\langle \alpha, \beta \rangle \leq 0$.*

Thus, Proposition 4.6.1 tells us that if α_i and α_j are not orthogonal, then the only possible length ratios are 1, $\sqrt{2}$ and $\sqrt{3}$. Furthermore, Proposition 4.6.2 says that these three cases correspond to angles of 120°, 135°, and 150°, respectively.

Two Dynkin diagrams are said to be *equivalent* if there is a one-to-one, onto map of the vertices of one to the vertices of the other that preserves the number of bonds and the direction of the arrows. Recall from Section 3.5 that any two bases for the same root system can be mapped into one another by the action of the Weyl group. This implies that the equivalence class of the Dynkin diagram is independent of the choice of base. As we will see, only graphs of certain very special forms arises as Dynkin diagram of a root system. The following characterizes root systems via their Dynkin diagram [91, p.270]

Theorem 4.6.3. *A root system is irreducible if and only if its Dynkin diagram is connected. Two root systems with equivalent Dynkin diagrams are equivalent. If R^* is the dual root system to R, then the Dynkin diagram of R^* is the same as that of R (with direction of each arrow reversed).*

So, the classification of irreducible root systems amounts to classifying all the connected diagrams that can arise as Dynkin diagrams of root systems. It is well known the classification of the Dynkin diagrams for the classical Lie algebras, $\mathfrak{sl}_n(\mathbb{C})$, $\mathfrak{so}_n(\mathbb{C})$, and $\mathfrak{sp}_n(\mathbb{C})$ (see Figure 4.1):

- A_n: The root system A_n is the root system of $\mathfrak{sl}_{n+1}(\mathbb{C})$, which has rank n.
- B_n: The root system B_n is the root system of $\mathfrak{so}_{2n+1}(\mathbb{C})$, which has rank n.
- C_n: The root system C_n is the root system of $\mathfrak{sp}_n(\mathbb{C})$, which has rank n.
- D_n: The root system D_n is the root system of $\mathfrak{so}_{2n}(\mathbb{C})$, which has rank n.

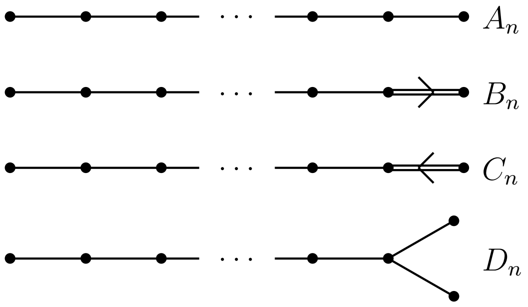

Figure 4.1: Dynkin diagrams for A_n, B_n, C_n and D_n.

Certain special things happen in low rank. In rank one, there is only one possible Dynkin diagram, reflecting that there is only one isomorphism class of complex semisimple Lie algebras in rank one. The Lie algebra $\mathfrak{so}_2(\mathbb{C})$ is not semisimple and the remaining three, $\mathfrak{sl}_2(\mathbb{C})$, $\mathfrak{so}_3(\mathbb{C})$, and $\mathfrak{sp}_1(\mathbb{C})$, are isomorphic. In rank two, the Dynkin diagram D_2 is disconnected, reflecting that $\mathfrak{so}_4(\mathbb{C}) \cong \mathfrak{sl}_2(\mathbb{C}) \oplus \mathfrak{sl}_2(\mathbb{C})$. Also, the Dynkin diagrams B_2 and C_2 are isomorphic, reflecting that $\mathfrak{so}_5(\mathbb{C}) \cong \mathfrak{sp}_2(\mathbb{C})$. In rank three, the Dynkin diagrams A_3 and D_3 are isomorphic, thus $\mathfrak{sl}_4(\mathbb{C}) \cong \mathfrak{so}_6(\mathbb{C})$. We may observe certain things about the short and long roots in root systems where more than one length of root occurs. The long roots in B_n form a root system by themselves, namely D_n. The short roots in B_n form a root system by themselves, namely $A_1 \times \cdots \times A_1$. In C_n, it is the reverse: the long roots form $A_1 \times \cdots \times A_1$ and the short roots form D_n.

In addition to the root systems associated to the classical Lie algebras, there are five 'exceptional' irreducible root systems, denoted G_2, F_4, E_6, E_7, and E_8, whose Dynkin diagrams are shown in Figure 4.2. For the construction of these exceptional root systems, see Humphreys [100]. Thus we have the following classification theorem for irreducible root systems [91, p.272]:

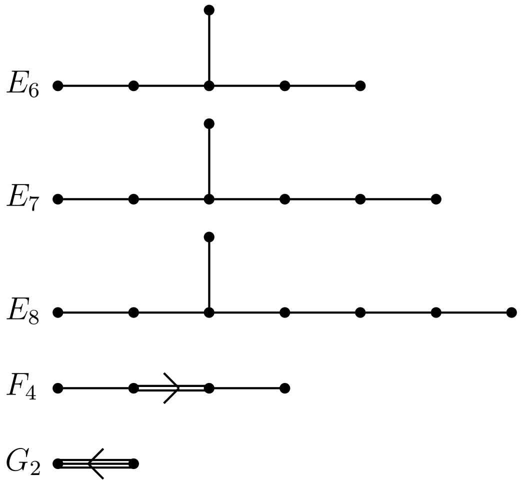

Figure 4.2: The exceptional Dynkin diagrams E_6, E_7, E_8, F_4 and G_2.

Theorem 4.6.4 (Classification of root systems). *Every irreducible root system is isomorphic to precisely one root system from the following:*

1. *The classical root systems A_n, $n \geq 1$;*
2. *The classical root systems B_n, $n \geq 2$;*
3. *The classical root systems C_n, $n \geq 3$;*
4. *The classical root systems D_n, $n \geq 4$;*
5. *The exceptional root systems G_2, F_4, E_6, E_7, and E_8.*

In the language of semisimple Lie theory (recall Chapter 3), suppose that $\mathfrak{h}$ is a Cartan subalgebra of a reductive semisimple Lie algebra $\mathfrak{g}$. Suppose also that Δ is the root system of $\mathfrak{g}$ with respect to $\mathfrak{h}$. We know that Δ is identified with the dual $\mathfrak{h}^*$ by means of the bilinear form $\langle \cdot, \cdot \rangle$, and the x_α (for $\alpha \in \Delta$) are the corresponding root vectors. Together with a basis of the root system Δ, the x_α's form a *Chevalley* basis of $\mathfrak{g}$. The sublattice

$$Q = \mathbb{Z}\Delta = \left\{ \sum n_i \alpha_i : n_i \in \mathbb{Z}, \ \alpha_i \in \Delta \right\} \tag{4.30}$$

of some lattice L, generated by Δ is the root lattice of $\mathfrak{g}$, and its dual

$$Q^\circ = \{\alpha \in \mathfrak{h} : \langle \alpha, Q \rangle \subseteq \mathbb{Z}\} \tag{4.31}$$

is the so called weight lattice. We now list examples of positive definite even lattices L such that $\delta = L_2$ spans $\mathfrak{h}$ and is indecomposable. Thus, the corresponding Lie algebra $\mathfrak{g}$ is simple. In each case, L is generated by Δ, that is, $L = Q = \mathbb{Z}\Delta$. The notations A_n, D_n and E_n correspond to the standard designations of the simple Lie algebras given above. In each case, $n = \operatorname{rank} Q = \dim \mathfrak{h}$, and recall that $\dim \mathfrak{g} = n + |\Delta|$.

For $\ell \geq 1$, denote by V_ℓ an ℓ-dimensional rational vector space equipped with a positive definite symmetric form $\langle \cdot, \cdot \rangle$ and an orthonormal basis $\{v_1, \ldots, v_\ell\}$.

Type A_n, $n \geq 1$: In V_{n+1} take

$$Q_{A_n} = \left\{ \sum_{i=1}^{n+1} m_i v_i : m_i \in \mathbb{Z},\ \sum m_i = 0 \right\}.$$

Then, $\Delta = \{\pm(v_i - v_j) : 1 \leq i < j \leq n+1\}$, $|\Delta| = n(n+1)$ and $\dim \mathfrak{g} = (n+1)^2 - 1$. As we have mentioned above, the case A_1 is the case $\mathfrak{g} = \mathfrak{sl}_2(\mathbb{C})$.

Type D_n, $n \geq 3$: In V_n take

$$Q_{D_n} = \left\{ \sum_{i=1}^{n} m_i v_i : m_i \in \mathbb{Z},\ \sum m_i \in 2\mathbb{Z} \right\}.$$

Then, $\Delta = \{\pm v_i \pm v_j : 1 \leq i < j \leq n\}$, $|\Delta| = 2n(n-1)$ and $\dim \mathfrak{g} = n(2n-1)$. As we have mentioned above, the case D_3 is the same as the case A_3. The same construction for $n = 2$ gives $A_1 \times A_1$.

Type E_8: In V_8 take

$$\begin{aligned} Q_{E_8} &= Q_{D_8} + \tfrac{1}{2}\mathbb{Z} \sum_{i=1}^{8} v_i \\ &= \left\{ \sum_{i=1}^{n} m_i v_i : \text{ either } m_1, \ldots, m_8 \in \mathbb{Z}, \text{ or } m_1, \ldots, m_8 \in \mathbb{Z} + \tfrac{1}{2}; \sum m_i \in 2\mathbb{Z} \right\}. \end{aligned}$$

Then, $\Delta = \{\pm v_i \pm v_j : 1 \leq i < j \leq 8\} \cup \left\{ \sum_{i=1}^{8} m_i v_i : m_i = \pm\frac{1}{2},\ \sum m_i \in 2\mathbb{Z} \right\}$, $|\Delta| = 240$ and $\dim \mathfrak{g} = 248$.

Type E_7: In V_8 take

$$Q_{E_7} = Q_{E_8} \cap \left\{ \sum_{i=1}^{8} m_i v_i : m_i \in \tfrac{1}{2}\mathbb{Z},\ \sum m_i \in 2\mathbb{Z} \right\}.$$

Then, $\Delta = \{\pm v_i \pm v_j : 1 \leq i < j \leq 6\} \cup \{\pm(v_7 - v_8)\} \cup \big\{ \sum_{i=1}^{6} m_i v_i : m_i = \pm(v_7 - v_8),\ m_i = \pm\frac{1}{2},\ \sum m_i \in 2\mathbb{Z}\big\}$, $|\Delta| = 126$ and $\dim \mathfrak{g} = 133$.

Type E_6: In V_8 take

$$Q_{E_6} = Q_{E_7} \cap \left\{ \sum_{i=1}^{8} m_i v_i :\ m_i \in \tfrac{1}{2}\mathbb{Z},\ m_6 = m_7 \right\}.$$

Then, $\Delta = \{\pm v_i \pm v_j : 1 \leq i < j \leq 5\} \cup \big\{ \pm \big(\sum_{i=1}^{5} m_i v_i + \frac{1}{2}(v_6 + v_7 - v_8)\big) : m_i = \pm\frac{1}{2},\ \sum m_i \in 2\mathbb{Z} - \frac{1}{2}\big\}$, $|\Delta| = 72$ and $\dim \mathfrak{g} = 78$.

4.7 Lie theory and Moonshine

McKay not only noticed (1.1), but also observed that

$$j(z)^{1/3} = q^{-1/3}(1 + 248q + 4,124q^2 + 34,752q^3 + \ldots). \tag{4.32}$$

The point is that 248 is the dimension of the defining representation of the E_8 simple Lie algebra, while $4,124 = 3,875 + 248 + 1$ and $34,752 = 30,380 + 3,875 + 2 \cdot 248 + 1$. Incidentally, $j^{1/3}$ is the Hauptmodul for the genus 0 congruence group $\Gamma(3)$. Thus Moonshine is related somehow to Lie theory.

McKay later found independent relationships with Lie theory [142], [17], [74], reminiscent of his famous A-D-E correspondence with finite subgroups of $SU_2(\mathbb{C})$. As mentioned in Chapter 1, $\mathbb{M}$ has two conjugacy classes of involutions. Let K be the smaller one, called '2A' in [36] (the alternative, class '2B', has almost 100 million times more elements). The product of any two elements of K will lie in one of nine conjugacy classes: namely, 1A, 2A, 2B, 3A, 3C, 4A, 4B, 5A, 6A, corresponding respectively to elements of orders 1, 2, 2, 3, 3, 4, 4, 5, 6. It is surprising that, for such a complicated group as $\mathbb{M}$, that list stops at only 6 —we call $\mathbb{M}$ a 6-transposition group for this reason—. The punchline: McKay noticed that those nine numbers are precisely the labels of the affine $\hat{E}_8$ Dynkin diagram (see Figure 4.3). Thus we can attach a conjugacy class of $\mathbb{M}$ to each vertex of the $\hat{E}_8$ diagram. An interpretation of the edges in the $\hat{E}_8$ diagram, in terms of $\mathbb{M}$, is unfortunately not known. We can't get the affine $\hat{E}_7$ labels in a similar way, but McKay noticed that an order two folding of affine $\hat{E}_7$ gives the affine $\hat{F}_4$ diagram, and we can obtain its labels using the Baby Monster $\mathbb{B}$ (the second largest sporadic). In particular, let K now be the smallest conjugacy class of involutions in $\mathbb{B}$ (also labeled '2A' in [36]); the conjugacy classes in K have orders 1, 2, 2, 3, 4 ($\mathbb{B}$ is a 4-transposition group), and these are the labels of the $\hat{F}_4$ affine Dynkin diagram. Of course we prefer $\hat{E}_7$ to $\hat{F}_4$, but perhaps that two-folding has something to do with the fact that an order-two central extension of $\mathbb{B}$ is the centraliser of an element $g \in \mathbb{M}$ of order two.

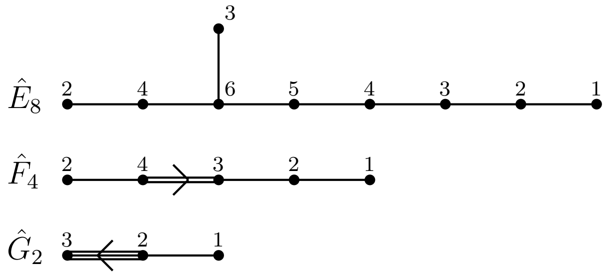

Figure 4.3: The affine Dynkin diagrams $\hat{E}_8$, $\hat{F}_4$ and $\hat{G}_2$ with labels.

Now, the triple-folding of the affine Lie algebra $\hat{E}_6$ is the affine $\hat{G}_2$. The Monster has three conjugacy classes of order three. The smallest of these ('3A' in Atlas notation) has a centraliser which is a triple cover of the Fischer group Fi'$_{24}$. Taking the smallest conjugacy class of involutions in Fi'$_{24}$, and multiplying it by itself, gives conjugacy classes with orders 1, 2, 3 (hence Fi'$_{24}$ is a 3-transposition group) and those not surprisingly are the labels of $\hat{G}_2$. Although we now understand (4.32) (see Chapter 9) and have proved the basic Conway-Norton conjecture (see Chapter 8), McKay's observation about $\hat{E}_8$, $\hat{F}_4$ and $\hat{G}_2$ diagrams still have no explanation. In [74] these patterns are extended, by relating various simple groups to the $\hat{E}_8$ diagram with deleted nodes.

Shortly after McKay's E_8 observation, Kac [111] and James Lepowsky [132] independently remarked that the unique level-1 highest-weight representation $L(\omega_0)$ of the affine Kac-Moody algebra $E_8^{(1)}$ has graded dimension $j(z)^{1/3}$. Since each homogeneous piece of any representation $L(\lambda)$ of the affine Kac-Moody algebra $X_\ell^{(1)}$ (in Kac's notation) must carry a representation of the associated finite-dimensional Lie group $X_\ell(\mathbb{C})$, and the graded dimensions (multiplied by an appropriate power of q) of an affine algebra are modular functions for some $G \subseteq SL_2(\mathbb{Z})$, this explained McKay's E_8 observation. His observation (1.1) took longer to clarify because so much of the mathematics needed was still to be developed.

Chapter 5

The Moonshine module

Following [67], we shall construct in this chapter a module with an associated vertex operator algebra $V^\natural$ on which the Monster group $\mathbb{M}$ acts as group of automorphisms. We shall call this the *Monster vertex algebra* or the *Moonshine module*. We cannot describe this vertex algebra in detail here —indeed a lengthy book is needed to do this—. However, we shall mention some of its most basic properties.
The vertex operator algebra associated with $V^\natural$ crowns the sequence of exceptional structures starting with the Golay error-correcting code and continuing with the Leech lattice. The corresponding sequence of their automorphisms consist of the Mathieu sporadic group M_{24}, the Conway sporadic group Co_0 and the Monster $\mathbb{M}$.

We begin this chapter with introductions to the two exceptional structures mentioned above. For more details, we refer to the papers [130]-[131], [29]-[33], to the books [86], and to the extensive collection [39]. The history of the discoveries related to to Monstrous Moonshine is reviewed in Chapter 1.

5.1 The Golay code

Let Ω a finite set with n elements. The power set $\wp(\Omega) = \{S : S \subseteq \Omega\}$ can be viewed as an $\mathbb{F}_2$-vector space under the operation $+$ of symmetric difference. By a *(binary linear) code* we shall understand an $\mathbb{F}_2$-subspace of $\wp(\Omega)$. An isomorphism of codes is defined in the obvious way. The cardinality $|C|$ of an element C of a code is called the *weight* of C. A code $\mathscr{C}$ is said to be of *type I* if

$$n \in 2\mathbb{Z}, |C| \in 2\mathbb{Z} \ \text{ for all } C \in \mathscr{C} \text{ and } \Omega \in \mathscr{C},$$

and $\mathscr{C}$ is said to be of *type II* if

$$n \in 4\mathbb{Z}, |C| \in 4\mathbb{Z} \ \text{ for all } C \in \mathscr{C} \text{ and } \Omega \in \mathscr{C}.$$

The type II codes will be seen as analogues of even lattices, and will help us construct them.

For a code $\mathscr{C}$, the *dual code* $\mathscr{C}^\circ$ is given by

$$\mathscr{C}^\circ = \{S \subseteq \Omega : |S \cap C| \in 2\mathbb{Z}, \text{ for all } C \in \mathscr{C}\}.$$

Thus, $\mathscr{C}^\circ$ is the annihilator of $\mathscr{C}$ in $\wp(\Omega)$ with respect to the natural nonsingular symmetric bilinear form

$$(S_1, S_2) \mapsto |S_1 \cap S_2| + 2\mathbb{Z} \tag{5.1}$$

on $\wp(\Omega)$. Hence

$$\dim_{\mathbb{F}_2} \mathscr{C}^\circ = n - \dim_{\mathbb{F}_2} \mathscr{C}.$$

We call $\mathscr{C}$ *self-dual* if $\mathscr{C}^\circ = \mathscr{C}$, in which case n is even and $\dim_{\mathbb{F}_2} \mathscr{C} = \frac{n}{2}$.
Consider the subspace

$$\mathcal{E}(\Omega) = \{S \subseteq \Omega : |S| \in 2\mathbb{Z}\}.$$

The map $q : \mathcal{E}(\Omega) \to \mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2$ given by $S \mapsto \frac{|S|}{2} + 2\mathbb{Z}$ is a quadratic form on $\mathcal{E}(\Omega)$, with associated bilinear form given by (5.1). In case $n \in 2\mathbb{Z}$, $\mathbb{F}_2\Omega$ is the radical form of q. A subspace of a space with a quadratic form is called *totally singular* if the form vanishes on it.

Remark 5.1.1. In case $n \in 4\mathbb{Z}$, the self-dual codes of type II correspond to the (maximal) totally singular subspaces of $\mathcal{E}(\Omega)/\mathbb{F}_2\Omega$ of dimension $\frac{n}{2} - 1$. Equivalently, the type II self-dual codes are the (maximal) totally singular subspaces of $\mathcal{E}(\Omega)$ of dimension $\frac{n}{2}$.
For a code $\mathscr{C}$, set

$$w(\mathscr{C}) = \sum_{C \in \mathscr{C}} q^{|C|} \in \mathbb{Z}[q]$$

the *weight distribution* of $\mathscr{C}$. A code with the following properties is called a *Hamming code* (see [67, p.300]):

Theorem 5.1.2. *There is a self-dual code of type II on an 8-element set Ω.*

It can be proved that the weight distribution of the constructed Hamming code is $1+14q^4+q^8$. The Hamming code is unique up to isomorphism (see, *e. g.* [136]).

Definition 5.1.3. A code with the following properties is called a *(binary) Golay code* [67, p.301]:

Theorem 5.1.4. *There is a self-dual code $\mathscr{C}$ of type II on a 24-element set, such that $\mathscr{C}$ has no elements of weight 4.*

It is not hard to see that the weight distribution of the constructed Golay code is $1 + 759q^8 + 2,576q^{12} + 759q^{16} + q^{24}$ (see [46]). Note that $759 = \binom{24}{5}/\binom{8}{5}$. The Golay code is unique up to isomorphism (see, *e. g.* [136]). The 759 elements of the Golay code of weight 8 are called *octads*. We mention some properties of the Golay code [67, p.302].

Proposition 5.1.5. *Let Ω be a 24-element set and let $\mathscr{C}$ be a Golay code in $\mathcal{E}(\Omega)$.*
(a) Every 5-element subset of Ω is included in a unique octad in $\mathscr{C}$.
(b) Let T_0 be a 4-element subset of Ω. Then T_0 lies in exactly 5 octads. These are of the form $T_0 \cup T_i$, for $i = 1, \dots, 5$, where $\Omega = \cup_{i=0}^5 T_i$ is a disjoint union of 4-element sets (called a sextet *in [32]). The union of any pair of T_i is an octad.*

From this, it follows directly that the octads generate the Golay code. The group of automorphisms of the Golay code $\mathscr{C}$ is called the *Mathieu group* M_{24}:

$$M_{24} = \operatorname{Aut} \mathscr{C}.$$

This is a nonabelian simple group (see Table 1.1). In fact, in the natural representation of M_{24} on $\wp(\Omega)$, the complete list of submodules is

$$0 \subseteq \langle \Omega \rangle \subseteq \mathscr{C} \subseteq \mathcal{E}(\Omega) \subseteq \wp(\Omega).$$

In particular, $\mathscr{C}/\langle\Omega\rangle$ and $\mathcal{E}(\Omega)/\mathscr{C}$ are faithful irreducible modules for M_{24}.

5.2 The Leech lattice

A self-dual code $\mathscr{C}$ of type II based on a set Ω gives rise to an even unimodular lattice as follows. Let

$$\mathfrak{h} = \bigoplus_{k \in \Omega} \mathbb{F}\alpha_k.$$

be a vector space with basis $\{\alpha_k : k \in \Omega\}$ and provide $\mathfrak{h}$ the symmetric bilinear form $\langle \cdot, \cdot \rangle$ such that

$$\langle \alpha_k, \alpha_\ell \rangle = \delta_{k\ell}, \quad \text{for } k, \ell \in \Omega.$$

For $S \subseteq \Omega$, set $\alpha_S = \sum_{k \in S} \alpha_k$. Define

$$Q = \bigoplus_{k \in \Omega} \mathbb{Z}\alpha_k,$$

and for a code $\mathscr{C}$ based on Ω define the positive definite lattice

$$L_0 = \left\{ \sum_{k \in \Omega} m_k \alpha_k : m_k \in \tfrac{1}{2}\mathbb{Z}, \{k | m_k \in \mathbb{Z} + \tfrac{1}{2}\} \in \mathscr{C} \right\}. \tag{5.2}$$

(there should be no confusion with notation in (4.24)). Then L_0 is even if and only if $|C| \in 4\mathbb{Z}$, for all $C \in \mathscr{C}$. Moreover, the dual lattice L_0° of L_0 (recall (4.26)) is the corresponding lattice based on the dual code $\mathscr{C}^\circ$:

$$L_0^\circ = \left\{ \sum_{k \in \Omega} m_k \alpha_k : m_k \in \tfrac{1}{2}\mathbb{Z}, \{k | m_k \in \mathbb{Z} + \tfrac{1}{2}\} \in \mathscr{C}^\circ \right\}.$$

Thus we have

Proposition 5.2.1. *A code $\mathscr{C}$ is self-dual of type II if, and only if, the corresponding lattice L_0 is even self-dual, or equivalently, even unimodular.*

Now consider the following modification of the lattice L_0 associated with a code $\mathscr{C}$, which we now assume contains Ω:

$$\begin{aligned} L_0' &= \sum_{C\in\mathscr{C}} \mathbb{Z}\tfrac{1}{2}\alpha_C + \sum_{k\in\Omega} \mathbb{Z}(\tfrac{1}{4}\alpha_\Omega - \alpha_k) \\ &= \sum_{C\in\mathscr{C}} \mathbb{Z}\tfrac{1}{2}\alpha_C + \sum_{k,\ell\in\Omega} \mathbb{Z}(\alpha_k - \alpha_\ell) + \mathbb{Z}(\tfrac{1}{4}\alpha_\Omega - \alpha_{k_0}), \end{aligned} \tag{5.3}$$

where k_0 is a fixed element of Ω. Since the lattice

$$L_0 \cap L_0' = \sum_{C\in\mathscr{C}} \mathbb{Z}\tfrac{1}{2}\alpha_C + \sum_{k,\ell\in\Omega} \mathbb{Z}(\alpha_k - \alpha_\ell) \tag{5.4}$$

has index 2 in both L_0 and L_0', then L_0' is unimodular if and only if L_0 is. A necessary and sufficient condition for L_0' to be even is that

$$n = |\Omega| \in 8(2\mathbb{Z}+1),$$

since

$$\langle \tfrac{1}{4}\alpha_\Omega - \alpha_k, \tfrac{1}{4}\alpha_\Omega - \alpha_k \rangle = \tfrac{n}{8} + 1, \tag{5.5}$$

for $k \in \Omega$, and we find

Proposition 5.2.2. *If $n \in 8(2\mathbb{Z}+1)$ and the code $\mathscr{C}$ is self-dual of type II, then the corresponding lattice L_0' is even unimodular.*

Definition 5.2.3. The *Leech lattice* is the even unimodular lattice

$$\Lambda = L_0'$$

for the case $n = 24$ and $\mathscr{C}$ the Golay code (recall Theorem 5.1.4).

The lattice Λ has no 'short' elements, *i. e.*, $\Lambda_2 = \emptyset$ (using notation (4.24)). This can be checked from equation (5.3) and the corresponding property of the Golay code, but we shall prove $\Lambda_2 = \emptyset$ and reconstruct Λ instead by using another principle, which will be an analogue for lattices of Remark 5.1.1 for codes.
Let L be an even unimodular lattice of rank n and with form $\langle\cdot,\cdot\rangle$. For our special purpose, now we provide L with the following rescaled form:

$$\langle \alpha, \beta \rangle_{1/2} = \tfrac{1}{2}\langle \alpha, \beta \rangle \quad \text{for } \alpha, \beta \in L \tag{5.6}$$

and by abuse of notation we drop the subscript 1/2. With respect to the new form $\langle\cdot,\cdot\rangle$, L has the following properties:

- $\langle \alpha, \alpha \rangle \in \mathbb{Z}$, for all $\alpha \in L$.
- $\big|\det[\langle \alpha_i, \alpha_j \rangle]_{i,j}\big| = \frac{1}{2^n}$, for a base $\{\alpha_1, \ldots, \alpha_n\}$ of L.

As we did in Section 4.4 (but keeping in mind the rescaled norm), set

$$\breve{L} = L/2L,$$

an n-dimensional vector space over $\mathbb{F}_2$, and write $\alpha \mapsto \breve{\alpha}$ for the canonical map. Since the original lattice L is even, the map $q_1 : \breve{L} \to \mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2$ given by

$$\breve{\alpha} \mapsto \langle \alpha, \alpha \rangle + 2\mathbb{Z}$$

defines a quadratic form on $\breve{L}$ with associated bilinear form $c_1 : \breve{L} \times \breve{L} \to \mathbb{F}_2$ determined by

$$(\breve{\alpha}, \breve{\beta}) \mapsto 2\langle \alpha, \beta \rangle + 2\mathbb{Z}.$$

These forms are nonsingular since the original lattice L is unimodular. From these definitions we have

Proposition 5.2.4. *Let M be a lattice such that $2L \subseteq M \subseteq L$. Then, M is even unimodular with respect to the new form $\langle \cdot, \cdot \rangle$ if, and only if, $\breve{M} = M/2L$ is a (maximal) totally singular subspace of $\breve{L}$ of dimension $\frac{n}{2}$.*

We shall apply this last principle to the direct sum of three copies of the root lattice Q_{E_8}, (recall Section 4.6) which is an even unimodular lattice of rank 8. For brevity, set

$$\Gamma = Q_{E_8}$$

and provide Γ with the rescaled form (5.6) as above. Using the Hamming code (Theorem 5.1.2), we shall first show that $\breve{\Gamma}$ contains complementary 4-dimensional totally singular subspaces.

We can describe the lattice Γ as follows:

$$\begin{aligned} \Gamma &= \sum_{k,\ell \in \Omega} \mathbb{Z}(\tfrac{1}{2}\alpha_k \pm \tfrac{1}{2}\alpha_\ell) + \mathbb{Z}\tfrac{1}{4}\alpha_\Omega \\ &= \left\{ \sum_{k \in \Omega} m_k \alpha_k : \text{ either } m_1, \ldots, m_8 \in \tfrac{1}{2}\mathbb{Z}, \text{ or } m_1, \ldots, m_8 \in \tfrac{1}{2}\mathbb{Z} + \tfrac{1}{4}; \sum m_k \in \mathbb{Z} \right\}. \end{aligned}$$

Now we identify Ω with the projective line over the 7-element field: $\Omega = \mathbb{P}^1(\mathbb{F}_7) = \mathbb{F}_7 \cup \{\infty\}$. Consider the sets of squares and non-squares

$$\begin{aligned} \mathcal{Q} &= \{x^2 : x \in \mathbb{F}_7\} = \{0, 1, 2, 4\}, \\ \mathcal{N} &= \Omega - \mathcal{Q} = \{3, 5, 6, \infty\}; \end{aligned}$$

and define subspaces

$$\begin{aligned} \mathscr{C}_1 &= \langle \mathcal{N} + i | i \in \mathbb{F}_7 \rangle, \\ \mathscr{C}_2 &= \langle -\mathcal{N} - i | i \in \mathbb{F}_7 \rangle, \end{aligned}$$

of $\mathscr{E}(\Omega)$. Then it follows that $\mathscr{C}_1$ and $\mathscr{C}_2$ satisfy the properties of Theorem 5.1.2. Now, consider the lattices

$$\begin{aligned} \Phi &= L_0 \text{ for } \mathscr{C} = \mathscr{C}_1, \\ \Psi &= L_0' \text{ for } \mathscr{C} = \mathscr{C}_2. \end{aligned}$$

Then, Φ and Ψ are even unimodular lattices by Propositions (5.2.1) and (5.2.2). Moreover, $2\Gamma \subseteq \Phi, \Psi \subseteq \Gamma$, and since $\mathscr{C}_1 + \mathscr{C}_2 = \mathscr{E}(\Omega)$, we see that

$$\Phi + \Psi = \Gamma.$$

Proposition (5.2.4) thus gives

Proposition 5.2.5. *We have a decomposition*

$$\breve{\Gamma} = \breve{\Phi} \oplus \breve{\Psi} \tag{5.7}$$

into 4-dimensional totally singular subspaces. In particular, $\Phi + \Psi = \Gamma$ *and* $\Phi \cap \Psi = 2\Gamma$.

Using the decomposition (5.7) we shall now reconstruct the Leech lattice by analogy with the construction of the Golay code in Theorem 5.1.4. Set

$$\Gamma^3 = \Gamma \oplus \Gamma \oplus \Gamma$$

the orthogonal direct sum of three copies of Γ, equipped with the modified form (5.6). In Γ^3, set

$$\Lambda = \{(\phi, \phi, 0) : \phi \in \Phi\} \oplus \{(\phi, 0, \phi) : \phi \in \Phi\} \oplus \{(\psi, \psi, \psi) : \psi \in \Psi\}. \tag{5.8}$$

Then we have [67, p.306]

Theorem 5.2.6. *The lattice Λ in (5.8) is an even unimodular lattice such that $\Lambda_2 = \emptyset$. Moreover, it coincides with the Leech lattice in Definition 5.2.3.*

The Leech lattice is the unique positive definite even unimodular lattice Λ of rank 24 with $\Lambda_2 = \emptyset$, up to isometry (see [153], [30]), and the E_8-root lattice is the unique positive definite even unimodular lattice of rank 8 up to isometry. Thus the lattices Γ, Φ and Ψ with its original bilinear form, are all isometric. In fact, the construction (5.8) expresses the Leech lattice as the non-orthogonal direct sum or three rescaled copies of the E_8-root lattice. Sometimes the Leech lattice Λ is denoted as Λ_{24}.

The group of isometries of the Leech lattice is called the *Conway group* Co_0 (or $\cdot_0$)

$$\begin{aligned} \mathrm{Co}_0 &= \mathrm{Aut}(\Lambda, \langle \cdot, \cdot \rangle) \\ &= \{g \in \mathrm{Aut}\,\Lambda : \langle g\alpha, g\beta \rangle = \langle \alpha, \beta \rangle \text{ for all } \alpha, \beta \in \Lambda\}. \end{aligned}$$

Its quotient by the central subgroup $\langle \pm 1 \rangle$ is called the *Conway group* Co_1

$$\mathrm{Co}_1 = \mathrm{Co}_0 / \langle \pm 1 \rangle.$$

We cite some basic facts about this groups [31, 32]:

- Co_0 equals its commutator subgroup: $\mathrm{Co}_0 = [\mathrm{Co}_0, \mathrm{Co}_0]$;
- $\mathrm{Cent}\,\mathrm{Co}_0 = \langle \pm 1 \rangle$;
- Co_1 is a nonabelian simple group (see Table 1.1);
- Co_1 acts faithfully and irreducibly on $\lambda/2\Lambda$.

Only as a comment, we have constructed the Leech lattice as the lattice L_0' based on the Golay code. The lattice L_0 based on the Golay code is also even unimodular, but $(L_0)_2 \neq \emptyset$. In fact, $(L_0)_2 = \{\pm\alpha_i : i \in \Omega\}$. This lattice L_0 is called the *Niemeier lattice of type* A_1^{24} and is sometimes written as $L_0 = N(A_1^{24})$. We have encountered a third even unimodular lattice of rank 24, namely, Γ^3. Altogether, there are 24 even unimodular lattices of rank 24, up to isometry, called the *Niemeier lattices* (see [153]).

Next, we shall describe and count the shortest nonzero elements of the Leech lattice. In the notation of the beginning of this section, for $S \subseteq \Omega$ let ϵ_S be the involution of $\mathfrak{h}$ given by

$$\epsilon_S : \alpha_k \mapsto \begin{cases} -\alpha_k & \text{if } k \in S \\ \alpha_k & \text{if } k \notin S \end{cases},$$

for $k \in \Omega$. It follows from relations (5.2), (5.3), (5.4) and (5.5) that Λ_4 is composed of three types of elements:

$$\Lambda_4 = \Lambda_4^1 \,\dot\cup\, \Lambda_4^2 \,\dot\cup\, \Lambda_4^3,$$

where

$$\begin{aligned} \Lambda_4^1 &= \{\tfrac{1}{2}\epsilon_S\alpha_C : C \in \mathscr{C}, |C| = 8, S \subseteq C, |S| \in 2\mathbb{Z}\}; \\ \Lambda_4^2 &= \{\pm\alpha_k \pm \alpha_\ell : k, \ell \in \Omega, k \neq \ell\}; \\ \Lambda_4^3 &= \{\epsilon_C(\tfrac{1}{4}\alpha_\Omega - \alpha_k) : C \in \mathscr{C}, k \in \Omega\}. \end{aligned}$$

Counting, we find that

$$\begin{aligned} |\Lambda_4| &= |\Lambda_4^1| + |\Lambda_4^2| + |\Lambda_4^3| && (5.9) \\ &= 759 \cdot 2^7 + \binom{24}{2} \cdot 2^2 + 24 \cdot 2^{12} && (5.10) \\ &= 196,560. && (5.11) \end{aligned}$$

5.3 The Monster vertex algebra $V^\natural$ and the Griess algebra $\mathscr{B}$

Now that we have the Leech lattice available, we can construct one of our main objects of study. We have already seen that the Leech lattice is the unique even unimodular lattice of rank 24, such that has no elements of norm 2.

First, using the Leech lattice we form the untwisted space

$$V_\Lambda = S(\tilde{\mathfrak{h}}^-) \otimes_{\mathbb{Z}} \mathbb{C}[\Lambda] \tag{5.12}$$

as in (4.20). Let $\hat{L}$ be a central extension of a lattice L by a finite cyclic group $\langle \kappa \rangle = \{k | k^s = 1\}$ of order s, and denote by

$$c_0 : L \times L \to \mathbb{Z}/s\mathbb{Z}$$

the associated commutator map, so that

$$aba^{-1}b^{-1} = \kappa^{c_0(\bar{a},\bar{b})}, \quad \text{for } a, b \in \hat{L}.$$

We make the following special choices: we fix the central extension

$$1 \longrightarrow \langle \kappa \rangle \longrightarrow \hat{\Lambda} \longrightarrow \Lambda \longrightarrow 1, \tag{5.13}$$

where $\kappa^2 = 1$, $\kappa \neq 1$ (*i. e.*, $s = 2$), and where the commutator map is the alternating $\mathbb{Z}$-bilinear map

$$c_0(\alpha, \beta) = \langle \alpha, \beta \rangle + 2\mathbb{Z}, \quad \text{for } \alpha, \beta \in \Lambda. \tag{5.14}$$

Then, taking a 2-th primitive root of unity $\xi = -1$, we have

$$\mathbb{C}[\Lambda] = \mathbb{C}[\hat{\Lambda}]/(\kappa + 1)\mathbb{C}[\hat{\Lambda}]$$

and

$$c(\alpha, \beta) = (-1)^{\langle \alpha, \beta \rangle}, \quad \text{for } \alpha, \beta \in \Lambda,$$

where $c : \Lambda \times \Lambda \to \mathbb{C}^\times$ is the map $(\alpha, \beta) \mapsto \xi^{c_0(\alpha,\beta)}$. So that, as operators on V_Λ, $\kappa = -1$ and

$$ab = (-1)^{\langle \bar{a}, \bar{b} \rangle} ba, \quad \text{for } a, b \in \hat{\Lambda}.$$

The automorphisms of $\hat{\Lambda}$ which induce the involution -1 on Λ are automatically involutions and are parametrized by the quadratic forms on $\Lambda/2\Lambda$ with associated form induced by (5.14). Among these, we fix the distinguished involution θ_0 determined by the canonical quadratic form q_1 given in (4.28)

$$q_1 : \Lambda/2\Lambda \to \mathbb{Z}/2\mathbb{Z}, \quad \text{with } \alpha + 2\Lambda \mapsto \tfrac{1}{2}\langle \alpha, \alpha \rangle + 2\mathbb{Z}.$$

Then, we have an involution $\theta_0 : \hat{\Lambda} \to \hat{\Lambda}$ given by

$$a \mapsto a^{-1}\kappa^{\langle \bar{a}, \bar{a} \rangle / 2}.$$

Besides the general properties $\theta_0^2 = 1$ and $\theta_0(a^2) = a^{-2}$, for all $a \in \hat{\Lambda}$, we observe that $\theta_0(a) = a^{-1}$ if $\bar{a} \in \Lambda_4$.

We also form the twisted space

$$V_\Lambda^T = S(\tilde{\mathfrak{h}}^-) \otimes_{\mathbb{Z}+\frac{1}{2}} T,$$

where T is any Λ-module such that $\kappa \cdot v = \xi v$, for all $v \in T$. Keeping in mind the choices above, we set

$$K = \{\theta_0(a)a^{-1} : a \in \hat{\Lambda}\} = \{a^2\kappa^{\langle \bar{a},\bar{a}\rangle/2} : a \in \hat{\Lambda}\},$$

which is a central subgroup of $\hat{\Lambda}$ such that $K = 2\Lambda$. Then, $\hat{\Lambda}/K$ is a finite group which is a central extension

$$1 \longrightarrow \langle\kappa\rangle \longrightarrow \hat{\Lambda}/K \longrightarrow \Lambda/2\Lambda \longrightarrow 1,$$

with commutator map induced by (5.14) and with squaring map the quadratic form q_1. Since Λ is unimodular, q_1 is non-singular, and $\hat{\Lambda}/K$ is an extraspecial 2-group with

$$|\hat{\Lambda}/K| = 2^{25}. \tag{5.15}$$

In fact, in the twisted space V_Λ^T we take T to be the canonical $\hat{\Lambda}$-module described in (5.12). Of course for $a \in \hat{\Lambda}$, we have $\theta_0(a) = a$ as operators on T.

Now we can define the Moonshine module —the space on which the Monster group will act—. Recall that θ_0 acts in a natural way on V_Λ given by

$$\theta_0 : x \otimes i(a) \mapsto \theta_0(x) \otimes i(\theta_0(a)), \quad \text{for } x \in S(\tilde{\mathfrak{h}}^-) \text{ and } a \in \hat{\Lambda}; \tag{5.16}$$

and on V_Λ^T, by

$$\theta_0 : x \otimes \tau \mapsto \theta_0(x) \otimes (-\tau) = -\theta_0(x) \otimes \tau, \quad \text{for } x \in S(\tilde{\mathfrak{h}}^-) \text{ and } \tau \in T. \tag{5.17}$$

Let $V_\Lambda^{\theta_0}$ and $(V_\Lambda^T)^{\theta_0}$ be the subspaces of V_Λ and V_Λ^T of θ_0-invariant elements. We know that for $v \in V_\Lambda^{\theta_0}$ the component operators of both the untwisted and the twisted vertex operators $Y(v,z)$ preserve the respective fixed spaces $V_\Lambda^{\theta_0}$ and $(V_\Lambda^T)^{\theta_0}$.

Definition 5.3.1. We define the *Moonshine module* to be the space

$$V^\natural = V_\Lambda^{\theta_0} \oplus (V_\Lambda^T)^{\theta_0}.$$

The symbol $\natural$ is for 'natural'. For $v \in V_\Lambda$, we form the vertex operator

$$Y(v,z) = Y_{\mathbb{Z}}(v,z) \oplus Y_{\mathbb{Z}+\frac{1}{2}}(v,z)$$

acting on the larger space

$$W_\Lambda = V_\Lambda \oplus V_\Lambda^T. \tag{5.18}$$

Similarly, for the component operators of $Y(v,z)$ we write $v_n = v_n \oplus v_n$ and $x_v(n) = x_v(n) \oplus x_v(n)$ on W_Λ, for $v \in V_\Lambda$, $n \in \mathbb{Q}$. Then

$$v_n \cdot V^\natural \subseteq V^\natural, \quad x_v(n) \cdot V^\natural \subseteq V^\natural,$$

if $v \in V_\Lambda^{\theta_0}$. In fact, $V^\natural$ can be given the structure of a vertex operator algebra, the *Monster vertex algebra*. $V^\natural$ has a conformal vector ω of central charge 24. Thus, the linear maps $L_i : V^\natural \to V^\natural$ given by $L_i = \omega_{i+1}$ satisfy

$$[L_i, L_j] = (i-j)L_{i+j} + 12\binom{i+1}{3}\delta_{i+j,0}\, 1_{V^\natural},$$

and thus give a representation of the Virasoro algebra in which the central element c is represented by $24 \cdot 1_{V^\natural}$ (recall (4.13)).

Because $\dim \mathfrak{h} = 24$ and $\Lambda_2 = \emptyset$, the space $V^\natural$ has some special structural features. First, $V^\natural$ is integrally graded, with degrees bounded below by -1:

$$V^\natural = \bigoplus_{n\in\mathbb{Z}} V_n^\natural,$$

with $V_n^\natural = 0$, for all $n < -1$. In particular,

$$\begin{aligned} V_{-1}^\natural &= \mathbb{F}_\ell(1), \\ V_0^\natural &= 0, \\ V_1^\natural &= \mathfrak{f} \oplus \mathfrak{p}, \end{aligned}$$

where, in more detail

$$\begin{aligned} \mathfrak{f} &= S^2(\mathfrak{h}) \oplus \sum_{\alpha\in\hat{\Lambda}_4} \mathbb{F}x_\alpha^+, \\ \mathfrak{p} &= \mathfrak{h} \otimes T. \end{aligned}$$

Definition 5.3.2. We define the *Griess module* to be the space

$$\mathscr{B} = V_1^\natural = \mathfrak{f} \oplus \mathfrak{p}.$$

We have counted the elements of Λ_4 in (5.11), and we find that

$$\begin{aligned} \dim V_{-1}^\natural &= 1, \\ \dim V_0^\natural &= 0, \\ \dim \mathscr{B} &= \dim V_1^\natural = 196,884; \end{aligned}$$

since

$$\dim \mathfrak{f} = 300 + \tfrac{1}{2}(196,560) = 300 + 98,280 = 98,580 \tag{5.19}$$

and

$$\dim \mathfrak{p} = 24 \cdot 2^{12} = 98,304. \tag{5.20}$$

In fact we have (recall Chapter 1)

$$\sum_{n\in\mathbb{Z}}(\dim V_n)q^n = J(z) = q^{-1} + 0 + 196,884q + 21,493,760q^2 + \dots,$$

where $q = e^{2\pi iz}$, $z \in \mathbb{H}$.

It is shown in the last four chapters of [67] that the Monster group $\mathbb{M}$ acts as a group of automorphisms of $V^\natural$. The subgroup of $\mathbb{M}$ preserving the subspaces $V_\Lambda^{\theta_0}$ and $(V_\Lambda^T)^{\theta_0}$ is the centralizer of an involution in $\mathbb{M}$. This is an extension of the extraspecial group $\hat{\Lambda}/K$ of order 2^{25} in (5.15) by Conway's sporadic group Co_1, which we have seen is related to the Leech lattice Λ. A crucial part of the Frenkel-Lepowsky-Meurman construction is to find an involution in the Monster $\mathbb{M}$ which acts on $V^\natural$, but which does not preserve the subspaces $V_\Lambda^{\theta_0}$ and $(V_\Lambda^T)^{\theta_0}$.

The conformal vector $\omega \in V^\natural$ lies in a 1-dimensional subspace $\mathbb{C}\omega \subseteq V_1^\natural$, invariant under the action of $\mathbb{M}$. The complementary submodule of $\mathbb{C}\omega$ in $V_1^\natural$ gives the smallest non-trivial representation of $\mathbb{M}$ of degree 196,883. This then, is the explanation of McKay's observation (1.1) considered Chapter 1. The Monster vertex algebra $V^\natural$ is a graded module whose graded components have dimension given by the coefficients of the $J(z)$ function, and the Monster $\mathbb{M}$ acts on each graded component. On $V^\natural$, the Jacobi identity takes the following simple form for vertex operators parametrized by $V_\Lambda^{\theta_0}$ [67, p.316]:

Theorem 5.3.3. *For $v \in V_\Lambda^{\theta_0}$, we have*

$$Y(v,z) = \sum_{n\in\mathbb{Z}} v_n z^{-n-1} \quad \textit{on } V^\natural,$$

that is, $Y(v,z)$ involves only integral powers of z. For $u, v \in V_\Lambda^{\theta_0}$, we have

$$\begin{aligned}[Y(u,z_1)\times_{z_0} Y(v,z_2)] &= z_0^{-1}\,\delta\Big(\frac{z_1-z_2}{z_0}\Big)Y(u,z_1)Y(v,z_2) - z_0^{-1}\,\delta\Big(\frac{z_2-z_1}{-z_0}\Big)Y(v,z_2)Y(u,z_1)\\ &= z_2^{-1}\,\delta\Big(\frac{z_1-z_0}{z_2}\Big)Y\big(Y(u,z_0)v,z_2\big)\end{aligned}$$

on $V^\natural$. In particular, $V_\Lambda^{\theta_0}$ is a vertex operator algebra of central charge 24 and $(V_\Lambda^T)^{\theta_0}$ is a $V_\Lambda^{\theta_0}$-module.

Moreover, we know the following [67, p.317]:

Theorem 5.3.4. *The space $\mathfrak{f}$ is a commutative nonassociative algebra with identity under the product*

$$u \times v = u_1 \cdot v,$$

and the bilinear form

$$\langle u, v\rangle = u_3 \cdot v$$

is nonsingular, symmetric and associative. The space $V^\natural$ is a graded module for the commutative affinization $\hat{\mathfrak{f}}$ of $\mathfrak{f}$ under the action $\pi : \hat{\mathfrak{f}} \to \operatorname{End} V^\natural$ defined by

$$\pi : \begin{cases} u \otimes t^n \mapsto x_u(n) & \textit{for } u \in \mathfrak{f},\ n \in \mathbb{Z} \\ \quad e \mapsto 1. \end{cases} \tag{5.21}$$

One of the main results obtained in the work of Frenkel, Lepowsky and Meurman [67] is an extension of this action to a representation

$$\hat{\mathscr{B}} \to \operatorname{End} V^\natural \tag{5.22}$$

of a larger commutative affinization by cross-bracket on $V^\natural$, where the space $\mathscr{B}$ is given the structure of a commutative nonassociative algebra with identity, and with a nonsingular symmetric associative form in the following way: the product × and the form $\langle \cdot, \cdot \rangle$ on $\mathscr{B}$ extend those on $\mathfrak{f}$. For $u \in \mathfrak{f}$ and $v \in \mathfrak{p}$, we use the product in Theorem 5.3.4 and the commutativity of this product on $\mathfrak{f}$ as motivation to define

$$u \times v = v \times u = u_1 \cdot v.$$

Similarly, we use the bilinear form in Theorem 5.3.4 and the symmetry of this form on $\mathfrak{f}$ as motivation to define

$$\langle u, v \rangle = \langle v, u \rangle = u_3 \cdot v = 0$$

(the fact that $u_3 \cdot v = 0$ is obtained by consideration of the gradation of $(V_\Lambda^T)^{\theta_0}$). Now, the identity element $\frac{1}{2}\omega$ on $\mathfrak{f}$ is also an identity element on $\mathscr{B}$. We define next a nonsingular symmetric bilinear form $\langle \cdot, \cdot \rangle$ on $\mathfrak{p} = \mathfrak{h} \otimes T$ by the formula

$$\langle h_1 \otimes \tau_1, h_2 \otimes \tau_2 \rangle = \tfrac{1}{2} \langle h_1, h_2 \rangle \langle \tau_1, \tau_2 \rangle \tag{5.23}$$

for $h_i \in \mathfrak{h}$, $\tau_j \in T$. Finally, we define a product × on $\mathfrak{p}$ so that $\mathfrak{p} \times \mathfrak{p} \subseteq \mathfrak{f}$, and uniquely determined by the nonsingularity of the form on $\mathfrak{f}$ and the associativity condition

$$\langle u, v \times w \rangle = \langle u \times v, w \rangle, \quad \text{for } u, v \in \mathfrak{p},\ w \in \mathfrak{f}. \tag{5.24}$$

The commutativity of this product on $\mathfrak{p}$ follows from the explicit formula for it given below.

Definition 5.3.5. The resulting nonassociative algebra $\mathscr{B}$ equipped with this form $\langle \cdot, \cdot \rangle$ is called the *Griess algebra*.

Here, the Griess algebra $\mathscr{B} = \mathfrak{f} \oplus \mathfrak{p}$ is actually a slight modification, with a natural identity element, of the algebra defined in [79], [80]. Summarizing,

Proposition 5.3.6. *The Griess algebra $\mathscr{B}$ is a commutative nonassociative algebra with identity element $\frac{1}{2}\omega \in \mathfrak{f}$, and the form $\langle \cdot, \cdot \rangle$ on $\mathscr{B}$ is nonsingular, symmetric and associative. We have*

$$\begin{aligned} \mathfrak{f} \times \mathfrak{f} &\subseteq \mathfrak{f}, \\ \mathfrak{f} \times \mathfrak{p} &\subseteq \mathfrak{p}, \\ \mathfrak{p} \times \mathfrak{p} &\subseteq \mathfrak{f}, \end{aligned}$$

with explicit formulas given by:

$$\begin{aligned} g^2 \times h^2 &= 4\langle g, h\rangle gh, \\ g^2 \times x_a^+ &= \langle g, \bar{a}\rangle^2 x_a^+, \\ x_a^+ \times x_b^+ &= \begin{cases} 0 & \textit{if } \langle \bar{a}, \bar{b}\rangle = 0, \pm 1 \\ x_{ab}^+ & \textit{if } \langle \bar{a}, \bar{b}\rangle = -2 \\ \bar{a}^2 & \textit{if } ab = 1 \end{cases}, \end{aligned}$$

for $g, h \in \mathfrak{h}$, $a, b \in \hat{L}_4$ and $x_a^+ = i(a) + i(\theta a) = i(a) + \theta i(a)$;

$$\begin{aligned} (x_a^+)_1 \cdot (h \otimes \tau) &= \tfrac{1}{8}(h - 2\langle \bar{a}, h\rangle \bar{a}) \otimes a \cdot \tau, \\ (g^2)_1 \cdot (h \otimes \tau) &= (\langle g, h\rangle + \tfrac{1}{8}\langle g, g\rangle h) \otimes \tau, \end{aligned}$$

for $g, h \in \mathfrak{h}$, $a \in \hat{L}_4$, $\tau \in T$;

$$\begin{aligned} (h_1 \otimes \tau_1) \times (h_2 \otimes \tau_2) &= \tfrac{1}{8}\big(2h_1 h_2 + \langle h_1, h_2\rangle \tfrac{1}{2}\omega\big)\langle \tau_1, \tau_2\rangle \otimes a \cdot \tau + \\ &\quad + \tfrac{1}{128} \sum_{a \in \hat{\Lambda}_4} \big(\langle h_1, h_2\rangle - 2\langle \bar{a}, h_1\rangle\langle \bar{a}, h_2\rangle\big)\langle \tau_1, a \cdot \tau_2\rangle x_a^+, \end{aligned}$$

for $h_i \in \mathfrak{h}$, $\tau_j \in T$. We also have

$$\langle \mathfrak{f}, \mathfrak{p}\rangle = 0,$$

and explicit formulas for the form on $\mathfrak{f}$ and $\mathfrak{p}$ given by:

$$\begin{aligned} \langle g^2, h^2\rangle &= 2\langle g, h\rangle^2, \\ \langle g^2, x_a^+\rangle &= 0, \\ \langle x_a^+, x_b^+\rangle &= \begin{cases} 0 & \textit{if } \bar{a} \neq \pm\bar{b} \\ 2 & \textit{if } ab = 1 \end{cases}, \end{aligned}$$

for g, h, a, b as above and (5.23)

$$\langle h_1 \otimes \tau_1, h_2 \otimes \tau_2\rangle = \tfrac{1}{2}\langle h_1, h_2\rangle\langle \tau_1, \tau_2\rangle.$$

The identity element satisfies $\langle \frac{1}{2}\omega, \frac{1}{2}\omega\rangle = 3$.

The significance of $\mathscr{B}$ is that Griess, who introduced this algebra, constructed a group of automorphisms of it, preserving the form $\langle \cdot, \cdot \rangle$, and showed this group to be a finite simple group: the Monster $\mathbb{M}$ (see [80]). In fact, Tits showed that the Monster is the full automorphism group of $\mathscr{B}$ (see [174], [175]). Having reconstructed the Griess algebra using properties of vertex operators, we shall also reconstruct the Monster using properties of vertex operators, and exhibit a natural action of it on $V^\natural$.

5.4 The construction of $\mathbb{M}$

We now sketch the discoveries and constructions of some of the generation three of the Happy Family sporadic simple groups and the pariahs, and make some comments on their properties. The first generation is related to the Golay code $\mathcal{C}$ and the Mathieu five sporadic groups. The second one, is related to the Leech lattice and the Conway sporadic groups. For a detailed discussion on this two other generations of the Happy Family see the book of Griess [86].

The third generation of the Happy Family [80] consist of eight simple groups (see Table 5.1) which are involved in the largest sporadic simple group, the Monster $\mathbb{M}$. The construction of $\mathbb{M}$ depends on construction and analysis of a certain commutative nonassociative algebra $\mathscr{B}$, of dimension 196,884, over the rational field $\mathbb{Q}$. This is the Griess algebra we already have studied in Section 5.3. This algebra $\mathscr{B}$ plays the role of the Golay code in the First generation, and the Leech lattice in the second.

Group	Discoverer (date)	First construction (date)
He	D. Held (1968)	G. Highman, J. McKay (1968?)
Fi_{22}	B. Fischer (1968)	B. Fischer (1969)
Fi_{23}	B. Fischer (1968)	B. Fischer (1969)
$\mathrm{Fi}_{24'}$	B. Fischer (1968)	B. Fischer (1969)
HN	K. Harada (1973)	S. Norton (1974)
Th	J. Thompson (1973)	P. Smith (1974)
$\mathbb{B}$	B. Fischer (1973)	J. Leon, C. Sims (1977)
$\mathbb{M}$	B. Fischer, R. Griess (1973)	R. Griess (January, 1980)

Table 5.1: Third generation of the Happy Family.

In a sequence of steps, we now proceed to define and establish the basic properties of a group C which will act naturally on $V^\natural$, $\mathscr{B}$ and $\hat{\mathscr{B}}$, and which moreover will act compatibly with the appropriate vertex operators. This group will be the centralizer of an involution in the Monster.

Starting with the central extension $\hat{\Lambda}$ (5.13) we first set

$$C_0 = \{g \in \mathrm{Aut}\,\hat{\Lambda} : \bar{g} \in \mathrm{Co}_0\}, \tag{5.25}$$

where $\bar{g}$ is the automorphism of the Leech lattice Λ induced by g and Co_0 is the isometry group of Λ (5.9). We know that $g\kappa = \kappa$ automatically. We have also mentioned that the sequence

$$1 \longrightarrow \mathrm{Hom}(\Lambda, \mathbb{Z}/2\mathbb{Z}) \xrightarrow{*} C_0 \xrightarrow{-} \mathrm{Co}_0 \longrightarrow 1 \tag{5.26}$$

is exact, where the pullback $\lambda^* : \hat{\Lambda} \to \hat{\Lambda}$ is given by

$$a \mapsto a\kappa^{\lambda a},$$

for $\lambda \in \mathrm{Hom}(\Lambda, \mathbb{Z}/2\mathbb{Z})$. Moreover, we have natural identifications

$$\mathrm{Hom}(\Lambda, \mathbb{Z}/2\mathbb{Z}) = \Lambda/2\Lambda = \mathrm{Inn}\,\hat{\Lambda},$$

so that the exact sequence (5.26) can be written

$$1 \longrightarrow \mathrm{Inn}(\hat{\Lambda}) \hookrightarrow C_0 \xrightarrow{*} \mathrm{Co}_0 \longrightarrow 1. \tag{5.27}$$

Now, C_0 induces a group of automorphisms of the extraspecial group $\hat{\Lambda}/K$ since C_0 preserves K, and we have a natural isomorphism $\varphi : C_0 \to \mathrm{Aut}(\hat{\Lambda}/K)$. We claim that

$$\mathrm{Ker}\,\varphi = \langle \theta_0 \rangle.$$

In fact, it is clear that $\theta_0 \in \mathrm{Ker}\,\varphi$. On the other hand, $\mathrm{Ker}\,\varphi \cap \mathrm{Inn}\,\hat{\Lambda} = 1$, since $K \cap \langle \kappa \rangle = 1$. Hence, $\mathrm{Ker}\,\varphi$ is isomorphic to its image in Co_0 by (5.27). But, $\overline{\mathrm{Ker}\,\varphi}$ acts trivially on $\lambda/2\Lambda$, so that by the faithfulness of the action of Co_1 on $\Lambda/2\Lambda$, we have $\overline{\mathrm{Ker}\,\varphi} \subseteq \langle \pm 1 \rangle$, proving the claim.

Set $C_1 = \varphi(C_0) \subseteq \mathrm{Aut}(\hat{(\Lambda)}/K)$. Then we have an exact sequence

$$1 \longrightarrow \mathrm{Inn}(\hat{\Lambda}) \longrightarrow C_1 \xrightarrow{*} \mathrm{Co}_1 \longrightarrow 1. \tag{5.28}$$

Here Co_1 acts in the natural way on $\mathrm{Inn}\,\hat{\Lambda} = \Lambda/2\Lambda$, and it follows from the properties of the Conway group Co_0 that C_1 equals its commutator subgroup and has trivial center:

- $C_1 = [C_1, C_1]$;
- $\mathrm{Cent}\, C_1 = 1$.

We recall (see [67]) that the extraspecial group $\hat{\Lambda}/K$, gives the exact sequence

$$1 \longrightarrow \mathbb{F}^\times \longrightarrow \mathcal{N}_{\mathrm{Aut}\,T}(\pi(\hat{\Lambda}/K)) \xrightarrow{\mathrm{int}} \mathrm{Aut}(\hat{\Lambda}/K) \longrightarrow 1, \tag{5.29}$$

where π denotes the faithful representation of $\hat{\Lambda}/K$ on T, $\mathcal{N}_{\mathrm{Aut}\,T}(\pi(\hat{\Lambda}/K))$ is the normalizer of $\pi(\hat{\Lambda}/K)$ on $\mathrm{Aut}\,T$, and

$$\mathrm{int}(g)(x) = gxg^{-1},$$

for $g \in \mathrm{Aut}\,T$, $x \in \hat{\Lambda}/K = \pi(\hat{\Lambda}/K)$ (T the Λ-module used in the twisted space V_Λ^T). Set

$$C_* = \{g \in \mathcal{N}_{\mathrm{Aut}\,T}(\pi(\hat{\Lambda}/K)) : \mathrm{int}(g) \in C_1\},$$

so that we have the commutative diagram with exact rows

$$\begin{tikzcd}
1 \arrow[r] & \mathbb{F}^\times \arrow[r] \arrow[d, equal] & C_* \arrow[r] \arrow[d] & C_1 \arrow[r] \arrow[d] & 1 \\
1 \arrow[r] & \mathbb{F}^\times \arrow[r] & \mathcal{N}_{\mathrm{Aut}\,T}(\pi(\hat{\Lambda}/K)) \arrow[r] & \mathrm{Aut}(\hat{\Lambda}/K) \arrow[r] & 1.
\end{tikzcd}$$

Also set $C_T = [C_*, C_*]$. We shall now show that C_T contains -1 and in fact all of $\pi(\hat{\Lambda}/K)$. Since

$$\mathrm{int}(\pi(\hat{\Lambda}/K)) = \Lambda/2\Lambda = \mathrm{Inn}\,\hat{\Lambda}, \tag{5.30}$$

we see that $\pi(\hat{\Lambda}/K) \subseteq C_*$, ans so $-1 = \pi(\kappa K) \in [\pi(\hat{\Lambda}/K), \pi(\hat{\Lambda}/K))] \subseteq C_T$. But since Co_1 acts irreducibly on $\Lambda/2\Lambda$, we have $\mathrm{Inn}\,\hat{\Lambda} = [\mathrm{Inn}\,\hat{\Lambda}, C_1]$, and it follows that

$$\pi(\hat{\Lambda}/K)) = [\pi(\hat{\Lambda}/K)), C_*] \subseteq C_T. \tag{5.31}$$

We claim that the sequence

$$1 \longrightarrow \langle \pm 1 \rangle \hookrightarrow C_T \overset{\mathrm{int}}{\longrightarrow} C_1 \longrightarrow 1 \tag{5.32}$$

is exact. By the properties of C_1, all we need to show it that $C_T \cap \mathbb{F}^\times = \langle \pm 1 \rangle$. To see this, we use the fact that the $\hat{\Lambda}/K$-module T has a $\mathbb{Q}$-form,

$$T \cong \mathrm{Ind}_{\hat{\Phi}/K}^{\hat{\Lambda}/K} \mathbb{Q}_{\psi_0} \otimes_{\mathbb{Q}} \mathbb{F},$$

where ψ_0 is any rational-valued character of $\hat{\Phi}/K$ such that $\psi_0(\kappa K) = -1$, and this gives us a $\hat{\Lambda}$-invariant $\mathbb{Q}$-subspace $T_{\mathbb{Q}}$ of T such that the canonical map

$$T_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{F} \to T$$

is an isomorphism. Let $C_{*,\mathbb{Q}} = C_* \cap \mathrm{Aut}\,T_{\mathbb{Q}}$. We have an exact sequence

$$1 \longrightarrow \mathbb{Q}^\times \longrightarrow C_{*,\mathbb{Q}} \longrightarrow C_1 \longrightarrow 1,$$

and so $C_* = C_{*,\mathbb{Q}}\mathbb{F}^\times$, and $C_T \subseteq C_{*,\mathbb{Q}}$, implying that $C_T \cap \mathbb{F}^\times \subseteq C_{*,\mathbb{Q}} \cap \mathbb{F}^\times = \mathbb{Q}^\times$. But since $\det C_T = 1$, we also have $C_T \cap \mathbb{F}^\times \subseteq \{\mu \in \mathbb{F}^\times : \mu^{2^{12}} = 1\}$, proving the claim.

Using (5.28) we have a map

$$\overline{\mathrm{int}} = {}^{-} \circ \mathrm{int} : C_T \to \mathrm{Co}_1,$$

and by (5.30) and (5.31), then $\pi(\hat{\Lambda}/K) \subseteq \mathrm{Ker}\,\overline{\mathrm{int}}$. Consideration on the order of C_T shows that the sequence

$$1 \longrightarrow \hat{\Lambda}/K \overset{\pi}{\longrightarrow} C_T \overset{\overline{\mathrm{int}}}{\longrightarrow} \mathrm{Co}_1 \longrightarrow 1 \tag{5.33}$$

is exact. Summarizing, we have an extension C_0 of Co_0 by $\Lambda/2\Lambda$ (5.27), an extension C_1 of Co_1 by $\Lambda/2\Lambda$ (5.28), and an extension C_T of Co_1 by the extraspecial group $\hat{\Lambda}/K$ (5.33). Now, form the pullback

$$\hat{C} = \{(g, g_T) \in C_0 \times C_T : \varphi(g) = \mathrm{int}(g_T)\},$$

so that we have the commutative diagram of surjections

$$\begin{tikzcd}
\hat{C} \arrow[r, "\pi_1"] \arrow[d, "\pi_2"'] & C_0 \arrow[d, "\varphi"] \\
C_T \arrow[r, "\mathrm{int}"] & C_1.
\end{tikzcd}$$

Set $\hat{\theta}_0 = (\theta_0, 1), \hat{\theta} = (1, -1) \in \hat{C}$. Then

$$\begin{aligned} \mathrm{Ker}\, \pi_1 &= \langle \hat{\theta} \rangle, \\ \mathrm{Ker}\, \pi_2 &= \langle \hat{\theta}_0 \rangle, \end{aligned}$$

and

$$\mathrm{Cent}\, \hat{C} = \langle \hat{\theta}_0 \rangle \times \langle \hat{\theta} \rangle = \mathrm{Ker}(\varphi \circ \pi_1),$$

since $\mathrm{Cent}\, C_1 = 1$ and $\theta_0 \in \mathrm{Cent}\, C_0$ from the definition.

We are finally ready to define the group C: Set

$$C = \hat{C}/\langle \hat{\theta}_0 \hat{\theta} \rangle. \tag{5.34}$$

Then, the diagram above enlarges to the commutative diagram of surjections

$$\begin{tikzcd}
\hat{C} \arrow[rr, "\pi_1"] \arrow[dd, "\pi_2"'] \arrow[dr, "\pi_0"] & & C_0 \arrow[dd, "\varphi"] \\
& C \arrow[dr, "\sigma"] & \\
C_T \arrow[rr, "\mathrm{int}"] & & C_1.
\end{tikzcd}$$

Also, $\mathrm{Ker}\, \pi_0 = \langle \hat{\theta}_0 \hat{\theta} \rangle$ and

$$\mathrm{Cent}\, C = \langle z | z^2 = 1 \rangle = \mathrm{Ker}\, \sigma,$$

where

$$z = \pi_0(\hat{\theta}_0) = \pi_0(\hat{\theta}) \tag{5.35}$$

(no confusion should arise between this notation and our formal variable notation). We have the exact sequence

$$1 \longrightarrow \langle z \rangle \hookrightarrow C \xrightarrow{\sigma} C_1 \longrightarrow 1.$$

Proceeding as in (5.33), we have a map $\bar{\sigma} = {}^{-} \circ \sigma : C \to \mathrm{Co}_1$ from (5.28). Moreover, there is a canonical embedding $\nu : \hat{\Lambda}/K \to \hat{C}$ given by

$$gK \mapsto (\mathrm{int}(g), \pi(g)),$$

since $\varphi(\mathrm{int}(g)) = \mathrm{int}(\pi(g))$. The result is an exact sequence

$$1 \longrightarrow \hat{\Lambda}/K \overset{\pi_0 \circ \nu}{\longrightarrow} C \overset{\bar{\sigma}}{\longrightarrow} \mathrm{Co}_1 \longrightarrow 1,$$

and we have proved

Proposition 5.4.1. *The group C is an extension of Co_1 by the extraspecial group $\hat{\Lambda}/K$. The nontrivial central element of $\hat{\Lambda}/K$ identifies with the nontrivial central element of C.*

Now that the group C is constructed we shall set up its canonical action on the Moonshine module $V^\natural$. First, we shall define an action of the larger group $\hat{C}$ on the larger space W_Λ (see (5.18)). For $g \in \mathrm{Co}_0$ and $Z = \mathbb{Z}$ or $\mathbb{Z} + \frac{1}{2}$, let g also denote the unique algebra automorphism

$$g : S(\tilde{\mathfrak{h}}_Z^-) \to S(\tilde{\mathfrak{h}}_Z^-),$$

(where $S(\tilde{\mathfrak{h}}_Z^-)$ is an abbreviation for $S(\tilde{\mathfrak{h}}^-) \otimes_Z T$, for the appropriate module T), such that g agrees with its natural action on $\tilde{\mathfrak{h}}^-$. For $g \in C_0$, let g also denote the operator

$$g : \mathbb{F}[\Lambda] \to \mathbb{F}[\Lambda] \ \text{ given by } \ i(a) \mapsto i(ga),$$

for $a \in \hat{\Lambda}$; note that this is well defined since $g\kappa = \kappa$. For $k = (g, g_T) \in \hat{C}$, let k also denote the operator

$$k = \bar{g} \otimes g \oplus \bar{g} \otimes g_T$$

on $W_\Lambda = S(\tilde{\mathfrak{h}}^-) \otimes_{\mathbb{Z}} \mathbb{F}[\Lambda] \oplus S(\tilde{\mathfrak{h}}^-) \otimes_{\mathbb{Z}+\frac{1}{2}} T$. This clearly gives a faithful representation of $\hat{C}$ on W_Λ. In fact, this representation is even faithful on a small subspace of W_Λ, for instance $T \oplus \mathfrak{p}$.

The action of $\hat{C}$ on W_Λ extends the action of θ_0 similar to that defined in (5.16) and (5.17) for the case W_Λ, in such a way that this operator corresponds to the element $\hat{\theta_0}\hat{\theta} = (\theta_0, -1)$ of $\hat{C}$. We have

$$\hat{C} \cdot V^\natural \subseteq V^\natural.$$

From the definitions of C and $V^\natural$ and the last paragraph, we see that C acts in a natural way on $V^\natural$: for $k = (g, g_T) \in \hat{C}$, $\pi_0(k)$ acts as the operator

$$\pi_0(k) = \bar{g} \otimes g \oplus \bar{g} \otimes g_T. \tag{5.36}$$

This action of C on $V^\natural$ is faithful, even on $\mathfrak{p} = \mathfrak{h} \otimes T$. The decomposition $V^\natural = V_\Lambda^{\theta_0} \oplus (V_\Lambda^T)^{\theta_0}$ in Definition 5.3.1 is the eigenspace decomposition with respect to the central involution z in C (5.35), and we introduce corresponding notation

$$\begin{aligned} V_\Lambda^{\theta_0} &= V^z = \{v \in \mathbb{V}^\natural : z \cdot v = v\}, \\ (V_\Lambda^T)^{\theta_0} &= V^{-z} = \{v \in \mathbb{V}^\natural : z \cdot v = -v\}. \end{aligned}$$

Note that $V^\natural = V^z \oplus V^{-z}$ and $C \cdot V^z \subseteq V^z$, $C \cdot V^{-z} \subseteq V^{-z}$.The actions of C on $V^\natural$ and of $\hat{C}$ on W_Λ preserve the homogeneous subspaces with respect to the gradings. The group C acts on the algebra $\mathscr{B} = V_1^\natural$ (Definition 5.3.2) and in fact preserves the summands $\mathfrak{f}$ and $\mathfrak{p}$. From the definition of the product $\times$ and the form $\langle \cdot, \cdot \rangle$ on $\mathscr{B}$ (see Proposition 5.3.6), we find [67, p.328]:

Proposition 5.4.2. *The group C acts faithfully as automorphisms of the Griess algebra $\mathscr{B}$ and as isometries of $\langle \cdot, \cdot \rangle$.*

In fact, in checking that the form (5.23) on $\mathfrak{p}$ is preserved by C, we use the fact that the form $\langle \cdot, \cdot \rangle$ on T is C_T-invariant:

$$\langle g\tau_1, g\tau_2 \rangle = \langle \tau_1, \tau_2 \rangle, \quad \text{for all } g \in C_T,\ \tau_i \in T.$$

Recall from Theorem 5.3.4 that $V^\natural$ is a graded module for the commutative affinization $\hat{\mathfrak{f}}$ of $\mathfrak{f}$. We shall relate this structure to the action of C. Given a commutative nonassociative algebra $\mathfrak{b}$ with a symmetric form and given a group G of linear automorphisms of $\mathfrak{b}$, we let G act as linear automorphisms of $\hat{\mathfrak{b}}$ by

$$\begin{aligned} g \cdot e &= e, \\ g \cdot (u \otimes t^n) &= (g \cdot u) \otimes t^n, \end{aligned}$$

for $g \in G$, $u \in \mathfrak{b}$ and $n \in \mathbb{Z}$. If G acts as algebra automorphisms and isometries of $\mathfrak{b}$, then G acts as algebra automorphisms of $\hat{\mathfrak{b}}$. Suppose now that V is a graded $\hat{\mathfrak{b}}$-module and that G acts as linear automorphisms of V, preserving each homogeneous subspace V_n. Then, we call V a *graded $(G, \hat{\mathfrak{b}})$-module* if

$$gxg^{-1} = g \cdot x \text{ as operators on } V,$$

for $g \in G$, $x \in \hat{\mathfrak{b}}$. By Proposition 5.4.2 we have

Proposition 5.4.3. *The Moonshine module $V^\natural$ is a graded $(C, \hat{\mathfrak{f}})$-module, and C acts as automorphisms of $\hat{\mathfrak{f}}$, and in fact of $\hat{\mathscr{B}}$.*

5.5 Improvements on the Construction

We have already seen in (5.36) and Proposition 5.4.2 that the group C acts as automorphisms on these main structures: the Griess algebra $\mathscr{B}$ and the Moonshine module $V^\natural$. In fact, the idea for complete the construction of the Monster group $\mathbb{M}$ lies in enlarge C to a group M (the Monster) of automorphisms and isometries of $\mathscr{B}$ (and hence automorphisms of $\hat{\mathscr{B}}$), and make $V^\natural$ into a $(M, \hat{\mathscr{B}})$-module. Furthermore, we should also define vertex operators $Y(u, z)$ on $V^\natural$, for all $v \in V^\natural$ in order to extend an action of C on $V^\natural$ (consequence of the action of $\hat{C}$ on W_Λ) of type

$$kY(v, z)k^{-1} = Y(kv, z), \quad \text{for all } k \in C,\ v \in V_\Lambda^{\theta_0},$$

to all M and $V^\natural$. We will not describe the complete construction of $\mathbb{M}$ (this requires in fact a complete book). For details on the complete construction, the reader may consult the paper of Griess [80], or the last three chapters of [67]. Although, we mention some notably properties.

Under the action of C, $\mathscr{B}$ breaks into the following invariant subspaces:

$$\mathbb{F}\omega,\ \{u \in S^2(\mathfrak{h}) : \langle u, \omega \rangle = 0\},\ \sum_{a \in \hat{\Lambda}_4} \mathbb{F}x_a^+,\ \mathfrak{h} \otimes T \tag{5.37}$$

of dimensions 1, 299, 98280 and 98304, respectively. Of course, C in fact fixes ω:

$$C \cdot \omega = \omega.$$

It can be shown that each of the invariants subspaces is absolutely irreducible under C; for instance, $\mathfrak{h} \otimes T$ is irreducible since T is irreducible under the extraspecial group $\hat{\Lambda}/K$ and $\mathfrak{h}$ is irreducible under Co_0. Before the Monster was proved to exist, it was postulated to be a finite simple group containing the group C as the centralizer of the involution $z \in C$ and it was believed to have a 196,883-dimensional irreducible module consisting of the direct sum of the las three C-modules listed in (5.37), or rather, abstract C-modules isomorphic to them. By 1976, Norton proved the existence of an invariant commutative nonassociative algebra and nonsingular associative symmetric bilinear form on this module if it and the Monster existed (*c.f.* [80]), though his method gave no description (he also made assumptions on the conjugacy classes). By constraining the possibilities for such an algebra and form on the direct sum of the C-modules, Griess was able to determine an algebra and a form admitting an automorphism outside the group C. The group generated by C and this automorphism had the required properties.

The original definition of the algebra structure in [80] was complicated, due mainly to sign problems. The existence of the mentioned irreducible representation of the Monster of degree 196,883 was predicted in 1974 by Griess [78]. With the Norton's improvement of the existence for such a nontrivial commutative nonassociative algebra structure on this module, and the study of the automorphisms outside of C, by surveying the actions of many subgroups of C on the space $\mathscr{B}$, the result was the system of structure constants in [80, Table 6.1], and the formula for an extra automorphism in [80, Table 10.2]. Prove that the linear automorphism so defined preserves the algebra structure was the hardest part of the construction in [80].

Many improvements on [80] were made by Tits [174, 175, 176], who showed that some definitions of [80] based on guesswork may be based on a more thorough analysis. A new style construction was made by Conway in [34], [39], using a Moufang loop to finesse the sign problems so prominent in the original version. This loop (a nonassociative group) has order 2^{13} and is a kind of 2-cover of the binary Golay code; its creation was an idea of Richard Parker (see [81]), which defines and gives the foundation of the class of loops called *code loops*. This theory is usually developed in the theory of p-locals in sporadic groups and Lie groups [81, 83, 84, 85].

The algebra $\mathscr{B}$ is not a classic nonassociative algebra. An algebra of dimension n satisfies a nontrivial polynomial identity of degree at most $n+1$; $\mathscr{B}$ satisfies no nontrivial identity in commuting variables of degree less than 6 [82]. In [80], the subgroup of $\operatorname{Aut}\mathscr{B}$ generated by C and the particular extra automorphism was identified as a simple group of the right order, thus proving the existence of a simple group of the right order and local properties. The full automorphism group of a finite dimensional algebra is an algebraic group. In [175], Tits showed that $\operatorname{Aut}\mathscr{B}$ was exactly the Monster $\mathbb{M}$. In [34], Conway gave a short argument with idempotents in $\mathscr{B}$ that $\operatorname{Aut}\mathscr{B}$ is finite and in [175], Tits identified the centralizer of an involution in $\operatorname{Aut}\mathscr{B}$ as C (not a larger group). The proof that the group order is right involves quoting harder theorems (see chapter 13 of [80]). In 1988, a uniqueness proof for the Monster was given by Griess, Meierfrankenfeld and Segev [87].

We close our discussion with this uniqueness result. First, we define a group of *Monster-type* to be a finite group G containing a pair of involutions z, t such that

- $C(z) \cong 2_{+}^{1+24}\mathrm{Co}_1$;
- $C(t)$ is a double cover of Fischer's $\{3,4\}$-transposition group (discovered by Fischer in 1973, and later called the Baby Monster $\mathbb{B}$).

Is follows that such a group is simple. See [87] for a fuller discussion of the hypotheses.

Theorem 5.5.1 (Uniqueness of the Monster). *A group of Monster-type is unique up to isomorphism.*

Chapter 6

Intermezzo: the j function

The j function plays an important role in modern number theory. It classifies the family of elliptic curves over the complex field $\mathbb{C}$. In fact, the j functions serves as the moduli space of this family of curves. In this and next chapters, our purpose is to detail in a better form some aspects of the original Conway-Norton conjecture. First, we will give here some background theory of modular forms and Hecke operators, and subsequently we outline some features of replicable functions introduced in [38].

6.1 The modular group

Let $\mathbb{H}$ denote the upper half-plane of $\mathbb{C}$ we have introduced in Chapter 1. That is $\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$. Let $SL_2(\mathbb{R})$ be the group of matrices $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with real entries, such that $ad - bc = 1$. We made $SL_2(\mathbb{R})$ act on $\mathbb{C}$ by: if $G = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, then

$$Gz = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}.$$

One can easily checks the formula

$$\mathrm{Im}(Gz) = \frac{\mathrm{Im}(z)}{|cz + d|^2},$$

from which can be showed that $\mathbb{H}$ is stable under the action of $SL_2(\mathbb{R})$. We have also seen in Chapter 1 that the element $-I = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ of $SL_2(\mathbb{R})$ acts trivially on $\mathbb{H}$. Thus we can consider the group $PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm I\}$ which operates —and in fact, acts faithfully—, and one can even show that it is the group of all analytic automorphisms of $\mathbb{H}$.

Let $SL_2(\mathbb{Z})$ be the subgroup of $SL_2(\mathbb{R})$ consisting of the matrices with integer entries. It is a discrete subgroup of $SL_2(\mathbb{R})$, thus it acts discontinually on $\mathbb{H}$.

Definition 6.1.1. The group $\Gamma = PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I\}$ is called the *modular group*. It is the image of $SL_2(\mathbb{Z})$ on $PSL_2(\mathbb{R})$. We denote by $\mathbb{H}/\Gamma$ the set of action orbits of Γ on $\mathbb{H}$.

For simplicity, if G is an element of $SL_2(\mathbb{Z})$, we will use the same symbol to denote its image in the modular group Γ. Let $S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ denote the elements of Γ

introduced in Chapter 1. We have seen that

$$\begin{aligned} Sz &= -\tfrac{1}{z}, \qquad & Tz &= z+1, \\ S^2 &= 1, \qquad & (ST)^3 &= 1. \end{aligned}$$

On the other hand, let D the subset of $\mathbb{H}$ formed of all points z such that $|z| \geq 1$ and $|\operatorname{Im}(z)| \leq 1/2$. The Figure 6.1 below represents the transforms of D by some of the elements of the group Γ.

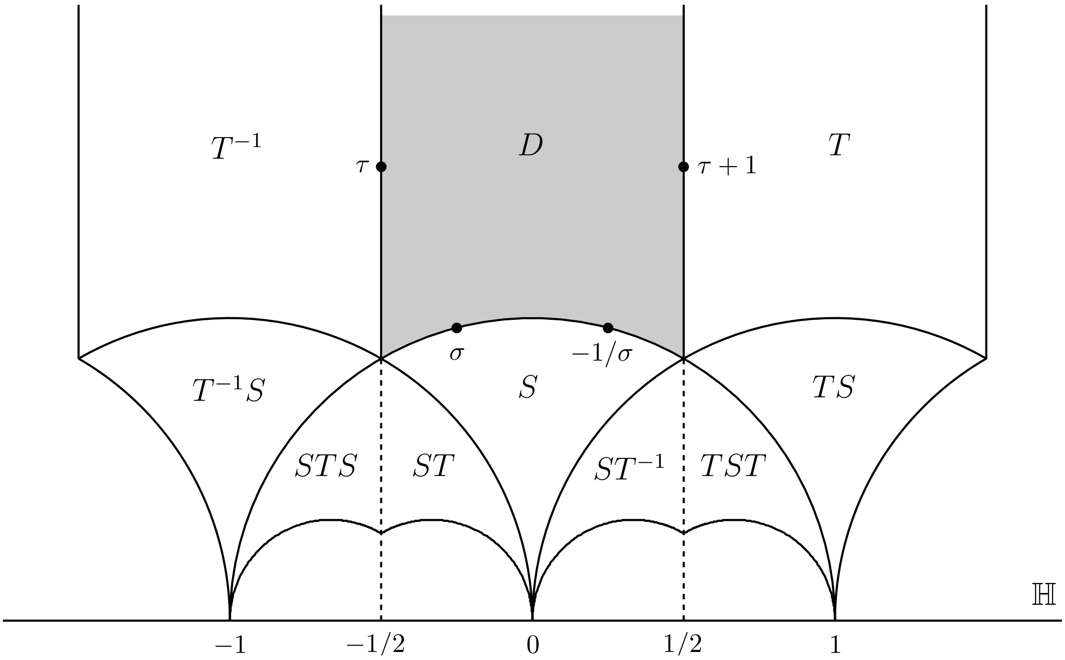

Figure 6.1: Fundamental domain D and some of its images by S and T.

We will show that D is a *fundamental domain* for the action of Γ on the half-plane $\mathbb{H}$. More precisely:

Theorem 6.1.2. *1. For every $z \in \mathbb{H}$ there exists $G \in \Gamma$ such that $Gz \in D$.*

2. Suppose that two distinct points z, z' of D are congruent modulo Γ. Then, $\operatorname{Re}(z) = \pm\frac{1}{2}$ *and $z' = z \pm 1$, or $|z| = 1$ and $z' = -\frac{1}{z}$.*

3. Let $z \in D$ and let $I(z) = \{G \in \Gamma : Gz = z\}$ be the stabilizer of z in Γ. One has $I(z) = \{I\}$, except in the following three cases:

- *$z = i$, in which case $I(z)$ is the group of order 2 generated by S,*
- *$z = \omega = e^{2\pi i/3}$, in which case $I(z)$ is the group of order 3 generated by ST,*
- *$z = -\omega^2 = e^{\pi i/3}$, in which case $I(z)$ is the group of order 3 generated by TS.*

Theorem 6.1.3. *The modular group* Γ *is generated by* S *and* T.

Proof. To prove Theorems 6.1.2 and 6.1.3, consider the subgroup Γ' of Γ, generated by S and T, and let $z \in \mathbb{H}$. If $G = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is an element of Γ', then $\operatorname{Im}(Gz) = \operatorname{Im}(z)/|cz+d|^2$. Since c and d are integers, the number of pairs (c, d) such that $|cz+d|$ is less that a given number is finite. This shows that there exist $G \in \Gamma'$ such that $\operatorname{Im}(Gz)$ is maximum. Choose now an integer n such that T^nGz has real part between $-\frac{1}{2}$ and $\frac{1}{2}$. The element $z' = T^nGz$ belongs to D; indeed, it suffices to show that $|z'| \geq 1$, but if $|z'| < 1$ then the element $-\frac{1}{z'}$ would have an imaginary part strictly greater than $\operatorname{Im}(z')$, which is impossible. Thus, the element T^nG of Γ has the desired property. This proves (1) in Theorem 6.1.2.

We now prove assertions (2) and (3). Let $z \in D$ and let $G = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ such that $Gz \in D$. Without lose of generality (replacing (z, G) with (Gz, G^{-1}) if necessary) we may suppose that $\operatorname{Im}(Gz) \geq \operatorname{Im}(z)$, *i. e.*, that $|cz+d| \leq 1$. This is clearly impossible if $|c| \geq 2$, leaving the cases $c = 0, 1, -1$. If $c = 0$, then $d = \pm 1$ and G is the translation by $\pm b$. Since $\operatorname{Re}(z)$ and $\operatorname{Re}(Gz)$ are both between $-\frac{1}{2}$ and $\frac{1}{2}$, this implies that either $b = 0$ and $G = I$, or $b = \pm 1$ in which case one of the numbers $\operatorname{Re}(z)$ or $\operatorname{Re}(Gz)$ must be equal to $-\frac{1}{2}$ and the other to $\frac{1}{2}$. If $c = 1$, the fact that $|z+d| \leq 1$ implies $d = 0$, except if $z = \omega$ (respectively if $z = -\omega^2$), in which case we can have $d = 0, 1$ (respectively $d = 0, -1$). The case $d = 0$ gives $|z| \leq 1$, hence $|z| = 1$; on the other hand, $ad - bc = 1$ implies $b = -1$, hence $Gz = a - \frac{1}{z}$ and the first part of discussion proves that $a = 0$, except in the cases $\operatorname{Re}(z) = \pm\frac{1}{2}$, *i. e.*, if $z = \omega, -\omega^2$, in which cases we have $a = 0, 1$ or $a = 0, -1$. The case $z = \omega$, $d = 1$ gives $a - b = 1$ and $G\omega = a - \frac{1}{\omega+1} = a + \omega$, hence $a = 0, 1$. We argue similar when $z = -\omega^2$, $d = -1$. Finally, the case $c = -1$ leads to the case $c = 1$ by changing the signs of a, b, c, d (which does not change G seen as an element of Γ). This completes the verification of (2) and (3).

To complete the proof of Theorem 6.1.3, it remains to prove that $\Gamma' = \Gamma$. Choose an element $z_0 \in \operatorname{int} D$ (for example $z_0 = 2i$), and let $z = Gz_0$. We have seen above that there exists an element $G' \in \Gamma'$ such that $G'z \in D$. The points $z_0, G'z = G'Gz_0$ of D are congruent modulo Γ (they lie on the same orbit in $\mathbb{H}/\Gamma$), and one of them is interior to D. By (2) and (3) of Theorem 6.1.2, it follows that these points coincide, so that $G'G = I$. Thus, $G = (G')^{-1} \in \Gamma'$, which completes the proof. □

Corollary 6.1.4. *The canonical map* $D \to \mathbb{H}/\Gamma$ *is surjective, and its restriction to* $\operatorname{int} D$ *is injective. In particular,* D *is a fundamental domain for the action of* Γ *on* $\mathbb{H}$.

Thus, D is a set intersecting each orbit of $\mathbb{H}/\Gamma$ just at one point. As a remark, one can show that $\langle S, T \mid S^2, (ST)^3 \rangle$ is a presentation of Γ, or, equivalently, that Γ is the free product of the cyclic group of order 2 generated by S and the cyclic group of order 3 generated by ST.

6.2 Modular forms

Definition 6.2.1. Let k be an integer. We say that a function $f : \mathbb{H} \to \mathbb{C}$ is *weakly modular* of weight $2k$ if f is meromorphic on the upper half-plane $\mathbb{H}$ and if it verifies the relation

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^{2k} f(z), \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}). \tag{6.1}$$

Let G be the image in Γ of $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. We have $\frac{d(Gz)}{dz} = (cz+d)^{-2}$. The relation (6.1) can then be written as

$$\frac{f(Gz)}{f(z)} = \left(\frac{d(Gz)}{dz}\right)^{-k}$$

or

$$f(Gz)d(Gz)^k = f(z)dz^k, \tag{6.2}$$

where $dz^k = dz \otimes \ldots \otimes dz$ (k times). It means that the tensor product of differential 1-forms $f(z)dz^k$ is invariant under Γ. Since Γ is generated by the elements S and T (Theorem 6.1.3), it suffices to check the invariance by S and by T. This gives:

Proposition 6.2.2. *Let f be a meromorphic function on $\mathbb{H}$. Then, f is weakly modular of weight $2k$ if, and only if, for all $z \in \mathbb{H}$ it satisfies the two relations:*

$$f(z+1) = f(z), \tag{6.3}$$
$$f(-1/z) = z^{2k} f(z). \tag{6.4}$$

Suppose the relation (6.3) is verified. We have already noticed that f can be expressed as a function of $q = e^{2\pi i z}$, function which we will denote by $\breve{f}$; it is meromorphic in the disk $|q| < 1$ with the origin removed. We can think the value $q = 0$ making $z = i\infty \in \overline{\overline{\mathbb{H}}}$, and, if $\breve{f}$ extends to a meromorphic (respectively holomorphic) function at the origin, we say, by abuse of language, that f is *meromorphic* (respectively *holomorphic*) *at infinity*. This means that $\breve{f}$ admits a Laurent expansion in a neighborhood of the origin

$$\breve{f}(q) = \sum_{-\infty}^{\infty} a_n q^n,$$

where the a_n's are zero for n small enough (respectively for $n < 0$).

Definition 6.2.3. If a weakly modular function f is meromorphic at infinity, we say that f is a *modular* function. When f is holomorphic at infinity, we set $f(\infty) = \breve{f}(0)$. This is the value of f at infinity.

Definition 6.2.4. A modular function which is holomorphic everywhere —including at infinity— is called a *modular form*; if such a function is zero at infinity, it is called a *cusp form* (or *Spitzenform*, or *forme parabolique*).

Thus, a modular form of weight $2k$ is given by a series $f(z) = \sum_{n \geq 0} a_n q^n = \sum_{n \geq 0} a_n e^{2\pi i n z}$, which converges for $|q| < 1$ (*i. e.*, for $\mathrm{Im}(z) > 0$) and that verifies the identity $f(-1/z) = z^{2k} f(z)$. It is a cusp form if $a_0 = 0$.

Example 6.2.5. If f and g are modular forms of weight $2k$, then any $\mathbb{C}$-linear combination $\alpha f + \beta g$ is a modular form of weight $2k$. Thus, the modular forms of weight $2k$ forms a $\mathbb{C}$-vector space.
If f and f' are modular forms of weight $2k$ and $2k'$, then the product fg is a modular form of weight $2k + 2k'$.

Example 6.2.6. We will see later that the function

$$q \prod_{n \geq 1} (1 - q^n)^{24} = q - 24q^2 + 252q^3 - 1472q^4 + \ldots, \quad q = e^{2\pi i z},$$

is a cusp form of weight 12.

6.3 Lattice functions and modular functions

Recall that a lattice in a real vector space V of finite dimension is a subgroup Λ of V satisfying one of the following equivalent conditions:

- Λ is discrete and V/Λ is compact;
- Λ is discrete and generates the $\mathbb{R}$-vector space V;
- There exists an $\mathbb{R}$-basis $e_1, \ldots, e_n$ of V which is a $\mathbb{Z}$-basis for Λ, that is Λ is the n-rank free abelian group $\Lambda = \mathbb{Z}e_1 \oplus \ldots \oplus \mathbb{Z}e_n$.

Let $\mathscr{L}$ be the set of lattices of $\mathbb{C}$ considered as a real vector space. Let M be the set of pairs (ω_1, ω_2) of elements of $\mathbb{C}^\times$ such that $\mathrm{Im}\left(\frac{\omega_1}{\omega_2}\right) > 0$. To such a pair we associate the lattice

$$\Lambda(\omega_1, \omega_2) = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$$

with basis $\{\omega_1, \omega_2\}$. We thus obtain a map $M \to \mathscr{L}$ which is clearly surjective. Let $G = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$ and let $(\omega_1, \omega_2) \in M$. We put

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} a\omega_1 + b\omega_2 \\ c\omega_1 + d\omega_2 \end{pmatrix}.$$

It is clear that $\{\omega_1', \omega_2'\}$ is a basis for $\Lambda(\omega_1, \omega_2)$. Moreover, if we set $z = \frac{\omega_1}{\omega_2}$ and $z' = \frac{\omega_1'}{\omega_2'}$ we have

$$z' = \frac{az + b}{cz + d} = Gz.$$

This shows that $\mathrm{Im}(z') > 0$, hence that (ω_1', ω_2') belongs to M. In fact, we have the following result

Proposition 6.3.1. *For two elements of M to define the same lattice it is necessary and sufficient that they are congruent modulo $SL_2(\mathbb{Z})$.*

Proof. We just saw that the condition is sufficient. Conversely, if (ω_1, ω_2) and (ω_1', ω_2') are two elements of M which define the same lattice, there exists an integer matrix $G = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ of determinant ± 1 which transforms the first basis into the second. If $\det(G) < 0$, the sign of $\mathrm{Im}(z')$ would be the opposite of $\mathrm{Im}(z')$ as one sees by an immediate computation. The two signs being the same, we have necessarily $\det(G) = 1$. □

Hence we can identify the set $\mathscr{L}$ of lattices of $\mathbb{C}$ with the quotient of $M/SL_2(\mathbb{Z})$. Make now $\mathbb{C}^\times$ act on $\mathscr{L}$ (respectively on M) by:

$$\Lambda \mapsto \lambda\Lambda \ \ \text{(respectively } (\omega_1, \omega_2) \mapsto (\lambda\omega_1, \lambda\omega_2)), \quad \text{for } \lambda \in \mathbb{C}^\times.$$

The quotient $M/\mathbb{C}^\times$ is identified with $\mathbb{H}$ by $(\omega_1, \omega_2) \mapsto z = \frac{\omega_1}{\omega_2}$ and this identification transforms the action of $SL_2(\mathbb{Z})$ on M into that of the modular group $\Gamma = PSL_2(\mathbb{Z})$ on $\mathbb{H}$. Thus:

Proposition 6.3.2. *The map $(\omega_1, \omega_2) \mapsto \frac{\omega_1}{\omega_2}$ gives by passing to the quotient, a bijection of $\mathscr{L}/\mathbb{C}^\times$ onto $\mathbb{H}/\Gamma$. (Thus, an element of $\mathbb{H}/\Gamma$ can be identified with a lattice of $\mathbb{C}$ defined up to a homothety.)*

In number theory it is frequent to associate to a lattice Λ of $\mathbb{C}$ the elliptic curve $E_\Lambda = \mathbb{C}/\Lambda$. It is easy to see that two lattices Λ and Λ' define isomorphic elliptic curves if and only if they are homothetic. This gives a third description of $\mathbb{H}/\Gamma = \mathscr{L}/\mathbb{C}^\times$: it is the set of isomorphism classes of elliptic curves.

Let us now pass to modular functions. Let F be a function defined on $\mathscr{L}$, with complex values, and let $k \in \mathbb{Z}$. We say that F is a *modular lattice function* of weight $2k$ if

$$F(\lambda\Lambda) = \lambda^{-2k} F(\Lambda), \quad \text{for all lattices } \Lambda \text{ and all } \lambda \in \mathbb{C}^\times. \tag{6.5}$$

Let F be such a function. If $(\omega_1, \omega_2) \in M$, we denote by $F(\omega_1, \omega_2)$ the value of F on the lattice $\Lambda(\omega_1, \omega_2)$. The formula (6.5) translates to:

$$F(\lambda\omega_1, \lambda\omega_2) = \lambda^{-2k} F(\omega_1, \omega_2). \tag{6.6}$$

Moreover, $F(\omega_1, \omega_2)$ is invariant by the action of $SL_2(\mathbb{Z})$ on M. Formula (6.6) shows that the product $\omega_2^{2k} F(\omega_1, \omega_2)$ depends only on $z = \frac{\omega_1}{\omega_2}$. There exists then a function f on $\mathbb{H}$ such that

$$F(\omega_1, \omega_2) = \omega_2^{-2k} f(\omega_1/\omega_2). \tag{6.7}$$

Writing that F is invariant by $SL_2(\mathbb{Z})$, we see that f satisfies the identity (6.1). Conversely, if f verifies (6.1), formula (6.7) associates to it a function F on $\mathscr{L}$ which is of weight $2k$. We can thus identify modular functions of weight $2k$ with some lattice functions of weight $2k$.

Example 6.3.3 (Example of modular functions: Eisenstein series).

Lemma 6.3.4. *Let Λ be a lattice in $\mathbb{C}$. The series*

$$\sum_{\gamma\in\Lambda}{}' \frac{1}{|\gamma|^{\sigma}}$$

is convergent for $\sigma > 2$. (The symbol Σ' signifies that summation runs over all nonzero elements of Λ).

Proof. We can proceed as with the series $\sum \frac{1}{n^{\sigma}}$, *i. e.*, majorize the series under consideration by a multiple of the double integral $\iint \frac{dxdy}{(x^2+y^2)^{\sigma/2}}$ extended over the plane without a disk with center at 0. The double integral is computed using the classical technique with polar coordinates. Another equivalent method consists in remarking that the number of elements of Λ such that $|\gamma|$ is between two consecutive integers n and $n+1$ is $O(n)$; the convergence of the series is then reduced to that of the series $\sum \frac{1}{n^{\sigma-1}}$. □

Now let k be an integer > 1. If Λ is a lattice of $\mathbb{C}$, put

$$G_k(\Lambda) = \sum_{\gamma\in\Lambda}{}' \frac{1}{\gamma^{2k}}. \tag{6.8}$$

This series is absolutely convergent, thanks to Lemma 6.3.4. Observe that G_k is a modular lattice function of weight $2k$. It is called the *Eisenstein series* of index k. As in the preceding section, we can view G_k as a function on M, given by

$$G_k(\omega_1, \omega_2) = \sum_{m,n}{}' \frac{1}{(m\omega_1 + n\omega_2)^{2k}}. \tag{6.9}$$

Here again the symbol Σ' means that the summation runs over all pairs of integers (m, n) distinct from $(0, 0)$. The function on $\mathbb{H}$ corresponding to G_k (by the procedure given in previous section) is denoted also by G_k. By formulas (6.7) and (6.9), we have

$$G_k(z) = \sum_{m,n}{}' \frac{1}{(mz + n)^{2k}}. \tag{6.10}$$

Proposition 6.3.5. *Let k an integer > 1. The Eisenstein series $G_k(z)$ is a modular form of weight $2k$. We have $G_k(\infty) = 2\zeta(2k)$, where ζ denotes the Riemann zeta function.*

Proof. The above arguments show that $G_k(z)$ is weakly modular of weight $2k$. We have to show that G_k is holomorphic everywhere. First suppose that z is contained in the fundamental domain D. Then

$$\begin{aligned} |mz+n|^2 &= (mz+n)(\overline{mz+n}) = (mz+n)(m\bar{z}+n) = m^2 z\bar{z} + 2mn\,\mathrm{Re}(z) + n^2 \\ &\geq m^2 - mn + n^2 = |m\omega - n|^2. \end{aligned}$$

By Lemma 6.3.4, the series $\sum{}' \frac{1}{|m\omega - n|^{2k}}$ is convergent. This shows that the series $G_k(z)$ converges in D, thus also (applying the result to $G_k(G^{-1}z)$ with $G \in \Gamma$) in each of the transforms GD of D by Γ. Since these sets cover $\mathbb{H}$, we see that G_k is holomorphic on $\mathbb{H}$. It remains to verify that G_k is holomorphic at infinity (and to find the value at this point). This reduces to prove that G_k has a limit for $\mathrm{Im}(z) \to \infty$. But one may suppose that z remains in the fundamental domain D; in view of the uniform convergence in D, we can make the passage to the limit term by term. The terms $\frac{1}{(mz+n)^{2k}}$ relative to $m \neq 0$ give 0; the others give $\frac{1}{n^{2k}}$. Thus

$$\lim_{z\to\infty} G_k(z) = \sum_n{}' \frac{1}{n^{2k}} = 2\sum_{n\geq 1} \frac{1}{n^{2k}} = 2\zeta(2k). \ \square$$

The Eisenstein series of lowest weights are G_2 and G_3, which are of weight 4 and 6. It is conveniently (because of the theory of elliptic curves) to replace these by some multiples:

$$g_2 = 60G_2, \quad g_3 = 140G_3.$$

We have $g_2(\infty) = 120\zeta(4)$ and $g_3(\infty) = 280\zeta(6)$. Using the known values of $\zeta(4)$ and $\zeta(6)$ (see for example Table 6.2), one finds that

$$g_2(\infty) = \tfrac{4}{3}\pi^4, \quad g_3(\infty) = \tfrac{8}{27}\pi^6.$$

If we put

$$\Delta = g_2^3 - 27g_3^2, \tag{6.11}$$

then we have $\Delta(\infty) = 0$; that is to say, Δ is a cusp form of weight 12.
In fact, all these stuff is related to the theory of elliptic curves. Let Λ be a lattice of $\mathbb{C}$ and let

$$\wp_\Lambda(u) = \frac{1}{u^2} + \sum_{\gamma\in\Lambda}{}' \Big(\frac{1}{(u-\gamma)^2} + \frac{1}{\gamma^2} \Big), \tag{6.12}$$

be the corresponding *Weierstrass function*. Then, $G_k(\Lambda)$ occur into the Laurent expansion of $\wp_\Lambda$:

$$\wp_\Lambda(u) = \frac{1}{u^2} + \sum_{k\geq 2} (2k-1)G_k(\Lambda)u^{2k-2}.$$

If we put $x = \wp_\Lambda(u)$ and $y = \wp'_\Lambda(u)$, we have

$$y^2 = 4x^3 - g_2x - g_3, \tag{6.13}$$

with $g_2 = 60G_2$, $g_3 = 140G_3$ as above. Up to a numerical factor, $\Delta = g_2^3 - 27g_3^2$ is equal to the *discriminant* of the polynomial $4x^3 - g_2x - g_3$. Usually one proves that the cubic defined by equation (6.13) in the projective plane is isomorphic to the elliptic curve $\mathbb{C}/\Lambda$. In particular, it is a nonsingular curve, and this shows that $\Delta \neq 0$.

6.4 The space of modular forms

Let f be a meromorphic function on $\mathbb{H}$, not identically zero, and let p be a point of $\mathbb{H}$. The integer n such that $\frac{f}{(z-p)^n}$ is holomorphic and non-zero at p is called the *order* of f at p and is denoted by $\nu_p(f)$. When f is a modular function of weight $2k$, the identity (6.1) shows that $\nu_p(f) = \nu_{Gp}(f)$ if G is in the modular group Γ. In other terms, $\nu_p(f)$ depends only on the image of p in $\mathbb{H}/\Gamma$. Moreover one can define $\nu_\infty(f)$ as the order for $q = 0$ of the function $\breve{f}(q)$ associated to f. Finally, we will denote by e_p the order of the stabilizer of the point p; we have that $e_p = 2$ (respectively $e_p = 3$) if p is congruent modulo Γ to i (respectively to $\omega = \frac{-1+\sqrt{3}i}{2}$), and $e_p = 1$ otherwise. Also, we have

Theorem 6.4.1. *Let f be a modular function of weight $2k$, not identically zero. Then*

$$\nu_\infty(f) + \sum_{p\in\mathbb{H}/\Gamma} \frac{1}{e_p}\nu_p(f) = \frac{k}{6}. \tag{6.14}$$

(We can also write this formula in the form

$$\nu_\infty(f) + \frac{1}{2}\nu_p(i) + \frac{1}{3}\nu_p(\omega) + {\sum_{p\in\mathbb{H}/\Gamma}}^{*} \frac{1}{e_p}\nu_p(f) = \frac{k}{6}, \tag{6.15}$$

where the symbol Σ^* means a summation over the points of $\mathbb{H}/\Gamma$ distinct from the classes of i and ω.)

Proof. Observe first that the sum written in Theorem 6.4.1 makes sense, *i. e.*, that f has only a finite number of zeros and poles modulo Γ. Indeed, since $\breve{f}$ is meromorphic, there exists $r > 0$ such that $\breve{f}$ has no zero nor pole for $0 < |q| < r$; and this means that f has no zero nor pole for $\mathrm{Im}(z) > \frac{1}{2\pi}\log\frac{1}{r}$. Now, the part D_r of the fundamental domain D defined by the inequality $\mathrm{Im}(z) < \frac{1}{2\pi}\log\frac{1}{r}$ is compact; since f is meromorphic in $\mathbb{H}$, it has only a finite number of zeros and of poles in D_r, hence our assertion.

To prove Theorem 6.4.1, we will integrate $\frac{1}{2\pi i}\frac{df}{f}$ on the boundary of D. More precisely:

1. Suppose that f has no zero nor pole on the boundary of D except possibly i, ω, and $-\omega^2$. There exists a contour γ as represented in Figure 6.2, whose interior contains a representative of each zero or pole of f not congruent to i or ω. By the residue theorem we have

   $$\frac{1}{2\pi i}\int_\gamma \frac{df}{f} = {\sum_{p\in\mathbb{H}/\Gamma}}^{*}\nu_p(f).$$

   On the other hand

   - the change of variables $q = e^{2\pi i z}$ transforms the arc EA into a circle η centered at $q = 0$, with negative orientation, and not enclosing any zero or pole of $\breve{f}$ except possibly 0. Hence

     $$\frac{1}{2\pi i}\int_E^A \frac{df}{f} = \frac{1}{2\pi i}\int_\eta \frac{df}{f} = -\nu_\infty(f).$$

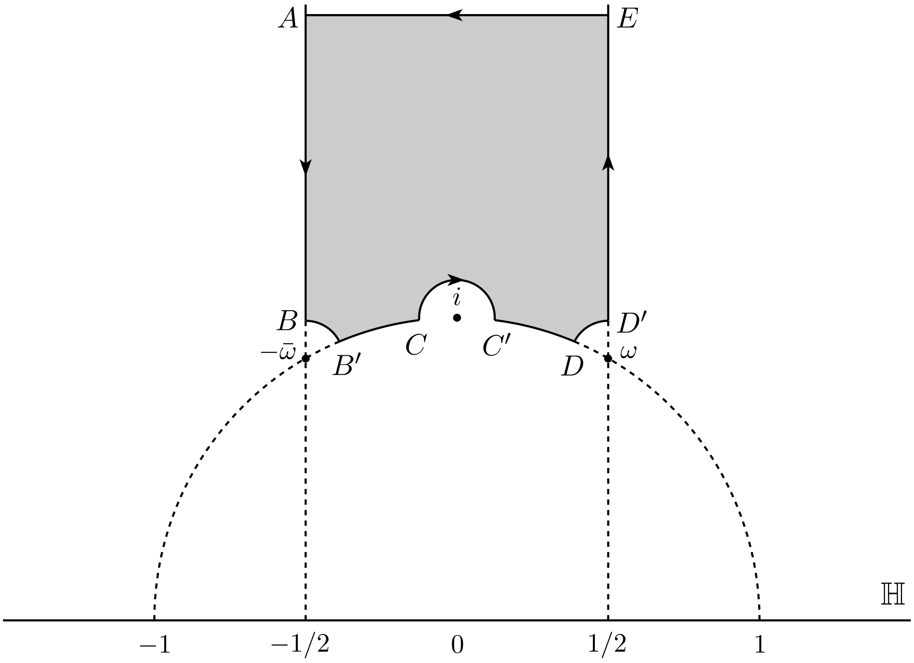

Figure 6.2: Integration on the boundary of region D.

- The integral of $\frac{1}{2\pi i}\frac{df}{f}$ on the circle which contains the arc BB', oriented negatively, has the value $-\nu_\omega(f)$. When the radius of this circle tends to 0, the angle $\angle B_\omega B'$ tends to $\frac{2\pi}{6}$. Hence

$$\frac{1}{2\pi i}\int_B^{B'}\frac{df}{f} \to -\frac{1}{6}\nu_\omega(f).$$

Similarly, when the radii of the arcs CC' and DD' tend to 0:

$$\frac{1}{2\pi i}\int_C^{C'}\frac{df}{f} \to -\frac{1}{2}\nu_i(f), \quad \text{and} \quad \frac{1}{2\pi i}\int_D^{D'}\frac{df}{f} \to -\frac{1}{6}\nu_{-\omega^2}(f).$$

- T transforms the arc AB into the arc ED'; since $f(Tz) = f(z)$, we get:

$$\frac{1}{2\pi i}\int_A^B\frac{df}{f} + \frac{1}{2\pi i}\int_{D'}^E\frac{df}{f} = 0.$$

- S transforms the arc $B'C$ onto the arc DC'; since $f(Sz) = z^{2k}f(z)$, we get:

$$\frac{df(Sz)}{f(Sz)} = 2k\frac{dz}{z} + \frac{df(z)}{f(z)},$$

hence

$$\begin{aligned}\frac{1}{2\pi i}\int_{B'}^{C}\frac{df}{f}+\frac{1}{2\pi i}\int_{C'}^{D}\frac{df}{f} &= \frac{1}{2\pi i}\int_{B'}^{C}\left(\frac{df(z)}{f(z)}-\frac{df(Sz)}{f(Sz)}\right)=\frac{1}{2\pi i}\int_{B'}^{C}-2k\frac{dz}{z}\\ &\to -2k\left(\frac{-1}{12}\right)=\frac{k}{6},\end{aligned}$$

when the radii of the arcs BB', CC', DD' tend to 0. Writing now that the two expressions we get for $\frac{1}{2\pi i}\int_\gamma \frac{df}{f}$ are equal, and passing to the limit, we find formula (6.15).

2. Suppose that f has a zero or a pole λ on the half line $\{z \in \mathbb{H} : \mathrm{Re}(z) = \frac{1}{2}, \mathrm{Im}(z) > \frac{\sqrt{3}}{2}\}$. We repeat the above proof with a contour modified in a neighborhood of λ and of $T\lambda$ as in Figure 6.3 (the arc circling around $T\lambda$ is the transform by T of the arc circling around λ)

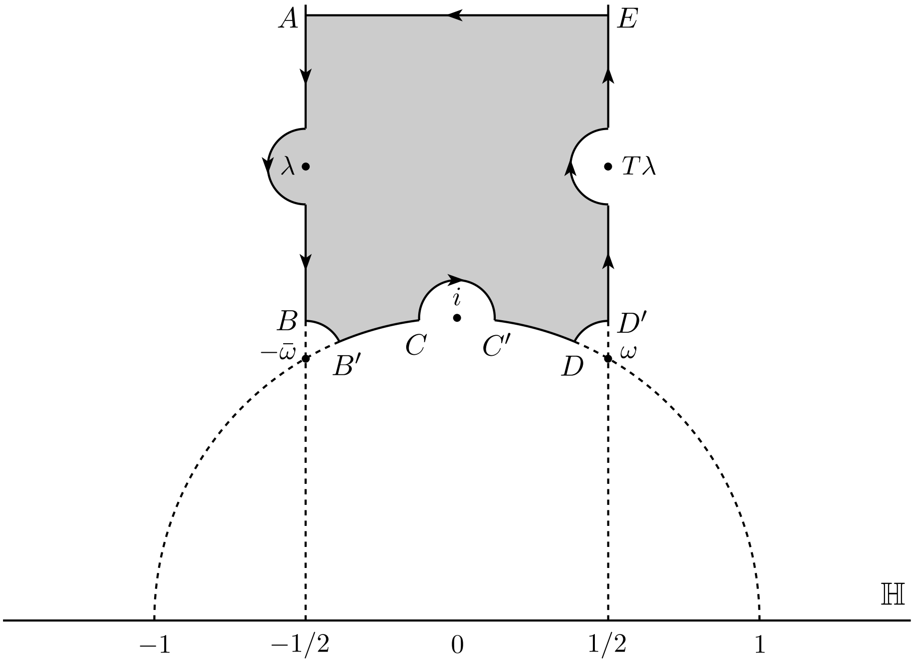

Figure 6.3: Integration on the boundary of region D.

We proceed in an analogous way if f has several zeros or poles on the boundary of D, concluding the proof. □

As a remark, we only mention that in fact this somewhat laborious proof could have been avoided by introducing a complex analytic structure on the compactification $\overline{\mathbb{H}}/\Gamma$ of $\mathbb{H}/\Gamma$, as we mention in Chapter 1 (see for example [20]).

If k is an integer, we denote by M_k (respectively M_k^0) the $\mathbb{C}$-vector space of modular forms of weight $2k$ (respectively of cusp forms of weight $2k$). By Definition 6.2.4, M_k^0 is the kernel of the linear form $f \mapsto f(\infty)$ on M_k. Thus we have $\dim M_k/M_k^0 \leq 1$. Moreover, for $k \geq 2$, the Eisenstein series G_k is an element of M_k such that $G_k(\infty) \neq 0$, by Proposition 6.3.5. Hence we have

$$M_k = M_k^0 \oplus \mathbb{C}\, G_k, \quad \text{for } k \geq 2.$$

Finally, recall that one denotes by Δ the element $g_2^3 - 27g_3^2$ of M_6^0, where $g_2 = 60G_2$ and $g_3 = 140G_3$.

Theorem 6.4.2. *1. We have $M_k = 0$ for $k < 0$ and $k = 1$.*

2. For $k = 0, 2, 3, 4, 5$, M_k is a vector space of dimension 1 with basis $1, G_2, G_3, G_4, G_5$; we have also $M_k^0 = 0$.

3. Multiplication by Δ defines an isomorphism of M_{k-6} onto M_k^0.

Proof. Let f be a nonzero element of M_k. All the terms on the left side of the formula (6.15)

$$\nu_\infty(f) + \frac{1}{2}\nu_p(i) + \frac{1}{3}\nu_p(\omega) + \sum_{p \in \mathbb{H}/\Gamma}{}^{*} \frac{1}{e_p}\nu_p(f) = \frac{k}{6}$$

are ≥ 0. Thus, we have $k \geq 0$ and also $k \neq 1$, since $\frac{1}{6}$ cannot be written in the form $n + \frac{n'}{2} + \frac{n''}{3}$, with $n, n', n'' \geq 0$. This proves (1).

Now apply equation (6.15) to $f = G_2$. We can write $\frac{2}{6}$ in the form $n + \frac{n'}{2} + \frac{n''}{3}$, $n, n', n'' \geq 0$, only $n = 0, n' = 0, n'' = 1$. This shows that $\nu_p(G_2) = 0$, for $p \neq \omega$ (modulo Γ). The same argument applies to G_3 and proves that $\nu_i(G_3) = 1$ and that all the others $\nu_p(G_3)$ are zero. This already shows that Δ is not zero at i, hence is not identically zero. Since the weight of Δ is 12 and $\nu_\infty(\Delta) \geq 1$, formula (6.15) implies that $\nu_p(\Delta) = 0$ for all $p \neq \infty$ and that $\nu_\infty(\Delta) = 1$. In other words, Δ does not vanish on $\mathbb{H}$ and has a simple zero at infinity. If f is an element of M_k^0 and if we set $g = f/\Delta$, it is clear that g is of weight $2k - 12$. Moreover, the formula

$$\nu_p(g) = \nu_p(f) - \nu_p(\Delta) = \begin{cases} \nu_p(f), & \text{if } p \neq \infty; \\ \nu_p(f) - 1, & \text{if } p = \infty; \end{cases}$$

shows that $\nu_p(g) \geq p$ for all p, thus that g belongs to M_{k-6}, which proves (3). Finally, if $k < 5$, we have $k - 6 < 0$ and $M_k^0 = 0$ by (1) and (3); this shows that $\dim M_k \leq 1$. Since $1, G_2, G_3, G_4$ and G_5 are nonzero elements of M_0, M_2, M_3, M_4, M_5, respectively, we have $\dim M_k = 1$ for $k = 0, 2, 3, 4, 5$, which proves (2). □

Corollary 6.4.3. *We have for any $k \geq 0$*

$$\dim M_k = \begin{cases} \lfloor k/6 \rfloor, & \text{if } k \equiv 1 \pmod 6; \\ \lfloor k/6 \rfloor + 1, & \text{if } k \not\equiv 1 \pmod 6. \end{cases} \tag{6.16}$$

Proof. Formula (6.16) is true for $0 \leq k < 6$ by (1) and (2). Moreover, the two expressions increase by one unit when we replace k by $k+6$ by (3). The formula is thus true for all $k \geq 0$. □

Corollary 6.4.4. *The space M_k has for basis the family of monomials $G_2^\alpha G_3^\beta$, with $\alpha, \beta \geq 0$ integers such that $2\alpha + 3\beta = k$.*

Proof. We show first that these monomials generate M_k. This is clear for $k \leq 3$ by (1) and (2). For $k \geq 4$, we argue by induction on k. Choose a pair (γ, δ) of integers ≥ 0 such that $2\gamma + 3\delta = k$ (this is possible for all $k \geq 2$). The modular form $g = G_2^\gamma G_3^\delta$ is not zero at infinity. If $f \in M_k$, there exists $\lambda \in \mathbb{C}$ such that $f - \lambda g$ is a cusp form, hence equal to Δh, with $h \in M_{k-6}$. One then applies the inductive hypothesis to h. It remains to see that the above monomials are linearly independent; if they were not, the function G_2^3/G_3^2 would verify a nontrivial algebraic equation with coefficients in $\mathbb{C}$, thus would be constant, which is absurd because G_2 is zero at ω but not G_3. □

As a remark, Let $M = \bigoplus M_k$ be the graded algebra which is the direct sum of the M_k and let $\varepsilon : \mathbb{C}[x, y] \to M$ be the homomorphism which maps $x \mapsto G_2$ and $y \mapsto G_3$. Corollary 6.4.4 is equivalent to saying that ε is an isomorphism. Hence, one can identify M with the polynomial algebra $\mathbb{C}[G_2, G_3]$.

6.5 The modular invariant j

We define the *j function* as

$$j(z) = 1728\frac{g_2^3(z)}{\Delta(z)} = \frac{1728 g_2^3(z)}{g_2^3(z) - 27 g_3^2(z)}. \tag{6.17}$$

Proposition 6.5.1. *The function j is a modular function of weight 0. Moreover, it is holomorphic on $\mathbb{H}$ and has a simple pole at infinity. It defines by passage to quotient a bijection of $\mathbb{H}/\Gamma$ onto $\mathbb{C}$.*

Proof. First assertion comes from the fact that g_2^3 and Δ are both of weight 12. The holomorphicity of j follows from the fact that $\Delta \neq 0$ on $\mathbb{H}$ and has a simple pole at infinity, while g_2 is nonzero at infinity. To prove the last assertion, one has to show that if $\lambda \in \mathbb{C}$, the modular form $f_\lambda = 1728 g_2^3 - \lambda\Delta$ has a unique zero modulo Γ. To see this, one applies formula (6.15) to f_λ and $k = 6$. The only decompositions of $\frac{k}{6} = 1$ in the form $n + \frac{n'}{2} + \frac{n''}{3}$, with $n, n', n'' \geq 0$ corresponds to

$$(n, n', n'') = (1, 0, 0) \text{ or } (0, 2, 0) \text{ or } (0, 0, 3).$$

This shows that f is zero at one and only one point of $\mathbb{H}/\Gamma$. □

Theorem 6.5.2. *Let f be a meromorphic function on $\mathbb{H}$. The following are equivalent:*

1. *f is a modular function of weight 0;*

2. *f is a quotient of two modular forms of the same weight;*

3. *f is a rational function of j.*

Proof. The implications (3) $\Rightarrow$ (2) $\Rightarrow$ (1) are immediate. We show that (1) $\Rightarrow$ (3). Let f be a modular function. Being free to multiply by a suitable polynomial in j, we can suppose that f is holomorphic on $\mathbb{H}$. Since Δ is zero at infinity, there exists an integer $n \geq 0$ such that $g = \Delta^n f$ is holomorphic at infinity. The function g is then a modular form of weight $12n$. By Corollary 6.4.4, we can write g as a linear combination of the $G_2^\alpha G_3^\beta$, with $2\alpha + 3\beta = 6n$. By linearity, we are reduced to the case $g = G_2^\alpha G_3^\beta$, *i. e.*, $f = G_2^\alpha G_3^\beta / \Delta^n$. But the relation $2\alpha + 3\beta = 6n$ shows that $p = \frac{\alpha}{3}$ and $q = \frac{\beta}{2}$ are integers and one has $f = G_2^{3p} G_3^{2q} / \Delta^{p+q}$. Thus, we only need to see that G_2^3/Δ and G_3^2/Δ are rational functions of j, which is obvious from the definitions of g_2, g_3 and Δ. □

As we stated in Chapter 1, it is possible to define in a natural way a structure of complex analytic manifold on the compactification $\overline{\mathbb{H}}/\Gamma$ of $\mathbb{H}/\Gamma$. Proposition 6.5.1 thus means that j defines an isomorphism of $\overline{\mathbb{H}}/\Gamma$ onto the Riemann sphere $\overline{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. As for Theorem 6.5.2, it amounts to the well known fact that the only meromorphic functions on $\overline{\mathbb{C}}$ are the rational functions.
The coefficient $1728 = 2^6 \cdot 3^3$ has been introduced in order that j has a residue equal to 1 at infinity. More precisely, the series expansion of next section shows that the q-expansion of j (recall equation (1.3)) is

$$j(z) = q^{-1} + 744 + \sum_{n \geq 1} c_n q^n, \tag{6.18}$$

where $q = e^{2\pi i z}$, $z \in \mathbb{H}$. One has $c_1 = 2^2 \cdot 3^3 \cdot 1823 = 196884$, $c_2 = 2^{11} \cdot 5 \cdot 2099 = 21493760$. All c_n are integers (this follows from the definition of j and the q-expansion formulas of g_2, g_3), and they enjoy remarkable divisibility properties, see for example [2] or [38].

6.6 Expansions at infinity

Eisenstein series and the j function are closely related to the Riemann zeta function. Only to give a basic idea, we will present some of this results, omitting most of the proofs. Interested reader can consult [117], [167].
Consider the *Bernoulli numbers B_k*, defined by the series

$$\frac{x}{e^x - 1} = 1 - \frac{x}{2} + \sum_{k \geq 1} (-1)^{k+1} B_k \frac{x^{2k}}{(2k)!}. \tag{6.19}$$

$B_1 = \frac{1}{6}$	$B_2 = \frac{1}{30}$	$B_3 = \frac{1}{42}$	$B_4 = \frac{1}{30}$	$B_5 = \frac{5}{66}$
$B_6 = \frac{691}{2730}$	$B_7 = \frac{7}{6}$	$B_8 = \frac{3617}{510}$	$B_9 = \frac{43867}{798}$	$B_{10} = \frac{283 \cdot 617}{330}$

Table 6.1: First Bernoulli numbers B_k.

There are a lot of properties related to Bernoulli numbers, (see [101]). For example, the B_k give the values of the Riemann zeta function for the positive even integers (and also for the negative odd integers):

Proposition 6.6.1. *If $k \geq 1$ is an integer, then*

$$\zeta(2k) = \frac{2^{2k-1}}{(2k)!} \pi^{2k} B_k. \tag{6.20}$$

Proof. The identity

$$z \cot z = 1 - \sum_{k \geq 1} B_k \frac{2^{2k} z^{2k}}{(2k)!} \tag{6.21}$$

follows from the definition of the B_k by putting $x = 2iz$. Moreover, taking the logarithmic derivative of

$$\sin z = z \prod_{n \geq 1} \Big(1 - \frac{z^2}{n^2 \pi^2}\Big), \tag{6.22}$$

we get:

$$z \cot z = 1 + 2 \sum_{n \geq 1} \frac{z^2}{z^2 - n^2 \pi^2} = 1 - 2 \sum_{n \geq 1} \sum_{k \geq 1} \frac{z^{2k}}{n^{2k} \pi^{2k}}. \tag{6.23}$$

Comparing (6.21) and (6.23), we get (6.20). □

Table 6.2 shows some values of the Riemann ζ function

$\zeta(2) = \frac{\pi^2}{2 \cdot 3}$	$\zeta(4) = \frac{\pi^4}{2 \cdot 3^2 \cdot 5}$	$\zeta(6) = \frac{\pi^6}{3^3 \cdot 5 \cdot 7}$	$\zeta(8) = \frac{\pi^8}{2 \cdot 3^3 \cdot 5^2 \cdot 7}$
$\zeta(10) = \frac{\pi^{10}}{3^5 \cdot 5 \cdot 7 \cdot 11}$	$\zeta(12) = \frac{691 \pi^{12}}{3^6 \cdot 5^3 \cdot 7^2 \cdot 11 \cdot 13}$	$\zeta(14) = \frac{2 \pi^{14}}{3^6 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13}$	$\ldots$

Table 6.2: Some values $\zeta(2k)$ for the Riemann ζ function.

We now give the Taylor expansion of the Eisenstein series $G_k(z)$ with respect to $q = e^{2\pi i z}$. Let us start with the well known formula

$$\pi \cot \pi z = \frac{1}{z} + \sum_{m \geq 1} \Big(\frac{1}{z+m} + \frac{1}{z-m}\Big).$$

We have on the other hand

$$\pi \cot \pi z = \pi \frac{\cos \pi z}{\sin \pi z} = \pi i \frac{q+1}{q-1} = \pi i - \frac{2\pi i}{1-q} = \pi i - 2\pi i \sum_{n \geq 0} q^n,$$

and comparing, we get:

$$\frac{1}{z} + \sum_{m\geq 1}\Big(\frac{1}{z+m} + \frac{1}{z-m}\Big) = \pi i - 2\pi i \sum_{n\geq 0} q^n. \tag{6.24}$$

By successive differentiations of (6.24), we obtain the following formula (valid for $k \geq 2$)

$$\sum_{m\in\mathbb{Z}} \frac{1}{(z+m)^k} = \frac{1}{(k-1)!}(-2\pi i)^k \sum_{n\geq 1} n^{k-1}q^n. \tag{6.25}$$

Denote now by $\sigma_k(n) = \sum_{d|n} d^k$ the sum of k-th powers of positive divisors of n.

Proposition 6.6.2. *For every integer $k \geq 2$, one has:*

$$G_k(z) = 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!}\sum_{n\geq 1}\sigma_{2k-1}(n)q^n. \tag{6.26}$$

Proof. To prove this, we expand $G_k(z) = \sum_{(n,m)}{}' \frac{1}{(nz+m)^{2k}} = 2\zeta(2k) + 2\sum_{n\geq 1}\sum_{m\in\mathbb{Z}}\frac{1}{(nz+m)^{2k}}$. Applying (6.25) with z replaced by nz, we get

$$G_k(z) = 2\zeta(2k) + 2\frac{2(-2\pi i)^{2k}}{(2k-1)!}\sum_{d\geq 1}\sum_{a\geq 1} d^{2k-1}q^{ad} = 2\zeta(2k) + 2\frac{(2\pi i)^{2k}}{(2k-1)!}\sum_{n\geq 1}\sigma_{2k-1}(n)q^n. \ \square$$

Definition 6.6.3. For $k \geq 1$, we define the *Eisenstein series* $E_k(z)$ by

$$E_k(z) = G_k(z)/2\zeta(2k), \tag{6.27}$$

where $G_k(z)$ is the Eisenstein series of index k defined in (6.8).

Corollary 6.6.4. *For $k \geq 2$, one has*

$$E_k(z) = 1 + \gamma_k \sum_{n\geq 1}\sigma_{2k-1}(n)q^n, \tag{6.28}$$

where $\gamma_k = (-1)^k\frac{4k}{B_k}$.

Proof. When taking the quotient of $G_k(z)$ by $2\zeta(2k)$ in equation (6.26), it is clear that $E_k(z)$ is given by (6.28). The coefficient γ_k is computed using Proposition 6.6.1

$$\gamma_k = \frac{(2\pi i)^{2k}}{(2k-1)!}\frac{1}{2\zeta(2k)} = \frac{(2\pi)^{2k}(-1)^k}{(2k-1)!}\frac{(2k)!}{2^{2k-1}\pi^{2k}B_k} = (-1)^k\frac{2k}{B_k}. \ \square$$

$E_2 = 1 + 240 \sum_{n\geq 1} \sigma_3(n) q^n$
$E_3 = 1 - 504 \sum_{n\geq 1} \sigma_5(n) q^n$
$E_4 = 1 + 480 \sum_{n\geq 1} \sigma_7(n) q^n$
$E_5 = 1 - 264 \sum_{n\geq 1} \sigma_9(n) q^n$
$E_6 = 1 + \frac{65520}{691} \sum_{n\geq 1} \sigma_{11}(n) q^n$
$E_7 = 1 - 24 \sum_{n\geq 1} \sigma_{13}(n) q^n$

Table 6.3: Expansion of the first Eisenstein series E_k.

Table 6.3 gives the q-expansion of some Eisenstein series E_k:

Since $g_2 = 60G_2$, $g_3 = 140G_3$, we have

$$g_2 = \frac{60(2\pi^4)}{2\cdot 3^2\cdot 5}E_2 = \frac{(2\pi)^4}{2^2\cdot 3}E_2, \tag{6.29}$$

$$g_3 = \frac{140(2\pi^6)}{3^3\cdot 5\cdot 7}E_3 = \frac{(2\pi)^6}{2^3\cdot 3^3}E_3. \tag{6.30}$$

Recall now that

$$\Delta = g_2^3 - 27g_3^2 = \frac{(2\pi)^{12}}{2^6\cdot 3^3}(E_2^3 - E_3^2) = (2\pi)^{12}\Big(q - 24q^2 + 252q^2 - 1472q^3 - \dots\Big).$$

In fact, we can write a compact form for the last expansion, due to Jacobi. We will not prove this here. The reader can look [117] or [167].

Theorem 6.6.5 (Jacobi)**.**

$$\Delta = (2\pi)^{12} q \prod_{n\geq 1}(1-q^n)^{24}. \tag{6.31}$$

Note that the cusp form $\eta(z) = (2\pi)^{-12}\Delta(z) = q\prod_{n\geq 1}(1-q^n)^{24}$, is the Dedekind function mentioned in Example 1.4.2. Usually $\tau(n)$ denote the coefficient of q^n in the expansion of $\eta(z)$, thus

$$\eta(z) = q\prod_{n\geq 1}(1-q^n)^{24} = \sum_{n\geq 1}\tau(n)q^n.$$

The function $n \mapsto \tau(n)$ is called the *Ramanujan's function*. Observe also that from the definition of j (6.17) and equations (6.29), (6.31) we can rewrite

$$j = 1728\frac{g_2^3}{\Delta} = 1728\left(\frac{(2\pi)^4}{12}\right)^3\frac{E_2^3}{(2\pi)^{12}\eta} = 1728\frac{(2\pi)^{12}}{1728}\frac{E_2^3}{(2\pi)^{12}\eta} = \frac{E_2^3}{\eta},$$

so we obtain the expression (1.3) for j given in Chapter 1.

A useful estimate for the coefficients of modular functions, which shows that the quotient $|a_n|/n^k$ remains bounded when $n \to \infty$, is given in the following result due to Hecke. For a proof see [165, p.94]

Theorem 6.6.6. *If $f = \sum a_n q^n$ is a cusp form of weight $2k$, then $a_n = O(n^k)$.*

Remark 6.6.7. The theory of quasi-modular forms extends the classical theory of modular forms, when it is equipped with the differencial $\partial/\partial z$. See for example [152].

6.7 Theta functions

Let V be a real vector space of finite dimension n endowed with an invariant measure μ. Let V^* be the dual of V. Let f be a rapidly decreasing smooth function on V (see[163]). The Fourier transform $\hat{f}$ of f is defined by the formula

$$\hat{f}(y) = \int_V e^{-2\pi i \langle x,y \rangle} f(x)\, d\mu(x). \tag{6.32}$$

This is a rapidly decreasing smooth function on V^*. Let now Λ be a lattice in V^*. We denote by Λ' the lattice in V^* dual to Λ; that is the set of $y \in V^*$ such that $\langle x, y \rangle \in \mathbb{Z}$, for all $x \in \Lambda$. Observe that Λ' may be identified with the $\mathbb{Z}$-dual of Λ (hence the terminology).

Proposition 6.7.1 (Poisson formula)**.** *Let $\nu = \mu(V/\Lambda)$. One has:*

$$\sum_{x \in \Lambda} f(x) = \frac{1}{\nu} \sum_{y \in \Lambda'} \hat{f}(y).$$

After replacing μ by $\nu^{-1}\mu$, we can assume that $\mu(V/\Lambda) = 1$. By taking a basis $e_1, \ldots, e_n$ of Λ, we identify V with $\mathbb{R}^n$, Λ with $\mathbb{Z}^n$, and μ with the product measure $dx_1 \cdots dx_n$. Thus we have $V^* = \mathbb{R}^n$, $\Lambda' = \mathbb{Z}^n$ and we are reduced to the classical Poisson formula (see [163]) We suppose now that V is endowed with a symmetric bilinear form $\langle x, y \rangle$ which is positive and non degenerate (*i. e.*, $\langle x, x \rangle > 0$ if $x \neq 0$). We identify V with V^* by means of this bilinear form. The lattice Λ becomes thus a lattice in V; and one has $y \in \Lambda$ if and only if $\langle x, y \rangle \in \mathbb{Z}$, for all $x \in \Lambda$. To a lattice Λ, we associate the following function defined on $\mathbb{R}^+$

$$\Theta_\Lambda(t) = \sum_{x \in V} e^{-\pi t \langle x,x \rangle}. \tag{6.33}$$

We choose the invariant measure μ on V such that, if $e_1, \ldots, e_n$ is an orthonormal basis of V, the unit cube defined by the e_i has volume 1. The volume ν of the lattice Λ is then defined by $\nu = \mu(V/\Lambda)$.

Proposition 6.7.2. *We have the identity*

$$\Theta_\Lambda(t) = t^{-n/2}\nu^{-1}\Theta_{\Lambda'}(t^{-1}).$$

Proof. Let $f = e^{-\pi\langle x,x\rangle}$. It is a rapidly decreasing smooth function on V. The Fourier transform $\hat{f}$ of f is equal to f. Indeed, choose an orthonormal basis of V and use this basis to identify V with $\mathbb{R}^n$; the measure μ becomes the measure $dx = dx_1 \cdots dx_n$ and the function f is

$$f = e^{-\pi(x_1^2+\ldots+x_n^2)}.$$

We are thus reduced to showing that the Fourier transform of $e^{-\pi x^2}$ is $e^{-\pi x^2}$, which is well known. We now apply Proposition 6.7.1 to the function f and to the lattice $t^{1/2}\Lambda$; the volume of this lattice is $t^{n/2}v$ and its dual is $t^{-1/2}\Lambda'$; hence we get the formula. □

We can give a matrix interpretation. Let $e_1, \ldots, e_n$ be a basis of Λ. Put $a_{ij} = \langle e_i, e_j\rangle$. The matrix $A = [a_{ij}]$ is positive, non degenerate and symmetric. If $x = \sum x_i e_i$ is an element of V, then

$$\langle x, x\rangle = \sum_{i,j} a_{ij}x_i x_j.$$

The function Θ_Λ can be written as

$$\Theta_\Lambda(t) = \sum_{x_i\in\mathbb{Z}} e^{-\pi t\sum a_{ij}x_i x_j}. \qquad (6.34)$$

The volume ν of Λ is given by $\nu = (\det A)^{1/2}$. This can be seen as follows: Let $\varepsilon_1, \ldots, \varepsilon_n$ be an orthonormal basis of V and put

$$\varepsilon = \varepsilon_1 \wedge \ldots \wedge \varepsilon_n, \quad e = e_1 \wedge \ldots \wedge e_n.$$

We have $e = \lambda\varepsilon$, with $|\lambda| = \nu$. Moreover, $\langle e, e\rangle = \det A\langle\varepsilon, \varepsilon\rangle$, and by comparing, we obtain $\nu^2 = \det A$. Let $B = [b_{ij}]$ be the matrix inverse to A. One checks immediately that the dual basis $\{e_i^*\}$ to $\{e_i\}$ is given by the formulas

$$e_i^* = \sum_j b_{ij}e_j, \quad \text{for } i = 1, \ldots, n.$$

The $\{e_i^*\}$ form a basis of Λ'. The matrix $[\langle e_i^*, e_j^*\rangle]$ is equal to B. This shows in particular that if $\nu' = \mu(V/\Lambda')$, then we have $\nu\nu' = 1$.

We will be interested in pairs (V, Λ) which have the following two properties:

1. The dual Λ' of Λ is equal to Λ.

2. We have $\langle x, x\rangle \equiv 0 \pmod 2$, for all $x \in \Lambda$.

Condition (1) says that one has $\langle x, y\rangle \in \mathbb{Z}$, for all $x, y \in \Lambda$ and that the form $\langle x, y\rangle$ defines an isomorphism of Λ onto its dual. In matrix terms, it means that the matrix $A = [\langle e_i, e_j\rangle]$ has integer coefficients and that its determinant equals 1, or equivalently, to $\nu = 1$. Condition (2) means that the diagonal terms of A are even. We have called these lattices even unimodular (see Section 4.4). Suppose that the pair (V, Λ) satisfies conditions (1) and (2) above, that is, Λ is even unimodular. Let $m \geq 0$ be an integer, and denote by Λ_m the set of elements $x \in \Lambda$ such that $\langle x, x\rangle = 2m$. It can be seen that $|\Lambda_m|$ is bounded by a polynomial in m (a crude volume argument gives for instance $|\Lambda_m| = O(m^{n/2})$). This shows that the series with integer coefficients

$$\theta_\Lambda(q) = \sum_{x\in\Lambda} q^{\langle x,x\rangle/2} = \sum_{m\in\mathbb{Z}} |L_m| q^m, \tag{6.35}$$

defined in equation (4.29), converges for $|q| < 1$. Thus one can define a function $\theta_\Lambda(z)$ on the half-plane $\mathbb{H}$ by the formula (6.35), with $q = e^{2\pi i z}$. The function θ_Λ is called the *theta function* of the quadratic module Λ. It is holomorphic on $\mathbb{H}$.

Theorem 6.7.3. *1. The dimension n of V is divisible by 8.*

2. Also, the function θ_Λ is a modular form of weight $n/2$.

Proof. We prove the identity

$$\theta_\Lambda(-1/z) = (iz)^{n/2}\theta_\Lambda(z). \tag{6.36}$$

Since the two sides are analytic in z, it suffices to prove this formula when $z = it$ with $t > 0$ real. We have

$$\theta_\Lambda(it) = \sum_{x\in\Lambda} e^{-\pi t\langle x,x\rangle} = \Theta_\Lambda(t).$$

Similarly, $\theta_\Lambda(-1/it) = \Theta_\Lambda(t^{-1})$. Formula (6.36) results thus from Proposition 6.7.2, taking into account that $\nu = 1$ and $\Lambda = \Lambda'$.

Now, to prove the first assertion, suppose that n is not divisible by 8; replacing Λ, if necessary, by $\Lambda \oplus \Lambda$ or $\Lambda \oplus \Lambda \oplus \Lambda$, we may suppose that $n \equiv 4 \pmod 8$. Formula (6.36) can then be written

$$\theta_\Lambda(-1/z) = (-1)^{n/4} z^{m/2}\theta_\Lambda(z) = -z^{n/2}\theta_\Lambda(z).$$

If we put $\omega(z) = \theta_\Lambda(z)\, dz^{n/4}$, we see that the differential form ω is transformed into $-\omega$ by $S : z \mapsto -1/z$. Since ω is invariant by $T : z \mapsto z + 1$, we see that ST transforms ω into $-\omega$, which is absurd because $(ST)^3 = 1$.

For (2), since n is divisible by 8, we can rewrite (6.36) in the form

$$\theta_\Lambda(-1/z) = z^{n/2}\theta_\Lambda(z) \tag{6.37}$$

which shows that θ_Λ is a modular form of weight $\frac{n}{2}$. □

Corollary 6.7.4. *There exists a cusp form f_Λ of weight $\frac{n}{2}$ such that $\theta_\Lambda = E_k + f_\Lambda$, where $k = \frac{n}{4}$.*

Proof. This follows from the fact that $\theta_\Lambda(\infty) = 1$, hence that $\theta_\Lambda - E_k$ is a cusp form. □

Corollary 6.7.5. *We have $|\Lambda_m| = \frac{4k}{B_k}\sigma_{2k-1}(m) + O(m^k)$, where $k = \frac{n}{4}$.*

Proof. This follows from Corollary 6.7.4, equation (6.28) and Theorem 6.6.6. □

Note that the 'error term' f_Λ is in general nonzero. However, Siegel has proved that the weighted mean of the f_Λ is zero. More precisely, let C_n be the set of classes (up to isomorphism) of even unimodular lattices Λ, and denote by g_Λ the order of the automorphism group of $\Lambda \in C_n$. One has:

$$\sum_{\Lambda \in C_n} \frac{1}{g_\Lambda} \cdot f_\Lambda = 0, \tag{6.38}$$

or equivalently

$$\sum_{\Lambda \in C_n} \frac{1}{g_\Lambda} \cdot \theta_\Lambda = M_n E_k, \quad \text{where } M_n = \sum_{\Lambda \in C_n} \frac{1}{g_\Lambda}. \tag{6.39}$$

Note that this is also equivalent to saying that the weighted mean of the θ_Λ is an eigenfunction of the Hecke operator $T(n)$ (for a proof see [165]).

Example 6.7.6 (The case $n = 8$)**.** Every cusp form of weight $\frac{n}{2} = 4$ is zero. Corollary 6.7.4 then shows that $\theta_\Lambda = E_2$, in other words, $|\Lambda_m| = 240\sigma_3(m)$ for all integers $m \geq 1$. This applies to the lattice Q_{E_8} constructed in Section 4.6.

Example 6.7.7 (The case $n = 16$)**.** For the same reason as above, we have $\theta_\Lambda = E_4 = 1 + 480\sum_m \sigma_7(m)q^m$. Here one may take $\Lambda = Q_{E_8} \oplus Q_{E_8}$ or $\Lambda = Q_{E_{16}}$; even though these two lattices are not isomorphic, they have the same theta function, *i. e.*, they represent each integer the same number of times. Note that the function θ associated to the lattice $Q_{E_8} \oplus Q_{E_8}$ is the square of the function θ of $Q_{E_{16}}$; we recover thus the identity:

$$\left(1 + 240\sum_{m\geq 1} \sigma_3(m)q^m\right)^2 = 1 + 480\sum_{m\geq 1}\sigma_7(m)q^m.$$

Example 6.7.8 (The case $n = 24$)**.** The space of modular forms of weight 12 is of dimension 2. It has for basis the two functions:

$$E_6 = 1 + \tfrac{65520}{691}\sigma_{11}(m)q^m,$$

and

$$\eta = (2\pi)^{-12}\Delta = q\prod_{m\geq 1}(1-q^m)^{24} = \sum_{m\geq 1}\tau(m)q^m.$$

The theta function associated with the lattice Λ can thus be written $\theta_\Lambda = E_6 + c_\Lambda \eta$, with $c_\Lambda \in \mathbb{Q}$. We have

$$|\Lambda_m| = \tfrac{65520}{691}\sigma_{11}(m) + c_\Lambda \tau(m), \quad \text{for } m \geq 1.$$

The coefficient c_Λ is determined by putting $m = 1$:

$$c_\Lambda = |\Lambda_m| - \tfrac{65520}{691}, \quad \text{for } m \geq 1.$$

For example:

- The Leech lattice Λ_{24} is such that $|(\Lambda_{24})_1| = 0$ (Λ_{24} has no 1-norm elements). Hence, $c_{\Lambda_{24}} = -\frac{65520}{691}$.
- The lattice $Q_{E_8} \oplus Q_{E_8} \oplus Q_{E_8}$ is such that $|(Q_{E_8})_1| = 3240$. Hence, $c_{Q_{E_8}} = \frac{432000}{691}$.

Chapter 7

Conway-Norton fundamental conjecture

This chapter is a preamble to the proof given by Borcherds. In fact, we develop some techniques and tools which appear in the Conway-Norton conjecture. We initially give some background theory on Hecke operators. Then we describe some properties of the congruence subgroups of $SL_2(\mathbb{Z})$ and their normalizers. Finally, following [38], we mention various *moonshine* properties that leaded Conway and Norton to formulate their conjecture, including the replication formulas, useful in the proof of Moonshine conjecture.

7.1 Hecke operators

Definition 7.1.1. Let E be a set and let X_E be the free abelian group generated by E. A *correspondence* on E (with integer coefficients) is a homomorphism $T : X_E \to X_E$. We can give T by its values on the elements x of E:

$$T(x) = \sum_{y \in E} n_y(x) y, \quad \text{where } n_y(x) \in \mathbb{Z}, \tag{7.1}$$

the $n_y(x)$ being zero for almost all y.

Let F be a numerical valued function on E. By $\mathbb{Z}$-linearity, it extends to a function (again denoted F), on X_E. The *transform* of F by T, denoted TF, is the restriction to E of the function $F \circ T$. With the notations of (7.1), we have

$$TF(x) = F(T(x)) = \sum_{y \in E} = n_y(x) F(y). \tag{7.2}$$

Example 7.1.2 (The operators $T(n)$)**.** Let $\mathscr{L}$ be the set of lattices of $\mathbb{C}$ (*c. f.* Section 6.3). Let $n \geq 1$ be an integer. We denote by $T(n)$ the correspondence on $\mathscr{L}$ which transforms a lattice to the sum (in $X_{\mathscr{L}}$) of its sublattices of index n. Thus we have:

$$T(n)\Lambda = \sum_{[\Lambda:\Lambda']} \Lambda', \quad \text{if } \Lambda \in \mathscr{L}. \tag{7.3}$$

The sum on the right-hand side of (7.3) is finite. Indeed, the lattices Λ' all contain $n\Lambda$ and their number is also the number of subgroups of order n of $\Lambda / n\Lambda = (\mathbb{Z}/n\mathbb{Z})^2$. If n is prime, one sees in fact that this number is equal to $n+1$ (number of points of the projective line over a field with n elements). We also use the homothety operators R_λ ($\lambda \in \mathbb{C}^\times$), defined by

$$R_\lambda \Lambda = \lambda\Lambda, \quad \text{if } \Lambda \in \mathscr{L}. \tag{7.4}$$

Definition 7.1.3. The operators $T(n)$ defined above are usually called *Hecke operators.*

It makes sense to compose the correspondences $T(n)$ and R_λ, since they are endomorphisms of the abelian group X.

Proposition 7.1.4. *The correspondences $T(n)$ and R_λ verify the identities*

1. *$R_\lambda R_\mu = R_{\lambda\mu}$, for all $\lambda, \mu \in \mathbb{C}^\times$.*

2. *$R_\lambda T(n) = T(n)R_\lambda$, for all $n \geq 1$ and all $\lambda \in \mathbb{C}^\times$.*

3. *$T(m)T(n) = T(mn)$, if $(m,n) = 1$.*

4. *$T(p^n)T(p) = T(p^{n+1}) + pT(p^{n-1})R_p$, for p prime, $n \geq 1$.*

Proof. Statements (1) and (2) are trivial. Note that (3) is equivalent to the following assertion: Let $m, n \geq 1$ be two relatively prime integers, and let Λ'' be a sublattice of a lattice Λ of index mn; there exists a unique sublattice Λ' of Λ, containing Λ'', such that $[\Lambda : \Lambda'] = n$ and $[\Lambda' : \Lambda''] = m$. This assertion follows itself from the fact that the group Λ/Λ'', which is of order mn, decomposes uniquely into a direct sum of a group of order m and a group of order n (Bezout's theorem). To prove (4), let r be a lattice. Then $T(p^n)T(p)\Lambda$, $T(p^{n+1})r\Lambda$ and $T(p^{n-1})R_p\Lambda$ are linear combinations of lattices contained in Λ and of index p^{n+1} in Λ (note that $R_p\Lambda$ is of index p^2 in Λ). Let Λ'' be such a lattice. In the above linear combinations it appears with coefficients a, b, c, say; we have to show that $a = b + pc$, *i. e.*, that $a = 1 + pc$ since b is equal to 1.
We have two cases: (i) Λ'' is not contained in $p\Lambda$. Then $c = 0$ and a is the number of lattices Λ', intermediate between Λ and Λ'', and of index p in Λ; such a lattice Λ' contains $p\Lambda$. In $\Lambda/p\Lambda$ the image of Λ' is of index p and it contains the image of Λ'', which is of order p (hence also of index p because $\Lambda/p\Lambda$ is of order p^2); hence there is only one Λ' which does the trick. This gives $a = 1$ and the formula $a = 1 + pc$ is valid. (ii) $\Lambda'' \subseteq p\Lambda$. We have $c = 1$; any lattice Λ' of index p in Λ contains $p\Lambda$, thus *a fortiori* Λ''. This gives $a = p + 1$ and $a = 1 + pc$ is again valid. □

Corollary 7.1.5. *For $n > 1$, the $T(pn)$ are polynomials in $T(p)$ and R_p.*

This follows from (4) by induction on n. Moreover,

Corollary 7.1.6. *The algebra generated by the R_λ and the $T(p)$, p prime, is commutative; it contains all the $T(n)$.*

We will study now the action of $T(n)$ on the lattice functions of weight $2k$. Let F be a function on $\mathscr{L}$ of weight $2k$ (recall Section 6.3). By definition

$$R_\lambda F = \lambda^{-2k} F, \quad \text{for all } \lambda \in \mathbb{C}^\times. \tag{7.5}$$

Let n be an integer. Property (2) in Proposition 7.1.4 shows that

$$R_\lambda(T(n)F) = T(n)(R_\lambda F) = \lambda^{-2k} T(n)F,$$

in other words, $T(n)F$ is also of weight $2k$. Assertions (3) and (4) on same proposition give:

$$\begin{aligned} T(m)T(n)F &= T(mn)F, \quad \text{if } (m,n) = 1; && (7.6)\\ T(p)T(p^n)F &= T(p^{n+1})F + p^{1-2k} T(p^{n-1})F, \quad \text{for } p \text{ prime}, \ n \geq 1. && (7.7)\end{aligned}$$

Let Λ be a lattice with basis $\{\omega_1, \omega_2\}$ and let $n \geq 1$ be an integer. The following lemma gives all the sublattices of Λ of index n:

Lemma 7.1.7. *Let S_n be the set of integer matrices $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, with $ad = n$, $a \geq 1$, $0 \leq b < d$. If $\sigma = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ is contained in S_n, let Λ_σ be the sublattice of Λ having for basis*

$$\omega_1' = a\omega_1 + b\omega_2, \quad \omega_2' = d\omega_2.$$

The map $\sigma \mapsto \Lambda_\sigma$ is a bijection of S_n onto the set $\Lambda(n)$ of sublattices of index n in Λ.

Proof. The fact that Λ_σ belongs to $\Lambda(n)$ follows from the fact that $\det \sigma = n$. Conversely let $\Lambda' \in \Lambda(n)$. We put

$$Y_1 = \Lambda/(\Lambda' + \mathbb{Z}\omega_2) \ \text{ and } \ Y_2 = \mathbb{Z}\omega_2/(\Lambda' \cap \mathbb{Z}\omega_2).$$

These are cyclic groups generated respectively by the images of ω_1 and ω_2. Let a and d be their orders. The exact sequence

$$0 \longrightarrow Y_2 \longrightarrow \Lambda/\Lambda' \longrightarrow Y_1 \longrightarrow 0$$

shows that $ad = n$. If $\omega_2' = d\omega_2$, then $\omega_2' \in \Lambda'$. On the other hand, there exists $\omega_1 \in \Lambda'$ such that

$$\omega_1' \equiv a\omega_1 \pmod{\mathbb{Z}\omega_2}.$$

It is clear that ω_1' and ω_2' form a basis of Λ'. Moreover, we can write ω_1' in the form

$$\omega_1' = a\omega_1 + b\omega_2, \quad \text{with } b \in \mathbb{Z},$$

where b is uniquely determined modulo d. If we impose on b the inequality $0 \leq b < d$, this fixes b, and also ω_1'. Thus, we have associated to every $\Lambda' \in \Lambda(n)$ a matrix $\sigma(\Lambda') \in S_n$, and one checks that the maps $\sigma \mapsto \Lambda_\sigma$ and $\Lambda' \mapsto \sigma(\Lambda')$ are inverses to each other; the lemma follows. □

Example 7.1.8. If p is a prime, the elements of S_p are the matrix $\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and the matrices $\left(\begin{smallmatrix} 1 & b \\ 0 & p \end{smallmatrix}\right)$, with $0 \leq b < p$.

7.2 Action of $T(n)$ on modular functions

Let k be an integer, and let f be a weakly modular function of weight $2k$. As we saw in Section 6.3, f corresponds to a function F of weight $2k$ on $\mathscr{L}$, satisfying equation (6.7)

$$F(\Lambda(\omega_1,\omega_2)) = \omega_2^{-2k} f(\omega_1/\omega_2). \tag{7.8}$$

We define $T(n)f$ as the function on $\mathbb{H}$ associated to the function $n^{2k-1}T(n)F$ on $\mathscr{L}$. (Note that the numerical coefficient n^{2k-1} gives formulas 'without denominators' in what follows.) Thus by definition:

$$T(n)f(z) = n^{2k-1}T(n)F(\Lambda(z,1)),$$

or else by Lemma 7.1.7

$$T(n)f(z) = n^{2k-1} \sum_{ad=n,\, 0\le b<d} \frac{1}{d^{2k}} f\Big(\frac{az+b}{d}\Big). \tag{7.9}$$

Proposition 7.2.1. *The function $T(n)f$ is weakly modular of weight $2k$. It is holomorphic on $\mathbb{H}$ if f is. We have:*

$$\begin{aligned} T(m)T(n)f &= T(mn), \quad \textit{if } (m,n)=1, & (7.10)\\ T(p)T(p^n)f &= T(p^{n+1})f + p^{2k-1}T(p^{n-1})f, \quad \textit{if } p \textit{ is prime},\ n\ge 1. & (7.11)\end{aligned}$$

Proof. Formula (7.9) shows that $T(n)f$ is meromorphic on $\mathbb{H}$, thus weakly modular; if in addition f is holomorphic, so is $T(n)f$. Formulas (7.10) and (7.11) follow from formulas (7.6) and (7.7) taking into account the numerical coefficient n^{2k-1} incorporated into the definition of $T(n)f$. □

Now suppose that f is a modular function (of weight 0), with q-expansion of the form

$$f(z) = \sum_{m\in\mathbb{Z}} H_m q^m, \quad q = e^{2\pi i z}.$$

Theorem 7.2.2. *$T(n)f$ is also a modular function with q-expansion*

$$T(n)f(z) = \sum_{m\in\mathbb{Z}} \sum_{s|(m,n)} s^{-1} H_{mn/s^2} q^m.$$

Proof. By definition, we have

$$T(n)f(z) = n^{-1} \sum_{ad=n,\, 0\le b<d} \sum_{\mu\in\mathbb{Z}} H_\mu e^{2\pi i\mu(az+b)/d}.$$

The sum

$$\sum_{0\le b<d} e^{2\pi i\mu b/d} = \begin{cases} 0, & \mu \not\equiv 0 \pmod{d};\\ d, & \mu\equiv 0 \pmod{d};\end{cases}$$

since it is the sum $1^\mu + \zeta_2^\mu + \ldots + \zeta_d^\mu$ over all d-roots of unity. Hence the sum is over multiples of d. Putting $\mu = td$,

$$T(n)f(z) = n^{-1} \sum_{ad=n,\, \mu=td} dH_{td} e^{2\pi i a\mu z/d} = n^{-1} \sum_{ad=n,\, t\in\mathbb{Z}} dH_{td} q^{at}.$$

Collecting powers of t and putting $m = at$, this gives

$$T(n)f(z) = \sum_{m\in\mathbb{Z}} q^m \sum_{a|(m,n)} \left(\tfrac{d}{n}\right) H_{\frac{m}{a}.\frac{n}{a}} = \sum_{m\in\mathbb{Z}} \sum_{a|(m,n)} a^{-1} H_{mn/a^2} q^m.$$

Since f s meromorphic at infinity, there exists an integer $N \geq 0$ such that $c_n = 0$ for all $m \leq -N$. The $c_{md/a}$ are thus zero for all $m \leq -nN$, which shows that $T(n)f$ is also meromorphic at infinity. Since it is weakly modular (by Proposition 7.2.1), it is also a modular function. □

7.3 Congruence subgroups of $SL_2(\mathbb{R})$

We shall describe in this section the genus 0 subgroups $G \subseteq SL_2(\mathbb{R})$ which appear in the Conway-Norton conjecture (see Chapter 1).

Let N be a positive integer and let

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : c \equiv 0 \pmod{N} \right\}.$$

This is called a *congruence subgroup* of $SL_2(\mathbb{Z})$. We shall describe the normalizer

$$\mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(N)),$$

i. e., the largest subgroup of $SL_2(\mathbb{R})$ in which $\Gamma_0(N)$ is normal. In describing this normalizer, the divisors of 24 play a crucial role. Let h be the largest divisor of 24 such that h^2 divides N, and consider the factorization

$$N = hn.$$

Here, n is a positive integer divisible by h. Let T be the subgroup of $SL_2(\mathbb{R})$ given by

$$T = \left\{ \begin{pmatrix} a & b/h \\ cn & d \end{pmatrix} \in SL_2(\mathbb{R}) : a, b, c, d \in \mathbb{Z}, ad - bcn/h = 1 \right\}.$$

It is readily seen that T is a subgroup of $SL_2(\mathbb{R})$ containing $\Gamma_0(N)$ and that $\Gamma_0(N)$ is normal in T. Hence

$$T \subseteq \mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(N)).$$

In fact, T is conjugate to $\Gamma_0(n/h)$ in $SL_2(\mathbb{R})$ since we have

$$\begin{pmatrix} h^{1/2} & 0 \\ 0 & h^{-1/2} \end{pmatrix} \begin{pmatrix} a & b/h \\ cn & d \end{pmatrix} \begin{pmatrix} h^{-1/2} & 0 \\ 0 & h^{1/2} \end{pmatrix} = \begin{pmatrix} a & b \\ cn/h & d \end{pmatrix}.$$

Now, T is not in general the full normalizer of $\Gamma_0(N)$. However, T is normal in this normalizer and the factor group is an elementary abelian 2-group. This can be understood as follows.

Definition 7.3.1. A *Hall divisor* of n/h is a divisor e such that $(e, n/he) = 1$.

The number of Hall divisors of n/h is a power of 2. The Hall divisors form an elementary abelian group $\mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2$ under the composition $e \star f = g$, where

$$g = \frac{e}{(e,f)} \cdot \frac{f}{(e,f)}.$$

For each Hall divisor e of n/h we may describe a coset of T in the normalizer of $\Gamma_0(N)$, which corresponds to e under the above isomorphism. This coset consists of all matrices of the form

$$\begin{pmatrix} ae^{1/2} & (b/h)e^{-1/2} \\ cne^{-1/2} & de^{1/2} \end{pmatrix},$$

for $a, b, c, d \in \mathbb{Z}$ and $ade - bc(n/h)e^{-1} = 1$. Moreover, $\mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(N))$ is the union of these cosets for all Hall divisors e of n/h (see [38]). Thus, the quotient of the normalizer by its normal subgroup T is isomorphic to the group of Hall divisors of n/h. Thus we have

$$\mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(N))/T \cong \mathbb{Z}_2 \times \ldots \times \mathbb{Z}_2,$$

and the number of factors on the right-hand side is the number of distinct prime divisors of n/h. We can now state a more precise form of the Conway-Norton conjecture 1.7.1

Conjecture 7.3.2 (**Moonshine Conjecture**). *(Conway-Norton, 1979). Let g be an element of the Monster group $\mathbb{M}$. Then, the McKay-Thompson series $T_g(z)$ is the normalized Hauptmodul*

$$J_G(z) : \overline{\mathbb{H}}/G \to \mathbb{CP}^1$$

for some genus 0 subgroup G of $SL_2(\mathbb{R})$ satisfying

$$T \subseteq G \subseteq \mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(N)),$$

for some N.

We call a discrete subgroup G of $SL_2(\mathbb{R})$ a subgroup of *moonshine-type* if it contains some congruence subgroup $\Gamma_0(N)$, and obeys

$$\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \in G \ \Rightarrow \ t \in \mathbb{Z}. \tag{7.12}$$

Of course, not all subgroups G satisfying the above condition will have genus 0. Thompson [173] proved that there are only finitely many modular groups of moonshine-type in each genus. Cummins [42] has found all of these of genus 0 and 1. In particular, there are precisely 6,486 genus 0 moonshine-type groups. Exactly 616 of these have Hauptmoduls with rational (in fact integral) coefficients; the remainder have cyclotomic integer coefficients. There are some natural equivalences (for example a Galois action) which collapse this number to 371, 310 of which have integral Hauptmoduls.
We illustrate this situation by considering the special case in which N is prime. So, let $N = p$ be a prime, and then we have $n = p$ and $h = 1$. It follows that $T = \Gamma_0(p)$ and

$$[\mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(p)) : T] = 2.$$

Now, it has been shown by Ogg (see [158]) that $\mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(p))$ has genus 0 if and only if

$$p \in \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 41, 47, 59, 71\},$$

the prime divisors of the order of the Monster group $\mathbb{M}$ (recall Chapter 1).

As mentioned in the introductory chapter, the central structure in the attempt to understand equations (1.7) is an infinite-dimensional $\mathbb{Z}$-graded module for the Monster, $V^\natural = V_{-1} \oplus V_1 \oplus V_2 \oplus \ldots$, with graded dimension $J(z)$ (see (1.8)). If we let ρ_n denote the n-th smallest irreducible $\mathbb{M}$-module, with dimension d_n numbered as in Table 1.2, then the first few subspaces will be $V_0 = \rho_0$, $V_1 = \{0\}$, $V_2 = \rho_0 \oplus \rho_1$, $V_3 = \rho_0 \oplus \rho_1 \oplus \rho_2$, $V_4 = 2\rho_0 \oplus 2\rho_1 \oplus \rho_2 \oplus \rho_3$ and so on. As we know from the representation theory for finite groups, a dimension can (and should) be replaced with the character. This gives us the graded traces (1.9)

$$T_g(z) = \sum_{n \in \mathbb{Z}} \mathrm{tr}(g|V_n) q^n.$$

or McKay-Thompson series for this module V. Of course, $T_e = J$.

We write $c_g(n)$ to be the coefficient of q^n in McKay-Thompson series T_g, that is

$$T_g(z) = q^{-1} + \sum_{n \geq 1} c_g(n) q^n, \quad \text{with } q = e^{2\pi i z}.$$

Conway-Norton conjecture (7.3.2) states that for each element g of the Monster $\mathbb{M}$, the McKay-Thompson series T_g is the Hauptmodul $J_{G_g}(z)$ for a genus 0 group $G_g \subseteq \Gamma$ of moonshine-type (recall (7.12)). These groups each contain $\Gamma_0(N)$ as a normal subgroup, for some N dividing $o(g) \cdot (24, o(g))$, $o(g)$ the order of g, and the quotient group $G_g/\Gamma_0(N)$ have exponent ≤ 2. So for each n the map $g \mapsto c_g(n)$ is a character $\mathrm{tr}(g|V_n)$ of $\mathbb{M}$. Conway and Norton [38] explicitly identify each of the groups G_g. The first 50 coefficients $c_g(n)$ of each T_g are given in [148]. Together with the recursions given in Section 7.5 below, this allows one to effectively compute arbitrarily many coefficients $c_g(n)$ of the Hauptmoduls.

It is also this that uniquely defines $V^\natural$, up to equivalence, as a graded $\mathbb{M}$-module.

There are around 8×10^{53} elements in the Monster, so naively we may expect about 8×10^{53} different Hauptmoduls T_g. However, a character evaluated at g and at any of his conjugates hgh^{-1} will always be equal, so $T_g = T_{hgh^{-1}}$. Hence there can be at most 194 distinct T_g (one for each conjugacy class). All coefficients $c_g(n)$ are integers (as are in fact most entries of the character table of $\mathbb{M}$). This implies that $T_g = T_h$ whenever the cyclic subgroups $\langle g \rangle$ and $\langle h \rangle$ are equal. In fact, the total number of distinct McKay-Thompson series T_g arising in Monstrous Moonshine turns out to be only 171. Of those many redundancies among the T_g, only one is unexpected —and unexplained—: the McKay-Thompson series of two unrelated classes of order 27, namely 27A and 27B (in Atlas notation), are equal. It would be interesting to understand what general phenomenon, if any, is responsible for $T_{27A} = T_{27B}$. But as we know from the theory of vertex algebras, the McKay-Thompson series $T_g(z)$ are actually specialisations of 1-point functions and as such are functions of not only z but of all $\mathbb{M}$-invariant vectors $v \in V^\natural$. What we call $T_g(z)$ is really the specialisation $T_g(z, 1)$ of this function $T_g(z, v)$.

Not all subgroups G of genus 0 satisfying $T \subseteq G \subseteq \mathcal{N}_{SL_2(\mathbb{R})}(\Gamma_0(N))$, for some N, correspond to McKay-Thompson series. If $h = 1$ almost all of them do, but there are three exceptions (*c.f.* [38]). It has recently shown by Conway (see [105]) that, apart from these exceptions when $h = 1$, there is a fairly simple characterization of the groups arising as G_g in Monstrous Moonshine:

Proposition 7.3.3. *A subgroup G of $SL_2(\mathbb{R})$ equals one of the modular groups G_g appearing in Conjecture 7.3.2 if, and only if*

1. *G is genus 0;*
2. *G has the form $\Gamma_0(n/h) + e, f, g, \ldots$;*
3. *the quotient of G by $\Gamma_0(nh)$ is a group of exponent ≤ 2;*
4. *each cusp $\mathbb{Q} \cup \{i\infty\}$ can be mapped to $\{i\infty\}$ by an element of $SL_2(\mathbb{R})$ that conjugates the group to one containing $\Gamma_0(nh)$.*

This is an observation made by examining the possible cases, although the significance of this condition is not yet understood. Also, notation $\Gamma_0(n/h) + e, f, g, \ldots$ in (2) corresponds to the full normaliser $\Gamma_0(N)$ in $PSL_2(\mathbb{R}$, obtained vy adjoinig to the group $\Gamma_0(n/h)$ its Atkin-Lehner involutions. See [38] or [105] for more details.

The subgroup G corresponding to a McKay-Thompson series $T_g(z)$ was conjectured explicitly by Conway and Norton for each $g \in \mathbb{M}$. The subgroup G is specified by giving the integer N and a subset of Hall divisors of n/h. In fact, N arises as the least positive integer such that

$$\begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix} T_g(z) = T_g(z),$$

and n arises as the order of g. Then, n divides N and the quotient $h = N/n$ divides 24. In fact, h^2 divides N. The subgroup G is given by

$$G = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} T_g(z) = \zeta T_g(z) \text{ for some } \zeta \in \mathbb{C} \text{ with } \zeta^h = 1 \right\}.$$

We give a few examples to illustrate the situation.

Example 7.3.4. $\mathbb{M}$ has 5 conjugacy classes of elements of order 10. They all arise from the congruence subgroup with $N = 10$. We have $h = 1$, $n = 10$, $n/h = 10$ and the Hall divisors of n/h form a group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. The five subgroups of this group are:

$$\begin{array}{lll} \{1,2,5,10\}, & \{1,2\}, & \\ \quad\{1,5\}, & \{1,10\}, & \{1\}. \end{array}$$

There are 5 corresponding subgroups G of the normalizer of $\Gamma_0(10)$ giving the 5 conjugacy classes of $\mathbb{M}$ or order 10.

Example 7.3.5. $\mathbb{M}$ has 6 conjugacy classes of elements of order 6. Five of them arise from the congruence subgroup with $N = 6$. In this case, we have $h = 1$, $n = 6$, $n/h = 6$ and the Hall divisors of n/h form a group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$. The five subgroups of this group are:

$$\begin{array}{lll} \{1,2,3,6\}, & \{1,2\}, & \\ \quad\{1,3\}, & \{1,6\}, & \{1\}. \end{array}$$

There are 5 corresponding subgroups G of the normalizer of $\Gamma_0(6)$. The remaining conjugacy class of elements of order 6 arises from the congruence subgroup with $N = 18$. In this case, we have $h = 3$, $n = 6$, $n/h = 2$, The Hall divisors of n/h form a group isomorphic to $\mathbb{Z}_2$. The unit subgroup $\{1\}$ is the one giving the required subgroup G of the normalizer of $\Gamma_0(18)$.

Example 7.3.6. $\mathbb{M}$ has 2 conjugacy classes of elements of order 78. They arise from the congruence subgroup with $N = 78$. In this case we have $h = 1$, $n = 78$, $n/h = 78$ and the Hall divisors of n/h form a group isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$. The two subgroups which we require here are:

$$\begin{array}{r} \{1,2,3,13,6,26,39,78\}, \\ \{1,6,26,39\}. \end{array}$$

These give rise to the two required subgroups G of the normalizer of $\Gamma_0(78)$.

These examples show some of the subtlety of the correspondences between subgroups of $SL_2(\mathbb{R})$ and conjugacy classes of $\mathbb{M}$. The full correspondence can be found in [38, Table 2].

7.4 Replicable functions

A conjecture in [38] that played an important role in proving the main Conway-Norton conjecture involves the *replication formulae*. Conway and Norton initially thought of the Hauptmoduls T_g as being intimately connected with $\mathbb{M}$; if so, then the group structure of $\mathbb{M}$ should somehow directly relate different T_g. Considering the power map $g \mapsto g^n$ leads to the following.

It was well known classically the following

Proposition 7.4.1. *The function $J(z)$ (equivalently, $j(z)$) has the property that*

$$K(z) = J(pz) + J\Big(\tfrac{z}{p}\Big) + J\Big(\tfrac{z+1}{p}\Big) + \ldots + J\Big(\tfrac{z+p-1}{p}\Big) \tag{7.13}$$

is a polynomial in $J(z)$, for any prime p.

Proof. The proof is straightforward, and is based on the principle that the easiest way to construct a function invariant with respect to some group G is by averaging it over the group: $\sum_{g\in G} f(gx)$. Here $f(x)$ is $J(pz)$ and G is the modular group $PSL_2(\mathbb{Z})$, and we average over finitely many cosets rather than infinitely many elements.

First, writing Γ for $PSL_2(\mathbb{Z})$, note that

$$\Gamma \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma \cup \bigcup_{i=0}^{p-1} \begin{pmatrix} 1 & i \\ 0 & p \end{pmatrix} \Gamma = \{A \in GL_2(\mathbb{Z}) : \det A = p\} = S_p. \tag{7.14}$$

Here, we have used the fact that we know all the matrices in the group S_p, according to Example 7.1.8, for p prime.

Now, applying the Hecke operator $T(p)$ to the modular form $J(z)$ (with weight $2k = 0$), by equation (7.9) we have

$$\begin{aligned} T(p)J(z) &= p^{2k-1} \sum_{ad=p,\, 0\le b<d} \frac{1}{d^{2k}} J\Big(\frac{az+b}{d}\Big) = \frac{1}{p} \sum_{G\in S_p} J(Gz) \\ &= \frac{1}{p}\bigg(J(pz) + J\big(\tfrac{z}{p}\big) + J\big(\tfrac{z+1}{p}\big) + \ldots + J\big(\tfrac{z+p-1}{p}\big)\bigg) \\ &= \tfrac{1}{p} K(z). \end{aligned}$$

Thus, $K(z) = pT(p)J(z)$. Since J is a modular form, Proposition 7.2.1 implies that $K(z)$ is also a modular form, hence a rational function on $\frac{Q(J(z))}{P(J(z))}$, by Theorem 6.5.2. Since the only poles of $J(z)$ are at the cusps, the same applies to $K(z)$. This implies that the denominator polynomial $P(J(z))$ must be trivial (recall that $J(\mathbb{H}) = \mathbb{C}$). Thus, $K(z)$ is a polynomial in $J(z)$. □

In particular, because $K(z)$ is a modular invariant (it is constant on orbits of $\mathbb{H}/\Gamma$, it must satisfy $K(z) = K(z+1)$ (see also Proposition 6.2.2). Thus, $K(z)$ must have a q-expansion

$$K(z) = \sum_{n\in\mathbb{Z}} \kappa_n q^n, \quad \text{where } q = e^{2\pi i z}.$$

More generally, the same argument says that

$$\sum_{ad=n,\, 0\leq b<d} J\Big(\frac{az+b}{d}\Big) = Q_n(J(z)), \tag{7.15}$$

is a polynomial in $J(z)$. In fact, Q_n is the unique polynomial for which $-q^{-n} + Q_n(J(z))$ has a q-expansion with only strictly positive powers of q (see Appendix A for more details). For example, $Q_1(x) = x$, $Q_2(x) = x^2 - 2c_1$, $Q_3(x) = x^3 - 3c_1x - 3c_2$, where $J(z) = \sum_n c_n q^n$. In fact, these equations (7.15) can be rewritten into recursions such as $a_4 = \binom{a_1}{2} + a_3$, or can be collected together into the remarkable remarkable identity originally due to Zagier [187], but discovered independently by Borcherds and others[1]:

$$p^{-1} \prod_{m\geq1,\, n\in\mathbb{Z}} (1 - p^m q^n)^{c_{mn}} = J(y) - J(z), \tag{7.16}$$

with $p = e^{2\pi i y}$, $q = e^{2\pi i z}$ and the powers c_{mn} are the coefficients of the q-expansion of the modular function $J(z)$. This is directly used in the proof of the Monstrous Moonshine conjecture.

Conway and Norton conjectured in [38] that these formulae have an analogue for any McKay-Thompson series T_g. In particular, (7.15) becomes

$$\sum_{ad=n,\, 0\leq b<d} T_{g^a}\Big(\frac{az+b}{d}\Big) = Q_{n,g}\big(T_g(z)\big), \tag{7.17}$$

where the $Q_{n,g}$ plays the same role as Q_n plays for J in (7.15). For example

$$\begin{aligned} T_{g^2}(2z) + T_g\Big(\tfrac{z}{2}\Big) + T_g\Big(\tfrac{z+1}{2}\Big) &= T_g(z)^2 - 2c_g(1), \\ T_{g^3}(3z) + T_g\Big(\tfrac{z}{3}\Big) + T_g\Big(\tfrac{z+1}{3}\Big) + T_g\Big(\tfrac{z+2}{3}\Big) &= T_g(z)^3 - 3c_g(1)T_g(z) - 3c_g(2). \end{aligned}$$

These are called the *replication formulae*. Again, these yield recursions like $c_g(4) = (c_g(1)^2 - c_{g^2}(1))/2 + c_g(3)$, or can be collected into the expression

$$p^{-1} \exp\Bigg(-\sum_{k>0} \sum_{m\geq1,\, n\in\mathbb{Z}} c_{g^k}(mn) \frac{p^{mk} q^{nk}}{k} \Bigg) = T_g(y) - T_g(z). \tag{7.18}$$

[1] D. Zagier (personal communication) has pointed that he was not the first one who proved this identity. This kind of relations was already known by many other authors, and also appeared in [38]. Indeed, he proved this formula and other similar relations in [187].

This looks a lot more complicated than (7.16), but we can glimpse the Taylor expansion of $\log(1-p^m q^n)$ there. In fact, taking $g=1$, the identity element of $\mathbb{M}$, equation (7.18) reduces to (7.16). Equation (7.18) is called the *Borcherds' identity*, and we will give its proof on Chapter 8.

Axiomatising (7.17) leads to Conway and Norton's notion of replicable function.

Definition 7.4.2. Let f be any function on $\mathbb{H}$ of the form $f(z) = q^{-1} + \sum_{n\geq 1} a_n q^n$, and write $f^{(1)} = f$ and $a_n^{(1)} = a_n$. Let $X_n = X_n(f)$ the unique monic polynomial of degree n such that the q-expansion of $-q^{-n} + X_n(f(z))$ has only positive powers of q. Use

$$\sum_{ad=n,\, 0\leq b<d} f^{(a)}\Big(\frac{az+b}{d}\Big) = X_n\big(f^{(1)}(z)\big),$$

to define recursively each $f^{(s)}$, for $s \geq 2$. In each $f^{(s)}$ has a q-expansion of the form $f^{(s)}(z) = q^{-1} + \sum\limits_{n\geq 1} a_n^{(s)} q^n$ —that is, no fractional powers of q arise—, then we call f a *replicable* function.

The reader can see Appendix A to get a better idea of these functions. Just as a matter of information, we have the following characterization of replicable functions:

Proposition 7.4.3. *Let f be a function of the form $q^{-1} + \sum_{n\geq 1} a_n q^n$, and define $X_n(f)$ as in Definition 7.4.2. Then, f is replicable if and only if $H_{m,n} = H_{r,s}$ holds whenever $mn = rs$ and $(m,n) = (r,s)$.*

Proof. The proof is not difficult. Taking the expansion form of the functions $X_n(f)$. If f is replicable, with replicates $f^{(s)} = q^{-1} + \sum_{n\geq 1} a_n^{(s)} q^n$, then

$$H_{m,n} = \sum_{s|(m,n)} \frac{1}{s} a^{(s)}_{mn/s^2},$$

(by Theorem 7.2.2), and the $H_{m,n} = H_{r,s}$ property manifests. The converse follows in a similar way. □

Equation (7.17) conjectures that the McKay-Thompson series are replicable. In particular, we have $(T_g)^{(s)} = T_{g^s}$. Cummins and Norton [44] proved that the Hauptmodul of any genus 0 modular group of moonshine-type is replicable, provided its coefficients are rational. Incidentally, if the coefficients a_n are irrational, then Definition 7.4.2 should be modified to include Galois automorphisms (see [40]).

Any function f obeying the replicable functions (7.17) will also obey *modular equations*, *i. e.*, a certain type of 2-variable polinomial identities satisfied by $f(x)$ and $f(nx)$. The simplest examples come from the exponential and cosine functions: note that for any $n \geq 1$, $\exp(nx) = (\exp(x))^n$ and $\cos(nx) = T_n(\cos(x))$, where T_n is a Tchebychev polynomial. It was known clasically that j (and hence J) satisfy a modular equation, for example, the invariant quantity in (7.13). Note that this property of J depends crucially on it being a Hauptmodul. Conversely, does the existence of modular equations force the Hauptmodul property? Unfortunately not; both the exponential and cosine trivially obey modular equations for each n (use Tchebychev polynomials for $\cos(nz)$). However, we have the following remarkable fact [124]: The only functions $f(z) = q^{-1} + a_1 q + a_2 q^2 + \ldots$ which obey modular equations for all n are J and the 'modular fictions' q^{-1} and and $q^{-1} \pm q$ (which are essentially exp, cos, and sin). More generally, Cummings [43] proved the following.

Theorem 7.4.4. *A function $B(q) = q^{-1} + \sum_{n \geq 1} a_n q^n$ which obeys a modular equation for all $n \equiv 1 \pmod N$, will either be of the form $B(q) = q^{-1} + a_1 q$, or will be a Hauptmodul for a modular group of Moonshine-type.*

The converse is also true [43]. This theorem is the desired algebraic interpretation of the genus 0 property. The denominator identity argument in Chapter 8 will tells us that each T_g obeys a modular equation for each $n \equiv 1$ modulo the order of g, so Theorem 7.4.4 then would conclude the proof of Monstrous Moonshine. Moreover, Norton has conjectured

Conjecture 7.4.5. *Any replicable function with rational coefficients is either a Hauptmodul for a genus 0 modular group of moonshine-type, or is one of the 'modular fictions' $f(z) = q^{-1} = e^{-2\pi i z}$, $f(z) = q^{-1} + q = 2\cos(2\pi z)$, $f(z) = q^{-1} - q = -2i\sin(2\pi z)$.*

This conjecture seems difficult and is still open.

7.5 The replication formulae

Let g be and element of the Monster group $\mathbb{M}$, and consider

$$T_g(z) = \sum_{n \in \mathbb{Z}} \mathrm{tr}(g|V_n^\natural) q^n, \quad \text{with } q = e^{2\pi i z},$$

the McKay-Thompson series of g. Let G be the genus 0 subgroup of $SL_2(\mathbb{R})$ associated to the conjugacy class of g by Conway and Norton. Let

$$T'_g(z) : \overline{\mathbb{H}}/G \to \mathbb{CP}^1,$$

be the canonical isomorphism of Riemann surfaces obtained from the action of G on the upper half-plane $\mathbb{H}$. Then, the Moonshine conjecture asserts that $T'_g(z) = T_g(z)$.

Conway and Norton noticed that the coefficients of $T'_g(z)$ satisfy certain recurrence formulas, called *replication formulae*. Let

$$T_g(z)' = q^{-1} + \sum_{n>1} c_n q^n, \quad \text{with } q = e^{2\pi i z}.$$

Then, the replication formulas express certain coefficients c_n in terms of smaller coefficients c_k, related either to g or g^2. In the case $g = 1$, such replication formulas for the coefficients of $J(z)$ had been obtained by Mahler [?]. In the case of arbitrary $g \in \mathbb{M}$, the following replication formulae were proved by Koike [118]. We write $c_g(n)$ to be the coefficient of q^n in $T_g'(z)$, that is

$$T_g(z)' = q^{-1} + \sum_{n>1} c_g(n)q^n, \quad \text{with } q = e^{2\pi iz}.$$

We then have four replication formulas:

$$\begin{aligned}
c_g(4k) &= c_g(2k+1) + \frac{1}{2}c_g(k)^2 - \frac{1}{2}c_{g^2}(k) + \sum_{j=1}^{k-1} c_g(j)c_g(2k-j); &(7.19)\\
c_g(4k+1) &= c_g(2k+3) - c_g(2)c_g(2k) + \frac{1}{2}c_g(2k)^2 + \frac{1}{2}c_{g^2}(2k) + \\
&\quad + \frac{1}{2}c_g(k+1)^2 - \frac{1}{2}c_{g^2}(k+1) + \sum_{j=1}^{k} c_g(j)c_g(2k+2-j) + \\
&\quad + \sum_{j=1}^{k-1} c_{g^2}(j)c_g(4k-4j) + \sum_{j=1}^{2k-1} (-1)^j c_g(j)c_g(4k-j); &(7.20)\\
c_g(4k+2) &= c_g(2k+2) + \sum_{j=1}^{k} c_g(j)c_g(2k+1-j); &(7.21)\\
c_g(4k+3) &= c_g(2k+4) - c_g(2)c_g(2k+1) - \frac{1}{2}c_g(2k+1)^2 + \frac{1}{2}c_{g^2}(2k+1) + \\
&\quad + \sum_{j=1}^{k+1} c_g(j)c_g(2k+3-j) + \sum_{j=1}^{k} c_{g^2}(j)c_g(4k+2-4j) + \\
&\quad + \sum_{j=1}^{2k} (-1)^j c_g(j)c_g(4k+2-j). &(7.22)
\end{aligned}$$

We note in particular that

$$c_g(4) = c_g(3) + \frac{1}{2}c_g(1)^2 - \frac{1}{2}c_{g^2}(1), \qquad (7.23)$$

but the second replication formula with $k = 1$ gives simply

$$c_g(5) = c_g(5).$$

The idea for obtaining such replication formulae is explained in [38, Section 8]. These replication formulae can be used to express $c_g(n)$, for all n, in terms only of $c_h(1)$, $c_h(2)$,

$c_h(3)$, $c_h(5)$, for various elements $h \in \mathbb{M}$, which are powers of g.

This fact gave Borcherds his strategy for proving the Moonshine conjecture. If the coefficients of the McKay-Thompson series $T_g(z)$ could be shown to satisfy the same replication formulae, and if their coefficients $c_g(1)$, $c_g(2)$, $c_g(3)$ and $c_g(5)$ agree with those of $T_g'(z)$, then it would follow that $T_g'(z) = T_g(z)$. In fact, Borcherds was able to obtain such replication formulae for $T_g(z)$ by means of the theory of infinite dimensional Lie algebras, as we shall explain in the next chapter.

Chapter 8

Borchherds' proof of Conway-Norton conjecture

Borcherds' proof of Moonshine conjecture makes use of the properties of a Lie algebra called the *Monster Lie algebra* [10]. This is an example of what is known as generalized Kac-Moody algebras, or Borcherds algebras [9, 11]. We shall first describe the properties of Borcherds algebras and, subsequently concentrate on the Monster Lie algebra.

8.1 Borcherds Lie algebras

Definition 8.1.1. A Lie algebra $\mathfrak{g}$ over $\mathbb{R}$ is called a *Borcherds algebra* if it satisfies the following axioms:
(i) $\mathfrak{g} = \bigoplus_{i \in \mathbb{Z}} \mathfrak{g}_i$ has a $\mathbb{Z}$-grading such that $\dim \mathfrak{g}_i$ is finite, for all $i \neq 0$ ($\dim \mathfrak{g}_0$ not need to be finite).
(ii) There exists a linear map $\omega : \mathfrak{g} \to \mathfrak{g}$ such that

- $\omega^2 = 1$, the identity map on $\mathfrak{g}$;
- $\omega(\mathfrak{g}_i) = \mathfrak{g}_{-i}$, for all $i \in \mathbb{Z}$;
- $\omega = -1$ on $\mathfrak{g}_0$.

(iii) $\mathfrak{g}$ has an invariant bilinear form $\langle \cdot, \cdot \rangle : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$, such that

- $\langle x, y \rangle = 0$, if $x \in \mathfrak{g}_i$, $y \in \mathfrak{g}_j$ and $i + j \neq 0$;
- $\langle \omega x, \omega y \rangle = \langle x, y \rangle$, for all $x, y \in \mathfrak{g}$;
- $-\langle x, \omega x \rangle > 0$, if $x \in \mathfrak{g}_i$, $i \neq 0$, $x \neq 0$.

These axioms imply that $\mathfrak{g}_0$ is abelian and that the scalar product $\langle \cdot, \cdot \rangle_0 : \mathfrak{g} \times \mathfrak{g} \to \mathbb{R}$, defined by

$$\langle x, y \rangle_0 = -\langle x, \omega y \rangle$$

is positive definite on $\mathfrak{g}_i$, for all $i \neq 0$. This $\langle \cdot, \cdot \rangle_0$ is called the *contravariant* bilinear form on $\mathfrak{g}$. We shall now give some examples of Borcherds algebras.

Example 8.1.2. Let $a = [a_{ij}]$, $i, j \in I$, be a symmetric matrix with $a_{ij} \in \mathbb{R}$. The index set I need not necessarily be to finite —we assume it is either finite or countably infinite—. Thus, our matrix a may be an infinite matrix. We assume that this matrix satisfies the conditions

- $a_{ij} \leq 0$, if $i \neq j$;
- if $a_{ii} > 0$, then $2\frac{a_{ij}}{a_{ii}} \in \mathbb{Z}$, for all $j \in I$.

There is a Borcherds algebra $\mathfrak{g}$ associated to the matrix a which is defined by generators and relations as follows. $\mathfrak{g}$ is generated by elements

$$e_i, f_j, h_{ij} \ \text{ for } i, j \in I,$$

subject to the relations

- $[e_i, f_j] = h_{ij}$,
- $[h_{ij}, h_{k\ell}] = 0$,
- $[h_{ij}, e_k] = \delta_{ij} a_{ik} e_k$,
- $[h_{ij}, f_k] = -\delta_{ij} a_{ik} f_k$,
- if $a_{ii} > 0$ and $i \neq j$, then
$$(\operatorname{ad} e_i)^n e_j = 0, \quad (\operatorname{ad} f_i)^n f_j = 0, \quad \text{where } n = 1 - 2\frac{a_{ij}}{a_{ii}},$$
- if $a_{ii} \leq 0$, $a_{jj} \leq 0$ and $a_{ij} = 0$, then
$$[e_i, e_j] = 0, \quad [f_i, f_j] = 0.$$

This Lie algebra $\mathfrak{g}$ can be graded by the condition

$$\deg e_i = n_i, \quad \deg f_i = -n_i,$$

for some $n_i \in \mathbb{Z}^+$. There is an involution $\omega : \mathfrak{g} \to \mathfrak{g}$ satisfying

$$\omega(e_i) = -f_i, \quad \omega(f_i) = -e_i.$$

There is also an invariant bilinear form on $\mathfrak{g}$ uniquely determined by the condition $\langle e_i, f_i \rangle = 1$, for all $i \in I$. We write $h_i = h_{ii}$. Then, $\langle e_i, f_i \rangle = h_i$ and

$$\langle h_i, h_j \rangle = \langle [e_i, f_i], h_j \rangle = \langle e_i, [f_i, h_j] \rangle = \langle e_i, a_{ii} f_i \rangle = a_{ij},$$

for all $i \neq j$. Thus, $\langle h_i, h_j \rangle = a_{ij}$, for all $i, j \in I$.
We therefore see that the Lie algebra $\mathfrak{g}$ satisfies the axioms of a Borcherds algebra (Definition 8.1.1). It is called the *universal Borcherds algebra* associated with the matrix $[a_{ij}]$.

The grading on $\mathfrak{g}$ can be chosen in many ways, depending on the choice of the positive integers n_i.

We next observe that any symmetrisable Kac-Moody algebra over $\mathbb{R}$ (recall Section 4.3) gives rise to a universal Borcherds algebra. For that, let $\mathfrak{g}$ be the Kac-Moody algebra over $\mathbb{R}$ with symmetrisable generalized Cartan matrix $A = [A_{ij}]$. Thus, there exists a diagonal matrix

$$D = \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix},$$

with each $d_i \in \mathbb{Z}^+$, such that the matrix DA is symmetric. Let $a = [a_{ij}]$ be given by

$$a_{ij} = \frac{d_i A_{ij}}{2}, \quad \text{for all } i, j.$$

Then we have $a_{ji} = a_{ij}$ and $a_{ii} = d_i$. Thus $a_{ij} \leq 0$ if $i \neq j$ and a_{ii} is a positive integer. Also,

$$2\frac{a_{ij}}{a_{ii}} = A_{ij} \in \mathbb{Z}.$$

Thus, the symmetric matrix $[a_{ij}]$ satisfies the conditions needed to construct a Borcherds algebra, and the universal Borcherds algebra with symmetric matrix $[a_{ij}]$ coincides with the subalgebra of the Kac-Moody algebra $\mathfrak{g}$ obtained by generators and relations prior to the adjunction of the commutative algebra of outer derivations (see Definition 4.3.3). The difference between a symmetrisable Kac-Moody algebra and a universal Borcherds algebra is that, in a Borcherds algebra:

1. The index set I may be countably infinite rather that finite;

2. The a_{ii}'s may not be positive and need not lie in $\mathbb{Z}$;

3. $2\frac{a_{ij}}{a_{ii}}$ is only assumed to lie in $\mathbb{Z}$ when $a_{ii} > 0$.

The center of a universal Borcherds algebra $\mathfrak{g}$ lies in the abelian subalgebra generated by the elements h_{ij} and contains all h_{ij} with $i \neq j$. In fact, it can be seen that $h_{ij} = 0$ unless the i-th and j-th columns of a are identical. If we factor out an ideal I of $\mathfrak{g}$ which lies in the center, then $\mathfrak{g}/I$ retains the structure of a Borcherds algebra. If we then adjoin to $\mathfrak{g}/I$ an abelian Lie algebra $\mathfrak{a}$ of outer derivations to give the Lie algebra

$$\mathfrak{g}^* = (\mathfrak{g}/I) \cdot \mathfrak{a},$$

where $\mathfrak{a} \subseteq (\mathfrak{g}^*)_0$ and $[e_i, x] \in \mathbb{R}e_i$, $[f_i, x] \in \mathbb{R}f_i$, for all $x \in \mathfrak{a}$, then $\mathfrak{g}^*$ retains the structure of a Borcherds algebra.

The converse is also true. Given any Borcherds algebra $\mathfrak{g}$, there is a unique universal Borcherds algebra $\mathfrak{g}_U$ and a homomorphism $f : \mathfrak{g}_U \to \mathfrak{g}$ (not necessarily unique), such that

- $\operatorname{Ker} f$ lies in the center of $\mathfrak{g}_U$,
- $\operatorname{Im} f$ is an ideal of $\mathfrak{g}$,
- $\mathfrak{g}$ is the semidirect product of $\operatorname{Im} f$ with a commutative Lie algebra of outer derivations lying in the 0-graded component of $\mathfrak{g}$ and preserving all subspaces $\mathbb{R}e_i$ and $\mathbb{R}f_i$.

The homomorphism f preserves the grading, involution and bilinear form.

8.2 The Borcherds character formula

Let $\mathfrak{g}$ be a universal Borcherds algebra. Recall from Section 4.3 that the root lattice Q of $\mathfrak{g}$ is the free abelian group with basis r_i, for $i \in I$, with symmetric bilinear form $Q \times Q \to \mathbb{R}$, given by

$$(r_i, r_j) \mapsto \langle r_i, r_j \rangle = a_{ij}.$$

The basis elements r_i are called the *simple roots*. We have a grading

$$\mathfrak{g} = \bigoplus_{\alpha \in Q} \mathfrak{g}_\alpha,$$

determined by $e_i \in \mathfrak{g}_{r_i}$, $f_i \in \mathfrak{g}_{-r_i}$. We have seen in Section 4.3 that an element $\alpha \in Q$ is called a *root* of $\mathfrak{g}$ if $\alpha \neq 0$ and $\mathfrak{g}_\alpha \neq 0$. The root α is called a *positive root* if α is a sum of simple roots. For any root α, either α or $-\alpha$ is positive. Let $\Phi = \Phi^+ \cup \Phi^-$ be the set of roots of $\mathfrak{g}$. We say that $\alpha \in \Phi$ is *real* if $\langle \alpha, \alpha \rangle > 0$, and *imaginary* if $\langle \alpha, \alpha \rangle \leq 0$.

Remember also from Section 3.5 that the *Weyl group* W of $\mathfrak{g}$ is the group of isometries of the root lattice Q generated by the reflections w_i corresponding to the simple real roots. We have from (3.5)

$$w_i(r_j) = r_j - 2\frac{\langle r_i, r_j \rangle}{\langle r_i, r_i \rangle} r_i = r_j - 2\frac{a_{ij}}{a_{ii}} r_i.$$

We recall that $2\frac{a_{ij}}{a_{ii}} \in \mathbb{Z}$, since $a_{ii} > 0$. Let $\mathfrak{h}$ be the abelian subalgebra of $\mathfrak{g}$ generated by the elements h_{ij}, for all $i, j \in I$. We have a map $Q \to \mathfrak{h}$ under which r_i maps to h_i, which is a homomorphism of abelian groups and preserves the scalar product. However, this map need not be injective.
If $\mathfrak{g}$ is any Borcherds algebra, the root system and Weyl group of $\mathfrak{g}$ is defined to be that of the corresponding universal Borcherds algebra.

We now introduce certain irreducible modules for a Borcherds algebra. Recall from Section 4.3 that if $\mathfrak{g}$ is a finite dimensional simple Lie algebra over $\mathbb{C}$, the irreducible finite dimensional $\mathfrak{g}$-modules are in 1-1 correspondence with dominant integral weights. Remember that a weight $\lambda \in \mathfrak{h}^*$ is dominant and integral if, and only if, $\lambda(h_i) \geq 0$ and $\lambda(h_i) \in \mathbb{Z}$, for all $i \in I$. The weight λ arises as the highest weight of this module, where $\lambda, \mu \in \mathfrak{h}^*$ satisfy

$\lambda \succ \mu$ if and only if $\lambda - \mu$ is a sum of simple roots.
Now these finite dimensional irreducible $\mathfrak{g}$-modules are also in 1-1 correspondence with antidominant integral weights, *i. e.*, weights $\lambda \in \mathfrak{h}^*$ satisfying $\lambda(h_i) \leq 0$ and $\lambda(h_i) \in \mathbb{Z}$, for all $i \in I$. For there is a unique lowest weight for the module, and this is antidominant and integral. In the case of Borcherds algebras, it is most convenient to consider lowest weight modules rather than highest weight modules.

Recall that if $\mathfrak{g}$ is a finite dimensional simple Lie algebra and λ is an antidominant integral weight, the corresponding finite dimensional irreducible lowest weight module M_λ has character given by the Weyl's character formula (4.16)

$$(\operatorname{char} M_\lambda) e^\rho \prod_{\alpha \in \Phi^+} (1 - e^\alpha) = \sum_{w \in W} \epsilon(w) w(e^{\lambda + \rho}),$$

where $\rho = -\sum_i \omega_i$ (this is an alternative way of writing (4.16), making use of the denominator identity (4.17)).
Next, suppose that $\mathfrak{g}$ is a symmetrisable Kac-Moody algebra and $\lambda \in \mathfrak{h}^*$ is a weight satisfying $\lambda(h_i) \leq 0$, $\lambda(h_i) \in \mathbb{Z}$, for all $i \in I$. Then, $\mathfrak{g}$ has a corresponding irreducible lowest module M_λ whose character is given by Kac's character formula (4.18)

$$(\operatorname{char} M_\lambda) e^\rho \prod_{\alpha \in \Phi} (1 - e^\alpha)^{\operatorname{mult} \alpha} = \sum_{w \in W} \epsilon(w) w(e^{\lambda + \rho}), \tag{8.1}$$

where $\rho \in \mathfrak{h}^*$ is any element satisfying $\rho(h_i) = -1$, for all $i \in I$. This time the sum and product may be infinite.

Finally suppose that $\mathfrak{g}$ is a Borcherds algebra. Let $\lambda \in Q \otimes \mathbb{R}$ be a weight satisfying

- $\langle \lambda, r_i \rangle \leq 0$, for all $i \in I$;
- $2 \frac{\langle \lambda, r_i \rangle}{\langle r_i, r_i \rangle} \in \mathbb{Z}$, for all i for which $\langle r_i, r_i \rangle > 0$.

Then, there is a corresponding irreducible lowest weight module M_λ whose character is given by *Borcherds character formula*

$$(\operatorname{char} M_\lambda) e^\rho \prod_{\alpha \in \Phi^+} (1 - e^\alpha)^{\operatorname{mult} \alpha} = \sum_{w \in W} \epsilon(w) w\left(e^{\lambda + \rho} \sum_{\alpha \in Q} \epsilon(\alpha) e^\alpha \right), \tag{8.2}$$

where $\rho \in Q \otimes \mathbb{R}$ satisfies

$$\langle \rho, r_i \rangle = -\tfrac{1}{2} \langle r_i, r_i \rangle, \quad \text{for all } i \text{ with } \langle r_i, r_i \rangle > 0,$$

and $\epsilon(\alpha) = (-1)^k$ if $\alpha \in Q$ is a sum of k orthogonal simple imaginary roots all orthogonal to λ, and $\epsilon(\alpha) = 0$ otherwise. This formula reduces to Kac's character formula (8.1) in the case of a symmetrisable Kac-Moody algebra, since in this case there are no simple imaginary roots and so

$$\sum_{\alpha\in Q}\epsilon(\alpha)e^{\alpha}=1.$$

In the special case $\lambda=0$, the module M_λ is the trivial 1-dimensional module and Borcherds' character formula becomes

$$e^{\rho}\prod_{\alpha\in\Phi^+}(1-e^{\alpha})^{\operatorname{mult}\alpha}=\sum_{w\in W}\epsilon(w)w\biggl(e^{s}\rho\sum_{\alpha\in Q}\epsilon(\alpha)e^{\alpha}\biggr). \tag{8.3}$$

This is called *Borcherds denominator identity.* It generalizes Kac's denominator identity (4.19), which was itself a generalization of Weyl's denominator identity (4.17). As we shall see, Borcherds' denominator identity plays a key role in the proof of the Moonshine conjecture.

8.3 The monster Lie algebra

We now consider an example of a Borcherds algebra, called the Monster Lie algebra [10], which is our second main object of study, and plays an important role in the proof of the Moonshine conjecture.

We start with the Monster vertex algebra $V^\natural$, constructed in Chapter 5. Recall that $V^\natural$ has a conformal vector of central charge 24. We will replace it with a vertex operator algebra of central charge 26. Let Π be the lattice of rank 2 whose symmetric bilinear scalar product $\Pi\times\Pi\to\mathbb{Z}$ is given by

$$\langle b_1,b_1\rangle=0,\quad\langle b_1,b_2\rangle=-1,\quad\langle b_2,b_2\rangle=0,$$

where b_1,b_2 is a basis of Π. Thus we have

$$\langle mb_1+nb_2,m'b_1+n'b_2\rangle=-mn'-m'n.$$

There is a vertex algebra V_Π associated with the lattice Π as in Section 4.5, and V_Π has a conformal vector of central charge 2.

The tensor product $V^\natural\otimes V_\Pi$ also has the structure of a vertex operator algebra. This vertex algebra has a conformal vector

$$\omega\otimes 1+1\otimes\omega_\Pi$$

of central charge 26, where ω and ω_Π are conformal vectors of $V^\natural$ and V_Π, respectively. Both vertex algebras $V^\natural$, V_Π have symmetric bilinear forms, and these define a symmetric bilinear form on $V^\natural \otimes V_\Pi$. We define the subspaces

$$\begin{aligned} P^1 &= \{v \in V^\natural \otimes V_\Pi : L_0(v) = v, L_i(v) = 0 \text{ for } i \geq 1\}; && (8.4)\\ P^0 &= \{v \in V^\natural \otimes V_\Pi : L_0(v) = 0, L_i(v) = 0 \text{ for } i \geq 1\}. && (8.5) \end{aligned}$$

Now recall from Remark 4.1.3 that the quotient $(V^\natural \otimes V_\Pi)/T(V^\natural \otimes V_\Pi)$ has the structure of a Lie algebra. The space $P^1/T(V^\natural \otimes V_\Pi) \cap P^1$ can be identified with a Lie subalgebra of $(V^\natural \otimes V_\Pi)/T(V^\natural \otimes V_\Pi)$. In fact, we have $TP^0 \subseteq P^1$ and

$$TP^0 = T(V^\natural \otimes V_\Pi) \cap P^1.$$

Thus, P^1/DP^0 has the structure of a Lie algebra (see [10]).

The symmetric bilinear form on $V^\natural \otimes V_\Pi$ induces such a form on P^1, and TP^0 lies in the radical of this form. Thus, we obtain a symmetric bilinear form on the Lie algebra P^1/TP^0. Let $\mathfrak{M}$ be the quotient of this Lie algebra by the radical of the bilinear form, that is,

$$\mathfrak{M} = \frac{P^1/TP^0}{\mathrm{Rad}\langle \cdot, \cdot \rangle}.$$

In fact, $\mathfrak{M}$ is itself a Lie algebra.

Definition 8.3.1. The Lie algebra $\mathfrak{M}$ is called the *Monster Lie algebra*.

Now the vertex algebra V_Π has a grading by elements of the lattice Π and this induces gradings of $V^\natural \otimes V_\Pi$, and then of its subquotient $\mathfrak{M}$ by elements of Π. We write

$$\mathfrak{M} = \bigoplus_{m,n \in \mathbb{Z}} \mathfrak{M}_{(m,n)}, \tag{8.6}$$

where (m, n) is the graded component corresponding to $mb_1 + nb_2 \in \Pi$. It was realized by Borcherds that a theorem from string theory, known as the *no-ghost theorem*, applies to this situation.

8.3.1 The no-ghost theorem

By the remarkable importance of applying this theorem in the proof of Moonshine conjecture, we mention a slight version of the no-ghost theorem (used by Borcherds in [12]). The idea of using the no-ghost theorem to prove results about Kac-Moody algebras appeared in Frenkel's paper [65], which also contains a proof of the no-ghost theorem. The original proof of Goddard and Thorn [76] works for the cases we need with only trivial modifications. For convenience we give a sketch of their proof.

Theorem 8.3.2 (The no-ghost theorem)**.** *Suppose that V is a vector space with a nonsingular bilinear form $\langle\cdot,\cdot\rangle$, and suppose that V is acted on by the Virasoro algebra of Definition 4.2.2 in such a way that*

- *the adjoint of L_i is L_{-i},*
- *the central element of the Virasoro algebra acts as multiplication by 24,*
- *any vector of V is a sum of eigenvectors of L_0 with nonnegative integral eigenvalues,*
- *and all the eigenspaces of L_0 are finite dimensional.*

We let V_{i-1} the subspace of V on which L_0 with has eigenvalue i. Assume that V is acted on by a group G which preserves all this structure. We let V_Π be the vertex algebra of the two dimensional even lattice Π (so that V_Π is Π-graded, has a bilinear form $\langle\cdot,\cdot\rangle$, and is acted on by the Virasoro algebra). We let P^1 be the subspace of $V\otimes V_\Pi$ as defined in (8.4), and we let P^1_α be the subspace of P^1 of degree $\alpha\in\Pi$. All these spaces inherit an action of G from the action of G on V and the trivial action of G on V_Π and $\mathbb{R}^2$.
Then, the quotient of P^1_α by the nullspace of its bilinear form is naturally isomorphic, as a G-module with an invariant bilinear form, to

$$\begin{cases} V_{-\langle\alpha,\alpha\rangle/2} & \text{if } \alpha\neq 0 \\ V_0\oplus\mathbb{R}^2 & \text{if } \alpha=0 \end{cases}.$$

The name *'no-ghost theorem'* comes from the fact that in the original statement of the theorem in [76], V was part of the underlying vector space of the vertex algebra of a positive definite lattice, so the inner product on V_{i-1} was positive definite, and thus, P^1_α had no ghosts (*i. e.*, vectors of negative norm) for $\alpha\neq 0$.

We give a sketch of proof taken from [12] and [76]. Fix some nonzero $\alpha\in\Pi$ and some norm 0 vector $w\in\Pi$, with $\langle\alpha,w\rangle\neq 0$. We use the following operators. We have an action of the Virasoro algebra on $V\otimes V_\Pi$ generated by its conformal vector. The operators L_i of the Virasoro algebra satisfy the relations (4.13)

$$[L_i,L_j]=(i-j)L_{i+j}+\tfrac{1}{2}\binom{i+1}{3}\delta_{i+j,0}\,26,$$

and the adjoint of L_i is L_{-i} (the 26 comes from the 24 in (4.13) plus the dimension of Π). We define operators K_i, for $i\in\mathbb{Z}$, by $K_i=v_{i-1}$, where v is the element $e^{-w}_{-2}e^w$ of the vertex algebra of Π, and e^w is an element of the group ring $\mathbb{R}[\Pi]$, corresponding to $w\in\Pi$, and e^{-w} is its inverse. These operators satisfy the relations

$$[L_i,K_j]=-jK_{i+j},\quad [K_i,K_j]=0,$$

since w has norm 0 and the adjoint of K_i is K_{-i}.

We define the following subspaces of $V\otimes V_\Pi$:

- H is the subspace of $V \otimes V_\Pi$, of degree $\alpha \in \Pi$. H^1 is its subspace of vectors h with $L_0(h) = h$.
- P is the subspace of H of all vectors h with $L_i(h) = 0$, for all $i > 0$. $P^1 = H^1 \cap P$.
- S, the space of spurious vectors, is the subspace of H of vectors perpendicular to P. $S^1 = H^1 \cap S$.
- $N = S \cap P$ is the radical of the bilinear form of P, and $N^1 = H^1 \cap N$.
- T, the transverse space, is the subspace of P annihilated by all the operators K_i, for $i > 0$, and $T^1 = H^1 \cap T$.
- K is the space generated by the action of the operators K_i, $i > 0$.
- Ve^α is the subspace $V \otimes e^\alpha$ of H.

We have the following inclusions of subspaces of H:

$$\begin{tikzcd}
S & & P & & K & \\
& N \arrow[ul] \arrow[ur] & & T \arrow[ul] \arrow[ur] & & Ve^{\alpha} \arrow[ul]
\end{tikzcd}$$

and we construct the isomorphism from $V_{-\langle\alpha,\alpha\rangle/2}$ to $P^1/N \cap P^1$ by zigzagging up and down this diagram; more precisely we show that Ve^α and T are both isomorphic to K modulo its nullspace, and then we show that T^1 is isomorphic to P^1 modulo its nullspace $P^1 \cap N$. In fact, the no-ghost theorem follows immediately from the next sequence of lemmas [12]:

Lemma 8.3.3. *If f is a vector of nonzero norm in T, then the vectors of the form*

$$L_{m_1} L_{m_2} \dots K_{n_1} K_{n_2} \dots (f)$$

for all sequences of integers with $0 > m_1 \geq m_2 \geq \dots$, $0 > n_1 \geq n_2 \geq \dots$, are linearly independent and span a space invariant under the operators K_i and L_i on which the bilinear form is nonsingular.

Lemma 8.3.4. *Th bilinear form on T is nonsingular, and K is the direct sum of T and the nullspace of K.*

Lemma 8.3.5. *Ve^α is naturally isomorphic to T.*

Lemma 8.3.6. *The associative algebra generated by the elements L_i, for $i < 0$, is generated by elements mapping S^1 into S.*

Lemma 8.3.7. *P^1 is the direct sum of T^1 and N^1.*

Recall from (8.6) that

$$\mathfrak{M} = \bigoplus_{m,n\in\mathbb{Z}} \mathfrak{M}_{(m,n)},$$

where (m, n) is the graded component corresponding to $mb_1 + nb_2 \in \Pi$. Applying the no-ghost theorem to this vertex algebra, this theorem asserts that for $\alpha \in \Pi$,

$$\mathfrak{M}_\alpha \text{ is isomorphic to } V^\natural_{-\langle\alpha,\alpha\rangle/2}, \quad \text{if } \alpha \neq 0,$$

where $V^\natural_i = \{v \in V^\natural : L_0(v) = (i + 1)v\}$. Let $\alpha = mb_1 + nb_2 \in \Pi$. Since $\langle\alpha, \alpha\rangle = -2mn$, we then have

$$\mathfrak{M}_{(m,n)} \cong V^\natural_{mn}, \quad \text{if } \alpha \neq (0, 0).$$

The no-ghost theorem also asserts in this situation that

$$\mathfrak{M}_{(0,0)} \cong \mathbb{R}^2.$$

The graded components of the monster Lie algebra $\mathfrak{M}$ can therefore be shown in the following table:

					⋮					
	0	0	0	0	0	$V^\natural_4$	$V^\natural_8$	$V^\natural_{12}$	$V^\natural_{16}$	
	0	0	0	0	0	$V^\natural_3$	$V^\natural_6$	$V^\natural_9$	$V^\natural_{12}$	
	0	0	0	0	0	$V^\natural_2$	$V^\natural_4$	$V^\natural_6$	$V^\natural_8$	
	0	0	0	$V^\natural_{-1}$	0	$V^\natural_1$	$V^\natural_2$	$V^\natural_3$	$V^\natural_4$	
$\cdots$	0	0	0	0	$\mathbb{R}^2$	0	0	0	0	$\cdots$
	$V^\natural_4$	$V^\natural_3$	$V^\natural_2$	$V^\natural_1$	0	$V^\natural_{-1}$	0	0	0	
	$V^\natural_8$	$V^\natural_6$	$V^\natural_4$	$V^\natural_2$	0	0	0	0	0	
	$V^\natural_{12}$	$V^\natural_9$	$V^\natural_6$	$V^\natural_3$	0	0	0	0	0	
	$V^\natural_{16}$	$V^\natural_{12}$	$V^\natural_8$	$V^\natural_4$	0	0	0	0	0	
					⋮					

Table 8.1: Graded components of the Monster Lie algebra $\mathfrak{M}$.

The group ring $\mathbb{R}[\Pi]$ of the lattice Π has an involution defined by

$$e^\alpha \mapsto (-1)^{\langle\alpha,\alpha\rangle/2} e^{-\alpha}, \quad \text{for } \alpha \in \Pi.$$

This gives rise to an involution on the vertex algebra $V_\Pi = S(\tilde{\mathfrak{h}}^-) \otimes \mathbb{R}[\Pi]$. This in turn, gives rise to an involution on the vertex algebra $V^\natural \otimes V_\Pi$, which acts trivially on $V^\natural$. This involution acts on the subquotient $\mathfrak{M}$ of $V^\natural \otimes V_\Pi$, giving a map $\omega : \mathfrak{M} \to \mathfrak{M}$ such that

- $\omega^2 = 1$,
- $\omega \mathfrak{M}_{(m,n)} = \mathfrak{M}_{(-m,-n)}$,
- $\omega = -1$ on $\mathfrak{M}_{(0,0)}$,
- $\langle \omega x, \omega y \rangle = \langle x, y \rangle$;

where $\langle \cdot, \cdot \rangle$ is the invariant bilinear form on $\mathfrak{M}$. Moreover, the contravariant form

$$\langle x, y \rangle_0 = -\langle x, \omega y \rangle, \quad \text{for } x, y \in \mathfrak{M},$$

is positive definite on $\mathfrak{M}_{(m,n)}$, for all $(m,n) \neq (0,0)$.

We can give $\mathfrak{M}$ a $\mathbb{Z}$-grading by means of the formula

$$\deg \mathfrak{M}_{(m,n)} = 2m + n.$$

The $\mathbb{Z}$-graded components are:

$\cdots$	-5	-4	-3	-2	-1	0	1	2	3	4	5	$\cdots$
$\cdots$	$V_2^\natural \oplus V_3^\natural$	$V_2^\natural$	$V_1^\natural$	0	$V_{-1}^\natural$	$\mathbb{R}^2$	$V_{-1}^\natural$	0	$V_1^\natural$	$V_2^\natural$	$V_2^\natural \oplus V_3^\natural$	$\cdots$

Table 8.2: $\mathbb{Z}$-graded components of the Monster Lie algebra $\mathfrak{M}$.

Hence, $\mathfrak{M}$ satisfies the axioms for a Borcherds algebra, having the necessary $\mathbb{Z}$-grading, involution, and invariant bilinear form giving rise to a positive definite contravariant form on non-zero graded components.

Let Q be the root lattice of the Borcherds algebra $\mathfrak{M}$, and let $\mathfrak{h}$ be the Cartan subalgebra of $\mathfrak{M}$. Then, $\mathfrak{h} = \mathfrak{M}_{(0,0)} = \mathbb{R}^2$ and we have a map $Q \to \mathfrak{h}$ as in Section 8.1, under which each simple root $r_i \in Q$ maps to $h_i \in \mathfrak{h}$, and preserving the scalar product. The elements of $\mathfrak{h}$ may be written in the form $mb_1 + nb_2$, for $m, n \in \mathbb{Z}$. The elements of $\mathfrak{h}$ which arise as images of simple roots in Q are those with (m,n) equal to

$$(1,-1),\ (1,1),\ (1,2),\ (1,3),\ (1,4),\ \ldots$$

Thus, $\mathfrak{M}$ has infinitely may simple roots (note that they are not linearly independent). Since

$$\langle mb_1 + nb_2, mb_1 + nb_2 \rangle = -2mn,$$

we see that $(1,-1)$ gives a real simple root and that all the other simple roots $(1,n)$, for $n \geq 1$, are imaginary.

We have pointed out in Section 8.1 that the map $Q \to \mathfrak{h}$ need not be injective, and in the present situation it is far from injective. Thus, there can be several simple roots in Q mapping to the same element $b_1 + nb_2$ of $\mathfrak{h}$. The number of simple roots mapping to a given element $b_1 + nb_2$ is called the *multiplicity* of $(1, n)$. This multiplicity is

$$\dim \mathfrak{M}_{(1,n)} = \dim V_n^\natural = c_n, \tag{8.7}$$

where c_n is the coefficient of q^n in the expansion of the normalized Hauptmodul (1.4)

$$J(z) = q^{-1} + \sum_{n \geq 1} c_n q^n, \quad \text{with } q = e^{2\pi i z}.$$

Thus,

- $(1, -1)$ has multiplicity 1,
- $(1, 1)$ has multiplicity 196,884,
- $(1, 2)$ has multiplicity 21,493,760;
- ...

and the sum of the simple root spaces in $\mathfrak{M}$ is isomorphic to the Moonshine module $V^\natural$. Hence, the Monster Lie algebra $\mathfrak{M}$ contains within it the Monster vertex algebra $V^\natural$ as the sum of its root spaces corresponding to the simple roots.

The symmetric matrix $[a_{ij}]$ corresponding to the Borcherds algebra $\mathfrak{M}$ is thus a countable matrix with many repeated rows and columns (see Table 8.3). It has the following form: Since $\mathfrak{M}$ has only one simple real root, its Weyl group W has order 2. Any root of $\mathfrak{M}$ maps to an element $mb_1 + nb_2 \in \mathfrak{h}$ such that $\mathfrak{M}_{(0,0)} \neq 0$ and $(m, n) \neq (0, 0)$. The multiplicity of (m, n) is then

$$\dim \mathfrak{M}_{(m,n)} = \dim V_{mn}^\natural = c_{mn}.$$

8.4 Denominator identities

In this section we will describe Borcherds' denominator identity and twisted denominator identity for the monster Lie algebra $\mathfrak{M}$. First, we obtain the Borcherds' denominator identity for the monster Lie algebra, supposing only we know completely the simple roots of $\mathfrak{M}$.

Remember from Section 8.1 the Borcherds identity (8.3)

$$e^\rho \prod_{\alpha \in \Phi^+} (1 - e^\alpha)^{\text{mult } \alpha} = \sum_{w \in W} \epsilon(w) w \bigg(e^\rho \sum_{\alpha \in Q} \epsilon(\alpha) e^\alpha \bigg),$$

	$(1,-1)$	$(1,1)$	$\cdots$	$(1,1)$	$(1,2)$	$\cdots$	$(1,2)$	$(1,3)$	$\cdots$
$(1,-1)$	2	0	$\cdots$	0	-1	$\cdots$	-1	-2	$\cdots$
$(1,1)$	0	-2	$\cdots$	-2	-3	$\cdots$	-3	-4	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$	
$(1,1)$	0	-2	$\cdots$	-2	-3	$\cdots$	-3	-4	$\cdots$
$(1,2)$	-1	-3	$\cdots$	-3	-4	$\cdots$	-4	-5	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$	
$(1,2)$	-1	-3	$\cdots$	-3	-4	$\cdots$	-4	-5	$\cdots$
$(1,3)$	-2	-4	$\cdots$	-4	-5	$\cdots$	-5	-6	$\cdots$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$	

Table 8.3: Cartan matrix for the monster Lie algebra $\mathfrak{M}$.

where $\rho \in Q \otimes \mathbb{R}$ is any vector satisfying

$$\langle \rho, r_i \rangle = -\tfrac{1}{2}\langle r_i, r_i \rangle, \ \forall i \in I.$$

Consider $Q \to \mathfrak{h}$ the homomorphism described in Section 8.1, mapping simple roots in Q to $\mathfrak{h}$. By abuse of language, we will also call these elements in $\mathfrak{h}$ of roots. In the case of the monster Lie algebra $\mathfrak{M}$, simple roots in $\mathfrak{h}$ are the elements $mb_1 + nb_2 \in \mathfrak{h}$, with $m, n \in \mathbb{Z}$ and $mn > 0$ or $mn = -1$. If there are k roots in Q mapping to the same root in $\mathfrak{h}$, we will say that this root in $\mathfrak{h}$ has multiplicity k. Also, we identify the root $mb_1 + nb_2$ with the pair $(m, n) \in \mathbb{Z}^2$. So, we know that the simple roots of $\mathfrak{M}$ are

$$(1,-1),\ (1,1),\ (1,2),\ (1,3), (1,4),\ \ldots$$

and we could take $\rho = (-1, 0)$, because

$$\begin{aligned} \langle (-1,0), (1,n) \rangle &= n, \\ \langle (1,n), (1,n) \rangle &= -2n, \quad \text{for all } n, \end{aligned}$$

so that $\langle \rho, r_i \rangle = -\frac{1}{2}\langle r_i, r_i \rangle$ for each simple root r_i. Also, we know that root (m, n) has multiplicity exactly c_{mn}.

Let $p = e^{(1,0)}$ and $q = e^{(0,1)}$. We have, $e^{\rho} = e^{-(1,0)} = p^{-1}$, so left-hand side of Borcherds' identity (8.3) is

$$p^{-1} \prod_{m>0,\, n \in \mathbb{Z}} (1 - p^m q^n)^{c_{mn}},$$

(because $e^{(m,n)} = e^{m(1,0)+n(0,1)} = (e^{(1,0)})^m(e^{(0,1)})^n = p^mq^n$). Remember also that for $\alpha \in Q$, $\epsilon(\alpha) = (-1)^k$, when α is the sum of k imaginary orthogonal simple roots, and $\epsilon(\alpha) = 0$ in other case.

For the monster Lie algebra $\mathfrak{M}$, there are no two imaginary orthogonal simple roots, because $\langle(1,m),(1,n)\rangle = -m-n < 0$, for all m, n. Thus, the elements $\alpha \in Q$ contributing to sum $\sum_\alpha \epsilon(\alpha)e^\alpha$ are $\alpha = 0$ with $\epsilon(\alpha) = 1$ and all the imaginary simple roots $(1,n) \in \mathfrak{h}$, $n \in \mathbb{Z}^+$. Since there are precisely c_n of these roots in Q (mapping to $(1,n)$) and all of them have $\epsilon(\alpha) = -1$, we have

$$\sum_{\alpha\in Q} \epsilon(\alpha)e^\alpha = 1 - \sum_{n>0} c_n pq^n. \tag{8.8}$$

Also, $|W| = 2$ and $W = \{1, s\}$, where $s(p) = q$ and $s(q) = p$. Thus, right-hand side of Borcherds' identity is

$$\begin{aligned}
\sum_{w\in W} \epsilon(w)w\Big(e^\rho \sum_{\alpha\in Q}\epsilon(\alpha)e^\alpha\Big) &= \sum_{w\in W}\epsilon(w)w\Big(p^{-1}\Big(1-\sum_{n>0}c_npq^n\Big)\Big) \\
&= p^{-1}\Big(1-\sum_{n>0}c_npq^n\Big) - q^{-1}\Big(1-\sum_{n>0}c_nqp^n\Big) \\
&= \Big(p^{-1}-\sum_{n>0}c_np^n\Big) - \Big(q^{-1}-\sum_{n>0}c_nq^n\Big) \\
&= j(p) - j(q). \qquad (8.9)
\end{aligned}$$

Combining (8.8) and (8.9), the denominator identity for the monster Lie algebra $\mathfrak{M}$ establishes Zagier's identity (7.16)

$$p^{-1}\prod_{m>0,\, n\in\mathbb{Z}} (1-p^mq^n)^{c_{mn}} = j(p) - j(q). \tag{8.10}$$

In fact, this identity were first proved by Borcherds in terms of the $\mathfrak{M}$ structure, and then were used for prove that the simple roots of $\mathfrak{M}$ are $(1,-1)$, $(1,1)$, $(1,2)$, $(1,3)$, ...

8.5 The twisted denominator identity

In order to complete the proof of Moonshine conjecture, we shall need a more general form of identity (8.10), named the twisted denominator identity. To explain it, we will give an outline of proof of Zagier's identity (8.10), and then we will generalize it.

Definition 8.5.1. Let U be a finite-dimensional real vector space, with a graded decomposition $U = \bigoplus_{\alpha \in L} U_\alpha$, for some lattice L. We define the *graded dimension* of U as the element in $\mathbb{R}[L]$ given by

$$\operatorname{gr}\dim U = \sum_{\alpha \in L} (\dim U_\alpha) e^\alpha,$$

where $\mathbb{R}[L]$ is the group algebra of L with basis e^α, for $\alpha \in L$.

Let $\wedge^0 U, \wedge^1 U, \wedge^2 U, \ldots$ be the exterior powers of U, that is

$$\wedge^k U = \{k\text{-linear alternate forms } \omega : U \times \ldots \times U \to \mathbb{R}\},$$

and $\wedge^0 U = \mathbb{R}$. We have a simple formula for the alternate sum

$$\sum_{k \geq 0} (-1)^k \operatorname{gr}\dim \wedge^k U,$$

given by

$$\sum_{k \geq 0} (-1)^k \operatorname{gr}\dim \wedge^k U = \prod_{\alpha \in L} (1 - e^\alpha)^{\dim U_\alpha}. \tag{8.11}$$

Note that the right-hand side can also be written as $\exp\Big(- \sum_{k>0} \frac{1}{k} \sum_{\alpha \in L} (\dim U_\alpha) e^{k\alpha} \Big)$, since

$$\begin{aligned} \exp\Big(\sum_{\alpha \in L} (\dim U_\alpha) \sum_{k>0} -\frac{1}{k} e^{k\alpha} \Big) &= \exp\Big(\sum_{\alpha \in L} (\dim U_\alpha) \log(1 - e^\alpha) \Big) \\ &= \prod_{\alpha \in L} (1 - e^\alpha)^{\dim U_\alpha}. \end{aligned}$$

Suppose now that U is a G-module, for some finite group G, and that G acts on each graded component U_α.

Definition 8.5.2. The *graded character* of U, as the map $\operatorname{gr}\operatorname{char} U : G \to \mathbb{R}[L]$ given by

$$g \longmapsto \sum_\alpha \operatorname{tr}(g|U_\alpha) e^\alpha.$$

Then, the alternate sum $\sum_{k \geq 0} (-1)^k \operatorname{gr}\dim \wedge^k U$ is given by the map

$$g \longmapsto \exp\Big(- \sum_{k>0} \frac{1}{k} \sum_\alpha \operatorname{tr}(g^k|U_\alpha) e^{k\alpha} \Big).$$

Observe that when $g = 1$, this map reduces to the alternate sum in (8.11). If U is an infinite dimensional vector space and $U = \bigoplus_{\alpha \in L} U_\alpha$ is a graded decomposition of U, such that each component U_α is finite dimensional, then the formulas are still valid (changing the sums for infinite sums).

Suppose now that $\mathfrak{g}$ is a Borcherds algebra with triangular decomposition

$$\mathfrak{g} = \mathfrak{n}^+ \oplus \mathfrak{h} \oplus \mathfrak{n}^-,$$

where $\mathfrak{n}^+ = \sum_{\alpha\in\Phi^+} \mathfrak{g}_\alpha$ and $\mathfrak{n}^+ = \sum_{\alpha\in\Phi^-} \mathfrak{g}_\alpha$ (note that $\mathfrak{n}^+$ can be infinite dimensional, but each $\mathfrak{g}_\alpha$ must be finite dimensional). Consider the exterior powers

$$\wedge^0 U,\ \wedge^1 U,\ \wedge^2 U,\ \ldots$$

We have a sequence

$$\ldots \xrightarrow{d_4} \wedge^3\mathfrak{n}^+ \xrightarrow{d_3} \wedge^2\mathfrak{n}^+ \xrightarrow{d_2} \wedge^1\mathfrak{n}^+ \xrightarrow{d_1} \wedge^0\mathfrak{n}^+ \xrightarrow{d_0} 0$$

with homology groups

$$H_0\mathfrak{n}^+ = \frac{\text{Ker } d_0}{\text{Im } d_1},\ H_1\mathfrak{n}^+ = \frac{\text{Ker } d_1}{\text{Im } d_2},\ H_2\mathfrak{n}^+ = \frac{\text{Ker } d_2}{\text{Im } d_3},\ \ldots$$

Observe that $\wedge^k\mathfrak{n}^+$ and $H_k\mathfrak{n}^+$ are graded vector spaces with finite dimensional graded components. By definition of the homology groups $H_k\mathfrak{n}^+$, we have

$$\sum_{k\geq 0}(-1)^k \operatorname{gr}\dim \wedge^k\mathfrak{n}^+ = \sum_{k\geq 0}(-1)^k \operatorname{gr}\dim H_k\mathfrak{n}^+.$$

Garland and Lepowsky [73] proved that for Kac-Moody algebras (and also it is verified for Borcherds algebras), $H_k\mathfrak{n}^+$ can be identified with a subspace of $\wedge^k\mathfrak{n}^+$ as follows:

$$\left(H_k\mathfrak{n}^+\right)_\alpha = \begin{cases} (\wedge_k\mathfrak{n}^+)_\alpha & \text{if } \langle \alpha+\rho, \alpha+\rho\rangle = \langle \rho, \rho\rangle \\ 0 & \text{in other case} \end{cases},$$

and this holds for each α in the root lattice of $\mathfrak{g}$.

Now, we specialize for the case $\mathfrak{g} = \mathfrak{M}$, the monster Lie algebra. The formula (8.11) gives

$$\sum_{k\geq 0}(-1)^k \operatorname{gr}\dim \wedge^k\mathfrak{M}^+ = \prod_{(m,n),\, m>0} (1-p^m q^n)^{c_{mn}},$$

where $p = e^{(1,0)}$, $q = e^{(0,1)}$ and $\mathfrak{M}^+ = \sum_{\alpha\in\Phi^+} \mathfrak{M}_\alpha$.

We also compute the alternate sum $\sum_{k\geq 0}(-1)^k \operatorname{gr}\dim H_k\mathfrak{m}^+$. We have, $\operatorname{gr}\dim \wedge^0\mathfrak{M}^+ = e^0$, *i. e.*, $\wedge^0\mathfrak{M}^+ = \mathbb{R}$ is 1-dimensional with weight e^0. Since $\langle 0+\rho, 0+\rho\rangle = \langle\rho,\rho\rangle$, we also have $\operatorname{gr}\dim H_0\mathfrak{M}^+ = e^0$. Next we have

$$\operatorname{gr}\dim \wedge^1\mathfrak{M}^+ = \operatorname{gr}\dim \mathfrak{M}^+ = \sum_{(m,n),\, m>0} c_{mn}p^m q^n.$$

Now we consider $\operatorname{gr}\dim H_1\mathfrak{M}^+$. We have $\rho = (-1,0)$. Let $\alpha = (m,n)$ be a root with $m > 0$. Then $\langle\alpha+\rho,\alpha+\rho\rangle = -2(m-1)n$ and $\langle\rho,\rho\rangle = 0$, so we get $\langle\alpha+\rho,\alpha+\rho\rangle = \langle\rho,\rho\rangle$ if and only if $m = 1$ or $n = 0$. Since $\mathfrak{M}^+$ has no roots with $n = 0$ and the roots (m,n) with $m = 1$ are precisely the simple roots, we obtain

$$\operatorname{gr}\dim H_1\mathfrak{M}^+ = \sum_{(1,n)} c_n p q^n = p\biggl(\sum_{n\in\mathbb{Z}} c_n q^n\biggr).$$

Now, the weights of $\wedge^2\mathfrak{M}^+$ are sums of two distinct weights in $\mathfrak{M}^+$. Since all weights of $\mathfrak{M}^+$ are of the form (m,n) with $m \geq 1$ then there are no weights (m,n) in $\wedge^2\mathfrak{M}^+$ with $m = 1$. However, there are some weights (m,n) with $n = 0$. These are precisely of the form $(1,-1)+(m-1,1)$, where $m-1 \geq 1$. It follows that

$$\operatorname{gr}\dim H_2\mathfrak{M}^+ = \sum_{m\geq 2} c_{m-1}p^m.$$

Note that for $\wedge^3\mathfrak{M}^+$ the weights are sums of three distinct weights of $\mathfrak{M}^+$. None has the form (m,n) with $m = 1$ or $n = 0$. Then, $H_3\mathfrak{M}^+ = 0$. Similarly, $H_k\mathfrak{M}^+ = 0$ for all $k \geq 3$. Thus we have

$$\begin{aligned}\sum_{k\geq 0}(-1)^k \operatorname{gr}\dim H_k\mathfrak{M}^+ &= e^0 - p\sum_{n\in\mathbb{Z}} c_n q^n + \sum_{m\geq 2} c_{m-1}p^m \\ &= e^0 + p\sum_{m\geq 1} c_m p^m - p\sum_{n\in\mathbb{Z}} c_n q^n \\ &= e^0 + p\biggl(\sum_{m\in\mathbb{Z}} c_m p^m - p^{-1}\biggr) - p\biggl(\sum_{n\in\mathbb{Z}} c_n q^n\biggr) \\ &= p\bigl(j(p) - j(q)\bigr).\end{aligned}$$

Therefore, we have derived the denominator identity (8.10)

$$\prod_{(m,n),\, m>0} (1-p^m q^n)^{c_{mn}} = p\bigl(j(p) - j(q)\bigr).$$

In order to obtain the twisted denominator identity, we consider $\mathfrak{M}^+$ as an $\mathbb{M}$-module for the Monster group $\mathbb{M}$. Then, each root space $(\mathfrak{M}^+)_\alpha$ is an $\mathbb{M}$-module. From that we have

$$\sum_{k\geq 0}(-1)^k \operatorname{gr}\dim \wedge^k \mathfrak{M}^+ = \sum_{k\geq 0}(-1)^k \operatorname{gr}\dim \wedge^k H_k\mathfrak{M}^+,$$

(observe that each subspace $(H_k\mathfrak{M}^+)_\alpha$ is also an $\mathbb{M}$-module). Moreover, the left-hand side $\sum_k(-1)^k \operatorname{gr}\dim \wedge^k\mathfrak{M}^+$ is the map

$$g \longmapsto \exp\bigg(-\sum_{k>0}\frac{1}{k}\sum_{\alpha\in\Phi^+} \operatorname{tr}(g^k|(\mathfrak{M}^+)_\alpha)e^{k\alpha}\bigg), \tag{8.12}$$

for $g \in \mathbb{M}$. If we replace all dimensions for characters in the formulas above involving $H_0\mathfrak{M}^+, H_1\mathfrak{M}^+$ and $H_2\mathfrak{M}^+$, we get that the sum $\sum_k(-1)^k \operatorname{gr}\operatorname{char} H_k\mathfrak{M}^+$ is the map

$$g \longmapsto p\bigg(\sum_{n\in\mathbb{Z}}\operatorname{tr}(g|V_n^\natural)p^n - \sum_{n\in\mathbb{Z}}\operatorname{tr}(g|V_n^\natural)q^n\bigg). \tag{8.13}$$

Comparing (8.12) and (8.13) we deduce that

$$p^{-1}\exp\bigg(-\sum_{k>0}\frac{1}{k}\sum_{(m,n),\, m>0} \operatorname{tr}(g^k|V_{mn}^\natural)p^{mk}q^{nk}\bigg) = \sum_{n\in\mathbb{Z}}\operatorname{tr}(g|V_n^\natural)p^n - \sum_{n\in\mathbb{Z}}\operatorname{tr}(g|V_n^\natural)q^n. \tag{8.14}$$

This is what we have called the *twisted denominator identity* for the monster Lie algebra $\mathfrak{M}$. Observe that if we write $c_g(n) = \operatorname{tr}(g|V_n^\natural)$ for the coefficient of q^n in the graded character $\sum_{n\in\mathbb{Z}}\operatorname{tr}(g|V_n^\natural)q^n$, then equation (8.14) is just

$$p^{-1}\exp\bigg(-\sum_{k>0}\sum_{(m,n),\, m>0}\frac{1}{k}c_{g^k}(mn)p^{mk}q^{nk}\bigg) = \sum_{n\in\mathbb{Z}}c_g(n)p^n - \sum_{n\in\mathbb{Z}}c_g(n)q^n,$$

or simply

$$p^{-1}\exp\bigg(-\sum_{k>0}\sum_{(m,n),\, m>0}\frac{1}{k}c_{g^k}(mn)p^{mk}q^{nk}\bigg) = T_g(y) - T_g(z),$$

where $y = e^{2\pi ip}, z = e^{2\pi iq}$. That is, exactly the form of equation (7.18).

8.6 Replication formulae again, and proof's end

The twisted denominator identity for the monster Lie algebra (8.14) is just what is needed to obtain the replication formulae (7.19)-(7.22). By comparing the coefficients of p^2 and p^4

in this identity and applying some elementary algebra, Borcherds derived the replication formulae. We omit the proof of this fact here, because the abundance of calculations. The complete derivation of the replication formulae is made in Appendix A.

This is almost sufficient for the proof of the Conway-Norton conjecture. In order to complete the proof, it remains to show that the coefficients $c_g(1)$, $c_g(2)$, $c_g(3)$ and $c_g(5)$ of the McKay-Thompson series $T_g(z)$, agree with those of $T'_g(z)$. These coefficients for $T'_g(z)$ were completely obtained by Conway and Norton in [38]. In order to obtain the coefficients for the graded characters $T_g(z)$, it is sufficient to know how the modules V_1, V_2, V_3 and V_5 of the Monster $\mathbb{M}$, with dimensions c_1, c_2, c_3 and c_5, respectively, decompose into irreducible modules. Observe that the only irreducible characters of $\mathbb{M}$ less or equal than $c_5 = \dim V_5$ are

$$\chi_0, \chi_1, \chi_2, \chi_3, \chi_4, \chi_5, \chi_6$$

(see Table 1.2), thus these are the only possible irreducible components of V_1, V_2, V_3 and V_5. Borcherds was able to proof that

$$\begin{aligned}
\dim V_1 &= \chi_0 + \chi_1, \\
\dim V_2 &= \chi_0 + \chi_1 + \chi_2, \\
\dim V_3 &= 2\chi_0 + 2\chi_1 + \chi_2 + \chi_3, \\
\dim V_5 &= 4\chi_0 + 5\chi_1 + 3\chi_2 + 2\chi_3 + \chi_4 + \chi_5 + \chi_6,
\end{aligned} \tag{8.15}$$

where as usual, $\chi_0, \chi_1, \ldots, \chi_6$ are the first 7 irreducible characters of $\mathbb{M}$ (denoted by d_i in Chapter 1). This was proved by finding 7 elements $g_1, g_2, \ldots, g_7$ of $\mathbb{M}$ for which the 7×7-matrix $[\chi_i(g_j)]$ is non-singular and by showing that the above equations (8.15) hold when evaluated at each g_i. They must then hold for all $g \in \mathbb{M}$.

We mention in conclusion, that the proof of the Conway-Norton conjecture for the Monster is by no means the sole achievement of Borcherds' work in [12]. Other sporadic simple groups are also discussed, including the Baby Monster $\mathbb{B}$, the Conway group Co_1, the Fischer group Fi'_{24}, the Harada-Norton group HN, the Held group He, and the Mathieu group M_{12}, and denominator identities for all these groups are also obtained. Thus, the topic of Monstrous Moonshine is by no means confined to the Monster $\mathbb{M}$.

Chapter 9

Concluding Remarks

We give in this chapter a quick sketch of further developments and conjectures. As can be seen, Moonshine is an area where it is much easier to conjecture than to prove.

9.1 Orbifolds

In string theory, the most tractable way to introduce singularities is by quotienting ('gauging') by a finite group. This construction plays a fundamental role for CFT and vertex operator algebras; it is the physics underlying what Norton calls *generalized Moonshine*. This is where finite group theory touches CFT. Let M be a manifold and G a finite group of symmetries of M. The set M/G of G-orbits inherits a topology from M, and forms a manifold-like space called an *orbifold*. Fixed points become conical singularities. For example, $\{\pm 1\}$ acts on $M = \mathbb{R}$ by multiplication. The orbifold $\mathbb{R}/\{\pm 1\}$ can be identified with the interval $[0, \infty)$. The fixed point at $x = 0$ becomes a singular point on the orbifold, that is, a point where locally the orbifold does not look like some open n-ball. Orbifolds were introduced into geometry in the 1950's as spaces with certain kind of singularities. They were introduced into string theory in [49], which greatly increased the class of background space-times in which the string could live and still be amenable to calculation. This section briefly sketches the corresponding construction for CFT; our purpose is to motivate some generalization of Moonshine conjecture.

About a third of the McKay-Thompson series T_g have some negative coefficients. We shall see Borcherds interpret them as dimensions of superspaces (which come with signs). In the important announcement [155], on a par with [38], Norton proposed that, although $T_g(-1/z)$ will not usually be another McKay-Thompson series, it will always have non-negative integer q-coefficients, and these can be interpreted as ordinary dimensions. In the process, he extended the $g \mapsto T_g$ assignment to commuting pairs $(g, h) \in \mathbb{M} \times \mathbb{M}$. In particular

Conjecture 9.1.1 (Norton). *To each such pair $g, h \in \mathbb{M}$, with $gh = hg$, we have a function $N_{(g,h)}(z)$, such that*

$$N_{(g^a h^c, g^b h^d)}(z) = \alpha N\Big(g, h; \tfrac{az+b}{cz+d}\Big), \quad \textit{for all} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}), \tag{9.1}$$

for some root of unity α (of order dividing 24, and depending on g, h, a, b, c, d). $N_{(g,h)}(z)$ is either constant, or generates the modular functions for a genus 0 subgroup of $SL_2(\mathbb{Z})$ containing some $\Gamma(N)$ (but otherwise not necessarily of moonshine-type). Constant $N_{(g,h)}(z)$ arise when all elements of the form $g^a h^b$ (with $(a, b) = 1$) are 'non-Fricke'. Each $N_{(g,h)}(z)$ has a $q^{1/N}$-expansion for that N; the coefficients of this expansion (not necessarily integers) are characters evaluated at h of some central extension of the centralizer $C_{\mathbb{M}}(g)$. Simultaneous conjugation of g, h leaves the Norton series unchanged: $N_{(aga^{-1},aha^{-1})}(z) = N_{(g,h)}(z)$.

We call $N_{(g,h)}(z)$ the *Norton series* associated to (g, h). An element $g \in \mathbb{M}$ is called *Fricke* if the group G_g contains an element sending 0 to $i\infty$ —the identity 1 is Fricke, as are 120 of the 171 G_g—. For example, when $\langle g, h\rangle \cong C_2 \times C_2$ and g, h, gh are all in class 2A, then $N_{(g,h)}(z) = q^{-1/2} - 492q^{1/2} - 22590q^{3/2} + \ldots$, while $N_{(g,g)}(z) = q^{-1/2} + 4372q^{1/2} - 96256q + \ldots$. The McKay-Thompson series are recovered by taking $g = 1$: $N_{(1,g)} = T_g$. This action (9.1) of $SL_2(\mathbb{Z})$ is related to its natural action on the fundamental group $\mathbb{Z}^2$ of the torus, as well as a natural action of the braid group, as we shall see in the next section. Norton arrived at his conjecture empirically, by studying the data of Queen (Section 9.3).
The basic tool we have for approaching Moonshine conjectures is the theory of vertex operator algebras, so we need to understand Norton's suggestion from that point of view. For reasons of space, we limit our discussion to $V^\natural$, but it generalizes. Given any automorphism $g \in \mathrm{Aut}(V^\natural)$, we can define g-twisted modules in a straightforward way [51]. Then for each $g \in \mathbb{M}$, there is a unique g-twisted module, call it $V^\natural(g)$,. More generally, given any automorphism $h \in \mathrm{Aut}(V^\natural)$ commuting with g, h will yield an automorphism of $V^\natural(g)$, so we can perform Thompson's graded characters (1.9) and define

$$\mathcal{Z}_{(g,h)}(z) = q^{-c/24}\,\mathrm{tr}_{V^\natural(g)}\, hq^{L_0}. \tag{9.2}$$

These $\mathcal{Z}_{(g,h)}$ can be thought of as the building blocks of the graded dimensions of various eigenspaces in $V^\natural(g)$; for example if h has order m, then the subspace of $V^\natural(g)$ fixed by automorphism h will have graded dimension $m^{-1}\sum_{i=1}^m \mathcal{Z}_{(g,h^i)}$. In the case of the Monster considered here, we have $\mathcal{Z}_{(g,h)} = N_{(g,h)}$.

At the level of algebra, this orbifold theory is analogous to the construction of twisted affine algebras from nontwisted ones. At the level of modular forms, it involves twists and shifts much like how $\theta_4(z) = \sum(-1)^n q^{n^2/2}$ and $\theta_2(z) = \sum q^{(n+1/2)^2}/2$ are obtained from $\theta_3(z)$. Far from an esoteric technical development, orbifolds are central to the whole theory, and a crucial aspect of Moonshine. The important paper [51] proves that, whenever the subgroup $\langle g, h\rangle$ generated by g and h is cyclic, then $N_{(g,h)}$ will be a Hauptmodul satisfying (9.1). One way this will happen of course is whenever the orders of g and h are coprime. Extending [51] to all commuting pairs (g, h) is one of the most pressing tasks in Moonshine. At least some aspects of orbifolds are more tractable in the subfactor framework (see for example [55], so further investigations in that direction should be fruitful. This orbifold construction is the same as was used to construct $V^\natural$ from $V_{\Lambda_{24}}$; $V^\natural$ is the sum of the ι-invariant subspace $V^\natural_+$ of $V_{\Lambda_{24}}$ with the ι-invariant subspace $V^\natural_-$ of the unique -1-twisted

module for $V_{\Lambda_{24}}$, where $\iota \in \mathrm{Aut}(\Lambda_{24})$ is some involution. The graded dimensions of $V^\natural_\pm$ are $2^{-1}(Z(\pm 1, 1) + Z(\pm 1, \iota))$, respectively, and these sum to J.
The orbifold construction is also involved in an interesting reformulation of the Hauptmodul property, due to Tuite [177]. Assume the following

Conjecture 9.1.2 (Uniqueness of $V^\natural$). *$V^\natural$ is the only vertex operator algebra with graded dimension J.*

Tuite argues from this that, for each $g \in \mathbb{M}$, T_g will be a Hauptmodul if and only if the only orbifolds of $V^\natural$ are $V_{\Lambda_{24}}$ and $V^\natural$ itself. In [104], this analysis is extended to some of Norton's series $N_{(g,h)}$, where the subgroup $\langle g, h\rangle$ is not cyclic (thus going beyond [51]), although again assuming the uniqueness conjecture. Recently [24], [25] (and other is preparation), Carnahan has outlined an approach to the generalized mooshine conjecture by using Borcherds' products.

9.2 Why the Monster?

In the work [189], Zhu introduced a particular algebra with special features that provided some results about the modularity of some functions arising on vertex operator algebras. The fact that $\mathbb{M}$ is associated with modular functions can be explained by it being the automorphism group of the Moonshine vertex operator algebra $V^\natural$ and the following theorem

Theorem 9.2.1 (Zhu's Theorem). *Suppose V is a C_2-cofinite weakly rational VOA (see [189] for the definitions), and let $\Phi(V)$ be the finite set of irreducible V-modules. Then, there is a representation π of $SL_2(\mathbb{Z})$ by complex matrices $\pi(A)$ indexed by V-modules $M, N \in \Phi(V)$, such that the one-point functions*

$$\chi_M(z, v) = \mathrm{tr}_M\, o(v) q^{L_0 - c/24} = q^{c/24} \sum_{n \geq 0} \mathrm{tr}_{M_{h+n}}\, o(v) q^{h+n},$$

obey

$$\chi_M\Big(\frac{az+b}{cz+d}, v\Big) = (cz+d)^n \sum_{N \in \Phi(V)} \pi \begin{pmatrix} a & b \\ c & d \end{pmatrix}_{MN} \chi_N(z, v),$$

for any $v \in V$ obeying $L_0 v = nv$, for some $n \in \mathbb{N}$.

What is so special about this group $\mathbb{M}$ that these modular functions T_g and $N_{(g,h)}$ should be Hauptmoduls? This is still open. One approach is due to Norton, and was first stated in [155]: the Monster is probably the largest (in a sense) group with the 6-transposition property. Recall that a k-transposition group G is one generated by a conjugacy class K of involutions, where the product gh of any two elements of K has order at most k. For example, taking K to be the transpositions in the symmetric group $G = S_n$, we find that S_n is 3-transposition.

A transitive action of $\Gamma = PSL_2(\mathbb{Z})$ on a finite set X with one distinguished point $x_0 \in X$, is equivalent to specifying a finite index subgroup Γ_0 of Γ. In particular, Γ_0 is the stabilizer $\{g \in \Gamma : g{\cdot}x_0 = x_0\}$ of x_0, X can be identified with the cosets $\Gamma_0\backslash\Gamma$, and x_0 with the coset Γ_0. To such an action, we can associate an interesting triangulation of the closed surface $\Gamma_0\backslash\overline{\mathbb{H}}$, called a (modular) *quilt*. The definition, originally due to Norton and further developed by Parker, Conway, and Hsu, is somewhat involved and will be avoided here (but you can see [98]). It is so-named because there is a polygonal 'patch' covering every cusp of $\Gamma_0\backslash\overline{\mathbb{H}}$, and the closed surface is formed by sewing together the patches along their edges 'seams' (Figure 9.1).

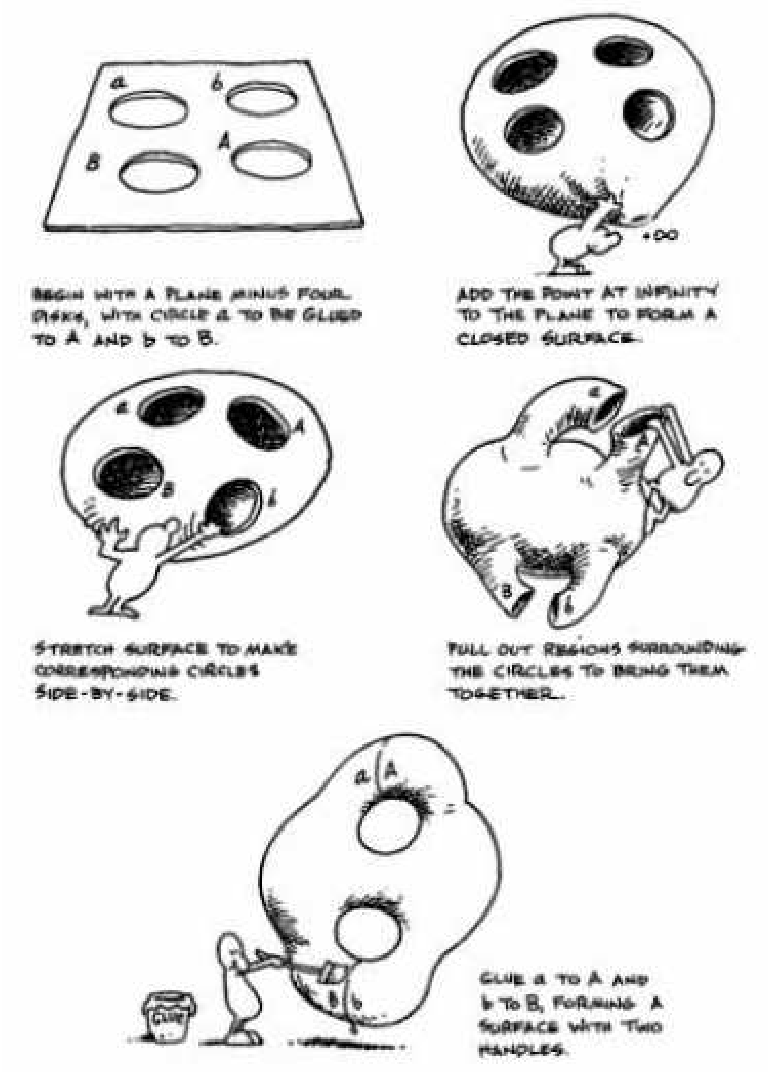
a
b
B
A
BEGIN WITH A PLANE MINUS FOUR DISKS, WITH CIRCLE a TO BE GLUED TO A AND b TO B.
+∞
ADD THE POINT AT INFINITY TO THE PLANE TO FORM A CLOSED SURFACE.
STRETCH SURFACE TO MAKE CORRESPONDING CIRCLES SIDE-BY-SIDE.
PULL OUT REGIONS SURROUNDING THE CIRCLES TO BRING THEM TOGETHER.
a A
b
GLUE
GLUE a TO A AND b TO B, FORMING A SURFACE WITH TWO HANDLES.

Figure 9.1: A 'friendly' process of compactifying and sewing a 4-punctured plane.

There are a total of $2n$ triangles and n seams in the triangulation, where n is the index

$|\Gamma_0\backslash\mathbb{H}| = |X|$. The boundary of each patch has an even number of edges, namely the double of the corresponding cusp width. The familiar formula

$$\gamma = \tfrac{n}{12} - \tfrac{n_2}{4} - \tfrac{n_3}{3} - \tfrac{n_\infty}{2} + 1$$

for the genus γ of $\Gamma_0\backslash\mathbb{H}$ in terms of the index n and the numbers n_i of Γ_0-orbits of fixed points of order i, can be interpreted in terms of the data of the quilt (see [98]), and we find in particular that if every patch of the quilt has at most six sides, then the genus will be 0 or 1, and genus 1 only exceptionally. In particular, we are interested in one class of these Γ-actions (actually an $SL_2(\mathbb{Z})$-action). For example, it is well known that the braid group B_3 has presentation

$$\langle \sigma_1, \sigma_2 \mid \sigma_1\sigma_2\sigma_1 = \sigma_2\sigma_1\sigma_2\rangle,$$

and center $Z = \langle(\sigma_1\sigma_2\sigma_1)^2\rangle$ (see [5]). It is related to the modular group by

$$B_3/Z \cong PSL_2(\mathbb{Z}), \quad B_3/\langle(\sigma_1\sigma_2\sigma_1)^4\rangle \cong SL_2(\mathbb{Z}).$$

Fix a finite group G (we are most interested in the choice $G = \mathbb{M}$). We can define a right action of B_3 on triples $(g_1, g_2, g_3) \in G^3$ by

$$(g_1, g_2, g_3)\sigma_1 = (g_1g_2g_1^{-1}, g_1, g_3), \quad (g_1, g_2, g_3)\sigma_2 = (g_1, g_2g_3g_2^{-1}, g_2). \tag{9.3}$$

We will be interested in this action on the subset of G^3 where all $g_i \in G$ are involutions. The action (9.3) is equivalent to a reduced version, where we replace (g_1, g_2, g_3) with $(g_1g_2, g_2g_3) \in G^2$. Then (9.3) becomes

$$(g, h)\sigma_1 = (g, gh), \quad (g, h)\sigma_2 = (gh^{-1}, h). \tag{9.4}$$

These B_3 actions come from specializations of the Burau and reduced Burau representations respectively [5], [114], and generalize to actions of B_n on G^n and G^{n-1}. We can get an action of $SL_2(\mathbb{Z})$ from the B_3 action (9.4) in two ways: either
(i) by restricting to commuting pairs (g, h); or
(ii) by identifying each pair (g, h) with all its conjugates (aga^{-1}, aha^{-1}).
Norton's $SL_2(\mathbb{Z})$ action of (9.1) arises from the B_3 action (9.4) when we perform both (i) and (ii).
The quilt picture was designed for this $SL_2(\mathbb{Z})$ action. The point of this construction is that the number of sides in each patch is determined by the orders of the corresponding elements g, h. If G is, say, a 6-transposition group (such as the Monster), and we take the involutions g_i from 2A, then each patch will have ≤ 6 sides, and the corresponding genus will be 0 (usually) or 1 (exceptionally). In this way we can relate the Monster with a genus 0 property.

Based on the actions (9.3) and (9.4), Norton anticipates some analogue of Moonshine valid for noncommuting pairs. Although they always seem to be modular functions, they

will no longer always be Hauptmoduls and their fixing groups won't always contain a $\Gamma(N)$. Conformal field theory considerations ('higher genus orbifolds') alluded in Section 9.1 suggest that more natural should be for example quadruples $(g, g', h, h') \in \mathbb{M}^4$ obeying $[g, h] = [h', g']$.
An important question is, how much does Monstrous Moonshine determine the Monster? How much of the $\mathbb{M}$ structure can be deduced from, for example, McKay's $\hat{E}_8$ Dynkin diagram observation, and/or the (complete) replicability of the McKay-Thompson series T_g, and/or Norton's conjectures 9.1.1, and/or Modular Moonshine in Section 9.4 below? A small start toward this is taken in [157], where some control on the subgroups of $\mathbb{M}$ isomorphic to $C_p \times C_p$ (p prime) was obtained, using only the properties of the series $N_{(g,h)}$. For related work, see [98].

9.3 Other finite groups: Mini-Moonshine

It is natural to ask about Moonshine for other groups. For example, the Hauptmodul for $\Gamma_0(2)^+$ looks like

$$J_{\Gamma_0(2)^+}(q) = q^1 + 4,372q + 96,256q^2 + 1,240,002q^3 + \ldots \qquad (9.5)$$

and we find the relations $4,372 = 4,371+1$, $96,256 = 96,255+1$, $1,240,002 = 1,139,374+4,371+2\cdot 1$, where 1, 4371, 96255, and 1139374 are all dimensions of irreducible representations of the Baby Monster $\mathbb{B}$. Thus we find 'Moonshine' for $\mathbb{B}$.

Of course any subgroup of $\mathbb{M}$ automatically inherits Moonshine by restriction, but this is not at all interesting. A more accessible sporadic is M_{24} (see for example[39]). Most constructions of the Leech lattice start with M_{24}, and most constructions of the Monster involve the Leech lattice. Thus we are led to the following natural hierarchy of (most) sporadics:

1. M_{24} (from which we can get M_{11}, M_{12}, M_{22}, M_{23}); which leads to
2. $\mathrm{Co}_0 \cong 2.\mathrm{Co}_1$ (from which we get HJ, HS, McL, Suz, Co_3, Co_2); which leads to
3. $\mathbb{M}$ (from which we get He, Fi_{22}, Fi_{23}, Fi'$_{24}$, HN, Th, $\mathbb{B}$).

It can thus be argued that we could approach problems in Monstrous Moonshine by first addressing in order M_{24} and Co_1, which should be much simpler. Indeed, the full vertex operator algebra orbifold theory (the complete analogue of Section 9.1) for M_{24} has been established in [52] (the relevant series $\mathcal{Z}_{(g,h)}$ had already been constructed in [140]). The orbifold theory for Co_1 though seems out of reach at present. Remarkably, that for the Baby Monster $\mathbb{B}$ is much more straightforward and has been worked out by Höhn [93].

Queen [159] established Moonshine for the following groups (all essentially centralizers of elements of $\mathbb{M}$): Co_0, Th, 3.2.Suz, 2.HJ, HN, $2.A_7$, He, M_{12}. In particular, to each element g of these groups, there corresponds a series $Q_g(z) = q^{-1} + \sum_{n\geq 0} a_n(g)q^n$, which is a Hauptmodul for some modular group of Moonshine-type, and where each $g \mapsto a_n(g)$ is a virtual character. For example, Queen's series Q_g for Co_0 is the Hauptmodul (1.10) for the genus 0 group $\Gamma_0(2)$. For Th, HN, He and M_{12} it is a proper character. Other differences with Monstrous Moonshine are that there can be a preferred nonzero value for the constant term a_0, and that although $\Gamma_0(N)$ will be a subgroup of the fixing group, it will not necessarily be normal. We will return to these results in the next section, where we will see that many seem to come out of the Moonshine for $\mathbb{M}$. About half of. Queen's Hauptmoduls Q_g for Co_0 do not arise as a McKay-Thompson series for $\mathbb{M}$. Norton conjecture 9.1.1 are a reinterpretation and extension of Queen's work.

Queen never reached $\mathbb{B}$ because of its size. However, the Moonshine (9.5) for $\mathbb{B}$ falls into her and Norton's scheme because equation (9.5) is the McKay-Thompson series associated to class 2A of $\mathbb{M}$, and the centralizer of an element in 2A is a double cover of $\mathbb{B}$.
There can not be a vertex operator algebra $V = \bigoplus V_n$ with graded dimension (9.5) and automorphisms in $\mathbb{B}$, because for example the $\mathbb{B}$-module V_3 does not contain V_2 as a submodule. However, Höhn deepened the analogy between $\mathbb{M}$ and $\mathbb{B}$ by constructing a vertex operator superalgebra $V\mathbb{B}^\natural$ of rank $c = 23\frac{1}{2}$, called the *shorter Moonshine module*, closely related to $V^\natural$ (see for example [94]). Its automorphism group is $C_2 \times \mathbb{B}$. Just as $\mathbb{M}$ is the automorphism group of the Griess algebra $\mathscr{B} = V_2^\natural$, so is $\mathbb{B}$ the automorphism group of the algebra $(V\mathbb{B}^\natural)_2$. Just as $V^\natural$ is associated to the Leech lattice Λ_{24}, so is $V\mathbb{B}^\natural$ associated to the *shorter Leech lattice* O_{23}, the unique 23-dimensional positive-definite self-dual lattice with no vectors of length 2 or 1 (see for example [39]). The automorphism group of O_{23} is $C_2 \times \mathrm{Co}_2$. A similar theory has recently appeared for Co_1 in [58]. There has been no interesting Moonshine for the remaining six sporadics (the pariahs J_1, J_3, Ru, O'N, Ly, J_4). There will be some sort of Moonshine for any group which is an automorphism group of a vertex operator algebra (so this means any finite group [53]). Many finite groups of Lie type should arise as automorphism groups of vertex operator algebras associated to affine algebras except defined over finite fields, but apparently all known examples of genus 0 Moonshine are limited to the groups involved with $\mathbb{M}$.

Lattices are related to groups through their automorphism groups, which are always finite for positive-definite lattices. The automorphism group $\mathrm{Aut}\,\Lambda_{24} = \mathrm{Co}_0$ of the Leech lattice has order about 8×10^{18}, and is a central extension by $\mathbb{Z}_2$ of Conway's simple group Co_1. Several other sporadic groups are also involved in Co_0, as we have seen. To each automorphism $\alpha \in \mathrm{Co}_0$, let θ_α denote the theta function of the sublattice of Λ_{24} fixed by α. Conway and Norton also associate with each automorphism α a certain function $\eta_\alpha(z)$ of the form $\prod_i \eta(a_i z)/\prod_j \eta(b_j z)$ built out of the Dedekind's η function (Example 1.4.2). Both θ_α and η_α are constant on each conjugacy class in Co_0, of which there are 167. [38] remarks that the ratio $\theta_\alpha/\eta_\alpha$ always seems to equal some McKay-Thompson series $T_{g(\alpha)}$. It turns out that this observation is not correct [126]. For each automorphism $\alpha \in \mathrm{Co}_0$, the subgroup of $SL_2(\mathbb{R})$ that fixes $\theta_\alpha/\eta_\alpha$ is indeed always genus 0, but for exactly 15 conjugacy classes in Co_0, $\theta_\alpha/\eta_\alpha$ is not the Hauptmodul. Nevertheless, this construction proved useful for establishing Moonshine for M_{24} [140]. Similarly, one can ask this for the E_8 root lattice, whose automorphism group is the Weyl group of the Lie algebra E_8 (of order 696,729,600). The automorphisms of the lattice E_8 that yield a Hauptmodul were classified in [162]. On the other hand, Koike established a Moonshine of this kind for the groups $PSL_2(\mathbb{F}_7)$, $PSL_2(\mathbb{F}_5) \cong \mathrm{Alt}_5$ and $PSL_2(\mathbb{F}_3)$, of order 168, 60 and 12, respectively [119, 121, 120, 122, 123].

9.4 Modular Moonshine

Consider an element $g \in \mathbb{M}$. We expect from [155], [159], [51], that there is a Moonshine for the centralizer $C_{\mathbb{M}}(g)$ of g in $\mathbb{M}$, governed by the g-twisted module $V^\natural(g)$. Unfortunately, $V^\natural(g)$ is not usually itself a vertex operator algebra, so the analogy with $\mathbb{M}$ is not perfect. Ryba and Borcherds [161], [19], [16] found it interesting that, for $g \in \mathbb{M}$ of prime order p, the Norton series $N_{(g,h)}$ can be transformed into a McKay-Thompson series (and has all the associated nice properties) whenever h is p-regular (that is, h has order coprime to p). This special behavior of p-regular elements suggested to him to look at modular representations. The basics of modular representations and Brauer characters are discussed in sufficient detail in [45].

A *modular representation* π of a group G is a representation defined over a field of positive characteristic p dividing the order $|G|$. Such representations possess many special features. For one thing, they are no longer completely reducible (so the role of irreducible modules as direct summands will be replaced with their role as composition factors). For another, the usual notion of character (the trace of representation matrices) loses its usefulness and is replaced by the more subtle Brauer character $\beta(\pi)$: a complex-valued class function on $\mathbb{M}$ which is only well-defined on the p-regular elements of G. We have, for example (see [16], [19], [161]).

Theorem 9.4.1. *Let $g \in \mathbb{M}$ be any element of prime order p, for any p dividing $|M|$. Then, there is a vertex operator superalgebra ${}^gV = \bigoplus_{n\in\mathbb{Z}} {}^gV_n$ defined over the finite field $\mathbb{F}_p$ and acted on by the centralizer $C_{\mathbb{M}}(g)$. If $h \in C_{\mathbb{M}}(g)$ is p-regular, then the graded Brauer character*

$$R_{(g,h)}(z) = q^{-1} \sum_{n\in\mathbb{Z}} \beta({}^gV_n)(h) q^n$$

equals the McKay-Thompson series $T_{gh}(z)$. Moreover, for g belonging to any conjugacy class in $\mathbb{M}$ except 2B, 3B, 5B, 7B, or 13B, this is in fact an ordinary vertex operator algebra (that is the 'odd' part vanishes), while in the remaining cases the graded Brauer characters of both the odd and even parts can separately be expressed using McKay-Thompson series.

By a vertex operator superalgebra, we mean there is a $\mathbb{Z}_2$-grading into even and odd subspaces, and for u, v both odd, the commutator in the locality axiom of Theorem 4.1.2 is replaced by an anticommutator. In the proof, the superspaces arise as cohomology groups, which naturally form an alternating sum. The centralizers $C_{\mathbb{M}}(g)$ in the theorem are quite nice; for example for g in classes 2A, 2B, 3A, 3B, 3C, 5A, 5B, 7A, 11A, respectively, these involve the sporadic groups $\mathbb{B}$, Co_1, $\mathrm{Fi'}_{24}$, Suz, Th, HN, HJ, He, and M_{12}. The proof for $p = 2$ is not complete at the present time. The conjectures in [161] concerning modular analogues of the Griess algebra for several sporadics follow from Theorem 9.4.1. Can these modular ${}^gV$ vertex operator algebras be interpreted as a reduction $\pmod p$ of (super)algebras in characteristic 0? Also, what about elements g of composite order? Borcherds has stated the following in [16]

Conjecture 9.4.2 (Borcherds)**.** *Choose any $g \in \mathbb{M}$ and let n denote its order. Then, there is a $\frac{1}{n}\mathbb{Z}$-graded superspace ${}^g\hat{V} = \bigoplus_{i \in (1/n)\mathbb{Z}} {}^g\hat{V}_i$ over the ring of cyclotomic integers $\mathbb{Z}[e^{2\pi i/n}]$. It is often (but probably not always) a vertex operator superalgebra; in particular, ${}^1\hat{V}$ is an integral form of the Moonshine module $V^\natural$. Each ${}^g\hat{V}$ carries a representation of a central extension of $C_{\mathbb{M}}(g)$ by C_n. Define the graded trace*

$$B_{(g,h)}(z) = q^{-1} \sum_{i \in \frac{1}{n}\mathbb{Z}} \mathrm{tr}(h | {}^g\hat{V}_i) q^i.$$

If $g, h \in \mathbb{M}$ commute and have coprime orders, then $B_{(g,h)}(z) = T_{gh}(z)$. If all q-coefficients of T_g are non-negative, then the 'odd' part of ${}^g\hat{V}$ vanishes, and ${}^g\hat{V}$ is the g-twisted module $V^\natural(g)$. If g has prime order p, then the reduction $\pmod p$ of ${}^g\hat{V}$ is the modular vertex operator superalgebra ${}^gV$ of Theorem 9.4.1.

When we say ${}^1\hat{V}$ is an integral form for $V^\natural$, we mean that ${}^1\hat{V}$ has the same structure as a vertex operator algebra, with everything defined over $\mathbb{Z}$, and tensoring it with $\mathbb{C}$ recovers $V^\natural$. This remarkable conjecture, which tries to explain Theorem 9.4.1, is completely open.

9.5 The geometry of Moonshine

Algebra is the mathematics of structure, and so of course it has a profound relationship with every area of mathematics. Therefore the trick for finding possible fingerprints of Moonshine in, say, geometry is to look there for modular functions. That search quickly leads to the elliptic genus.

For details see for example [95], [164], or [170]. All manifolds here are compact, oriented and differentiable. In Thom's cobordism ring Ω, elements are equivalence classes of cobordant manifolds, addition is connected sum, and multiplication is Cartesian product. The universal elliptic genus $\phi(M)$ is a ring homomorphism from $\mathbb{Q} \otimes \Omega$ to the ring of power series in q, which sends n-dimensional manifolds with spin connections to a weight $\frac{n}{2}$ modular

form of $\Gamma_0(2)$ with integer coefficients. Several variations and generalizations have been introduced, *e. g.*, the Witten genus assigns to spin manifolds with vanishing first Pontryagin class a weight $\frac{n}{2}$ modular form of $SL_2(\mathbb{Z})$ with integer coefficients.
We have noticed several deep relationships between elliptic genera Moonshine. For instance, the important rigidity property of the Witten genus with respect to any compact Lie group action on the manifold, is a consequence of the modularity of the characters of affine algebras [135]. The elliptic genus of a manifold M has been interpreted as the graded dimension of a vertex operator superalgebra constructed from M [169]. Seemingly related to this, [21] recovered the elliptic genus of a Calabi-Yau manifold X from the sheaf of vertex algebras in the chiral de Rham complex $\mathcal{MSV}$[139] attached to X. Unexpectedly, the elliptic genus of even-dimensional projective spaces P^{2n} has non-negative coefficients and in fact equals the graded dimension of some vertex algebra [138]; this suggests interesting representation theoretic questions in the spirit of Monstrous Moonshine. In physics, elliptic genera arise as partition functions of $N = 2$ superconformal field theories [183]. Mason's constructions [140] associated to Moonshine for the Mathieu group M_{24} have been interpreted as providing a geometric model (elliptic system) for elliptic cohomology $\mathrm{Ell}^*(BM24)$ of the classifying space of M_{24} [170], [54]. The Witten genus (normalized by η^8) of the Milnor-Kervaire manifold M_0^8, an 8-dimensional manifold built from the E_8 diagram, equals $j^{1/3}$ [95] (recall (4.32)).
Another interesting fact is that a Borcherds algebra can be associated with any even Lorentzian lattice, and also with any Calabi-Yau manifold [92]. Of course it is a broad enough class that almost all of them will be uninteresting; an intriguing approach to identifying the interesting ones is by considering the so called automorphic products [14, 15].

Hirzebruch's 'prize question' [95] asks for the construction of a 24-dimensional manifold M with Witten genus J (after being normalized by η^{24}). We would like $\mathbb{M}$ to act on M by diffeomorphisms, and the twisted Witten genera to be the McKay-Thompson series T_g. It would also be nice to associate Norton series $N_{(g,h)}$ to this Moonshine manifold. Constructing such a manifold is perhaps the remaining 'Holy Grail' of Monstrous Moonshine. Hirzebruch's question was partially answered by Mahowald and Hopkins [137], who constructed a manifold with Witten genus J, but could not show that it would support an effective action of the Monster. Related work is by Aschbacher [1], who constructed several actions of $\mathbb{M}$ on, for example, 24-dimensional manifolds (but none of which could have genus J), and Kultze [125], who showed that the graded dimensions of the subspaces $V^\natural_\pm$ of the Moonshine module are twisted $\hat{A}$-genera of the Milnor-Kervaire manifold M_0^8 (the $\hat{A}$-genus is the specialization of elliptic genus to the cusp $i\infty$).

There has been a second conjectured relationship between geometry and Monstrous Moonshine. *Mirror symmetry* says that most Calabi-Yau manifolds come in closely related pairs. Consider a 1-parameter family X_z of Calabi-Yau manifolds, with mirror X^* given by the resolution of an orbifold X/G for G finite and abelian. Then the Hodge numbers $h^{1,1}(X)$ and $h^{2,1}(X^*)$ will be equal, and more precisely the moduli space of (complexified) Kähler structures on X will be locally isometric to the moduli space of complex structures on X. The 'mirror map' $z(q)$, which can be defined using the Picard-Fuchs equation [151], gives a canonical map between those moduli spaces. For example, $x1^4 + x_2^4 + x_3^4 + x_4^4 + z^{-1/4}x_1x_2x_3x_4 = 0$ is such a family of $K3$ surfaces (that is Calabi-Yau 2-folds), where $G = C_4 \times C_4$. Its mirror map is given by

$$z(q) = q - 104q^2 + 6,444q^3 - 311,744q^4 + 13,018,830q^5 - 493,025,760q^6 + \dots \quad (9.6)$$

Lian and Yau [133] noticed that the reciprocal $\frac{1}{z(q)}$ of the mirror map in (9.6) equals the McKay-Thompson series $T_g(z) + 104$ for g in class 2A of $\mathbb{M}$. After looking at several other examples with similar conclusions, they proposed their *Mirror-Moonshine conjecture*: The reciprocal $1/z$ of the mirror map z of a 1-parameter family of $K3$ surfaces with an orbifold mirror will be a McKay-Thompson series (up to an additive constant).

A counterexample (and more examples) are given in [180]. In particular, although there are relations between mirror symmetry and modular functions (see for example [89] and [92]), there does not seem to be any special relation with the Monster. Doran [56] demystifies the Mirror-Moonshine phenomenon by finding necessary and sufficient conditions for $1/z$ to be a modular function for a modular group commensurable with $SL_2(\mathbb{Z})$.

9.6 Moonshine and physics

The physical side (perturbative string theory, or equivalently conformal field theory) of Moonshine was noticed early on, and has profoundly influenced the development of Moonshine and vertex operator algebras. This is a very rich subject, which we can only superficially touch on. The book [64], with its extensive bibliography, provides an introduction but will be difficult reading for many mathematicians (as will this section). The treatment in [70] is more accessible and shows how naturally vertex operator algebras arise from the physics. This effectiveness of physical interpretations is not magic; it merely tells us that many of our finite-dimensional objects are seen much more clearly when studied through infinite-dimensional structures. Of course Moonshine, which teaches us to study the finite group $\mathbb{M}$ via its infinite-dimensional module $V^\natural$, fits perfectly into this picture.

A *conformal field theory* is a quantum field theory on 2-dimensional space-time, whose symmetries include the conformal transformations. In string theory the basic objects are finite curves —called *strings*— rather than points (particles), and the conformal field theory lives on the surface traced by the strings as they evolve (colliding and separating) through time. Each conformal field theory is associated with a pair V_L, V_R of mutually commuting vertex operator algebras, called its *chiral algebras* [4]. For example, strings living on a compact Lie group manifold (the so-called Wess-Zumino-Witten model) will have chiral algebras given by affine algebra vertex operator algebras. The space $\mathcal{H}$ of states for the conformal field theory carries a representation of $V_L \otimes \overline{V}_R$, and many authors have (somewhat optimistically) concluded that the study of conformal field theories reduces to that of vertex operator algebra representation theory. Rational vertex operator algebras correspond to the important class of rational conformal field theories, where $\mathcal{H}$ decomposes into a finite sum $\oplus M_L \otimes M_R$ of irreducible modules. The Virasoro algebra of Section (4.2) arises naturally in conformal field theory through infinitesimal conformal transformations. The vertex operator $Y(\phi, z)$, for the space-time parameter $z = e^{t+ix}$, is the quantum field which creates from the vacuum $|0\rangle \in \mathcal{H}$ the state $|\phi\rangle \in \mathbb{H}$ at time $t = -\infty : |\phi\rangle = \lim_{z \to 0} Y(\phi, z)|0\rangle$. In particular, Borcherds' definition [8, 13] of vertex operator algebras can be interpreted as an axiomatisation of the notion of chiral algebra in conformal field theory, and for this reason alone is important.

In conformal field theory, the Hauptmodul property of Moonshine is hard to interpret, and a less direct formulation, like that in [177], is needed. However, both the statement and proof of Zhu's Theorem are natural from the conformal field theory framework (see [70]); for example, the modularity of the series T_g and $N_{(g,h)}$ are automatic in conformal field theory. This modularity arises in conformal field theory through the equivalence of the Hamiltonian formulation, which describes concretely the graded spaces we take traces on (and hence the coefficients of our q-expansions), and the Feynman path formalism, which interprets these graded traces as sections over moduli spaces (and hence makes modularity manifest). Beautiful reviews are sketched in [184], [183]. More explicitly, the Virasoro action on moduli spaces discussed in section 4.2 of [71] gives rise to a system of partial differential equations (the Knizhnik-Zamolodchikov equations). According to conformal field theory, the vertex operator algebra characters will satisfy those equations for a torus with one puncture, and their modularity (that is, Zhu's Theorem) arises from the monodromy of those equations.
Because $V^\natural$ is so mathematically special, it may be expected that it corresponds to interesting physics. Certainly it has been the subject of some speculation. There will be a $c = 24$ rational conformal field theory whose chiral algebra V_L and state space $\mathcal{H}$ are both $V^\natural$, while V_R is trivial (this is possible because $V^\natural$ is holomorphic). This conformal field theory is nicely described in [48]; see also [50]. The Monster is the symmetry of that conformal field theory, but the Bimonster $\mathbb{M} \wr C_2$ will be the symmetry of a rational conformal field theory with $\mathcal{H} = V^\natural \otimes \overline{V}^\natural$. The paper [41] finds a family of D-branes for the latter theory which are in one-to-one correspondence with the elements of $\mathbb{M}$, and their 'overlaps' $\langle\langle g||q^{\frac{1}{2}(L_0+\overline{L}_0+c/24)}||h\rangle\rangle$ equal the McKay-Thompson series $T_{g^{-1}h}$. However, we still lack any explanation as to why a conformal field theory involving $V^\natural$ should yield interesting physics.

Almost every facet of Moonshine finds a natural formulation in conformal field theory, where it often was discovered first. For example, the no-ghost theorem (Theorem 8.3.2) of Brower-Goddard-Thorn was used to great effect in [12] to understand the structure of the Monster Lie algebra $\mathfrak{M}$. On a finite-dimensional manifold M, the index of the Dirac operator D in the heat kernel interpretation is a path integral in supersymmetric quantum mechanics, that is, an integral over the free loop space $\mathcal{L}M = \{\gamma : S^1 \to M\}$; the string theory version of this is that the index of the Dirac operator on $\mathcal{L}M$ should be an integral over $\mathcal{L}(\mathcal{L}M)$, that is over smooth maps of tori into M, and this is just the elliptic genus, and explains why it should be modular. The orbifold construction of [51] comes straight from conformal field theory (although construction of $V^\natural$ in [66] predates conformal field theory orbifolds by a year and in fact influenced their development in physics). That said, the translation process from physics to mathematics of course is never easy; Borcherds' definition [8] is a prime example.

From this standpoint, what is most exciting is what has not yet been fully exploited. String theory tells us that conformal field theory can live on any surface Σ. The vertex operator algebras, including the geometric vertex operator algebras of [99], capture conformal field theory in genus 0. The graded dimensions and traces considered above concern conformal field theory quantities (conformal blocks) at genus 1: $z \mapsto e^2\pi iz$ maps $\mathbb{H}$ onto a cylinder, and the trace identifies the two ends. There are analogues of all this at higher genus [188] (though the formulas can rapidly become awkward). For example, the graded dimension of the $V^\natural$ conformal field theory in genus 2 is computed in [178], and involves for instance Siegel theta functions. The orbifold theory in Section 9.1 is genus 1: each 'sector' (g, h) corresponds to a homomorphism from the fundamental group $\mathbb{Z}^2$ of the torus into the orbifold group G (for example $G = \mathbb{M}$); g and h are the targets of the two generators of $\mathbb{Z}^2$ and hence must commute. More generally, the sectors will correspond to each homomorphism $\varphi : \pi_1(\Sigma) \to G$, and to each we will get a higher genus trace $\mathcal{Z}(\varphi)$, which will be a function on the Teichmüller space $\mathcal{T}_g$ (generalizing the upper half-plane $\mathbb{H}$ for genus 1). The action of $SL_2(\mathbb{Z})$ on the $N_{(g,h)}$ generalizes to the action of the mapping class group on $\pi_1(\Sigma)$ and $\mathcal{T}_g$. See for example [6] for some thoughts in this direction. Recently, Witten [186] has also related the Monster with 3-dimensional gravity in black holes, although the paper is still abundant in conjectures.

9.7 Conclusion

There are different basic aspects to Monstrous Moonshine: (i) why modularity enters at all; (ii) why in particular we have genus 0; and (iii) what it has to do with the Monster. Today, we understand (i) best. There will be a Moonshine-like relations between any (subgroup of the) automorphism group of any rational vertex operator algebra, and the characters χ_M, and the same can be expected to hold of the orbifold characters $\mathcal{Z}$ in Section 9.1.

To prove the genus 0 property of the McKay-Thompson series T_g, we needed recursions obtained one way or another from the Monster Lie algebra $\mathfrak{M}$, and from these we apply Theorem 7.4.4. These recursions are very special, but so presumably is the 0 genus property. The suggestion of [28], though, is that we may be able to considerably simplify this part of the argument.

Every group known to have rich Moonshine properties is contained in the Monster. To what extent can we derive $\mathbb{M}$ from Monstrous Moonshine? The understanding of this seemingly central role of $\mathbb{M}$ is the poorest of those three aspects.

The central role that vertex operator algebras play in our current understanding of Moonshine should be clear from this review (that is basically Zhu's Theorem). The excellent review [54] makes this point even more forcefully. However, it can be (and has been) questioned whether the full and difficult machinery of vertex operator algebras is really needed to understand this; that is, whether we really have isolated the key conjunction of properties needed for Moonshine to arise. Conformal field theory has been an invaluable guide thus far, but perhaps we are a little too steeped in its lore.
In particular, what it is really needed is a second independent proof of the Moonshine conjectures. One tempting possibility is the heat kernel; its general role in modularity concerns is emphasized in [109], and is also the central ingredient in the Knizhnik-Zamolodchikov equations for affine algebras [96]. A heat kernel probably plays an analogous role in the Knizhnik-Zamolodchikov equations associated to $V^\natural$, but does it relate to the genus 0 property? Another possibility is the braid group B_3, whose fingerprints are all over the mathematics and physics of Moonshine.

Moonshine (in its more general sense) is a relation between algebra and number theory, and its impact on algebra has been dramatic (for example: vertex operator algebras, $V^\natural$, Borcherds-Kac-Moody algebras). Its impact on number theory has been far less so. This may merely be a temporary accident due to the backgrounds of most researchers (including the mathematical physicists) working to date in the area. Gannon [71] has suggested that the most exciting prospects for the future of Moonshine are in the direction of number theory. Hints of this future can be found in for example [15], [47], [57], [90], [150], [185] and [18]. Other recent advances and discussions are [144], [63], [37], [146], and [143].

References

[1] M. G. Aschbacher. Finite groups acting on homology manifolds. *Pacific J. Math.*, Special Issue:3–36, 1997.

[2] A. L. Atkin and J. N. L'Brien. On divisibility properties of the coefficients of the j-function. *Trans. Amer. Math. Soc.*, 126, 1967.

[3] A. L. Atkin and J. Lehner. Hecke operators on $\Gamma_0(m)$. *Math. Ann.*, 185:134–160, 1970.

[4] A. Belavin, A. Polyakov, and A. Zamolodchikov. Infinite conformal symmetries in two dimensional quantum field theory. *Nucl. Phys.*, B241:333–380, 1984.

[5] J. S. Birman. *Braids, links and mapping class groups.* Princeton University Press, Princeton, 1974.

[6] P. Bántay. Higher genus moonshine. *Moonshine, the Monster, and related topics. Contemp. Math.*, 193:1–8, 1996.

[7] R. E. Borcherds. Vertex algebras. *Preprint.*

[8] R. E. Borcherds. Vertex algebras, Kac-Moody algebras and the Monster. *Proc. Natl. Acad. Sci. USA*, 83:3068–3071, 1986.

[9] R. E. Borcherds. Generalized Kac-Moody algebras. *J. Algebra*, 115:501–512, 1988.

[10] R. E. Borcherds. The monster Lie algebra. *Adv. Math.*, 83:30–47, 1990.

[11] R. E. Borcherds. Central extensions of generalized Kac-Moody algebras. *J. Algebra*, 140:330–335, 1991.

[12] R. E. Borcherds. Monstrous moonshine and monstrous Lie superalgebras. *Invent. Math.*, 109:405–444, 1992.

[13] R. E. Borcherds. Sporadic groups and string theory. In *First European Congress of Math. (Paris, 1992), I*, pages 411–421, Basel, 1994. Birkhäuser.

[14] R. E. Borcherds. Automorphic forms on $O_{(s+2,2)}(\mathbb{R})$ and generalized Kac-Moody algebras. In *Proceedings of the International Congress of Mathematicians (Zürich, 1994)*, pages 744–752, Basel, 1995. Birkhäuser.

[15] R. E. Borcherds. Automorphic forms with singularities on Grassmannians. *Invent. Math.*, 132:491–562, 1998.

[16] R. E. Borcherds. Modular moonshine III. *Duke Math. J.*, 93:129–154, 1998.

[17] R. E. Borcherds. What is moonshine? In *Proceedings of the International Congress of Mathematicians (Berlin, 1998)*, pages 607–615, Bielefeld, 1998. Documenta Mathematica.

[18] R. E. Borcherds. Problems in Moonshine. In *First International Congress of Chinese Mathematics (Beijing, 1998)*, pages 3–10, Providence, 2001. American Mathematical Society.

[19] R. E. Borcherds and A. J. E. Ryba. Modular moonshine II. *Duke Math. J.*, 83:435–459, 1996.

[20] A. Borel. Seminar on Complex Multiplication. *Lecture Notes in Mathematics*, 21. Springer-Verlag, 1966.

[21] L. A. Borisov and A. Libgober. Elliptic genera of toric varieties and applications to mirror symmetry. *Invent. Math.*, 140:453–485, 2000.

[22] T. Bröcker and T. Dieck. *Representations of Compact Lie groups.* Springer-Verlag, Berlin, 1985.

[23] W. Burnside. *Theory of groups of finite order.* Cambridge University Press, Cambridge, 1911.

[24] S. Carnahan. Generalized Moonshine I: genus zero functions. *arXiv:0812.3440v2*, 2008.

[25] S. Carnahan. Generalized Moonshine II: Borcherds products. *arXiv:0908.4223v2*, 2009.

[26] A. A. Castro. *Curso de equações diferenciais ordinárias.* preprint, Rio de Janeiro, 2009.

[27] C. Chevalley. Sur certains groupes simples. *Tohoku Math. J.*, 7:14–66, 1955.

[28] H. Cohn and J. McKay. Spontaneous generation of modular invariants. *Math. Comp.*, 65:1295–1309, 1996.

[29] J. H. Conway. A perfect group of order 8,315,553,613,086,720,000 and the sporadic simple groups. *Proc. Natl. Acad. Sci. USA*, 61:398–400, 1968.

[30] J. H. Conway. A characterization of Leech's lattice. *Invent. Math.*, 7:137–142, 1969.

[31] J. H. Conway. A group of order 8,315,553,613,086,720,000. *Bull. London Math. Soc.*, 1:79–88, 1969.

[32] J. H. Conway. Three lectures on exceptional groups. In *Finite Simple Groups*, pages 215–247, 1971.

[33] J. H. Conway. The miracle octad generator. In *Topics in group theory and computation*, pages 62–68, 1977.

[34] J. H. Conway. A simple construction for the Fischer-Griess monster group. *Invent. Math.*, 79:513–540, 1985.

[35] J. H. Conway. The monster group and its 196884-dimensional space. In *Sphere Packings, Lattices and Groups*, pages 555–567, 1988.

[36] J. H. Conway, R. T. Curtis, S. P. Norton, and *et. al. Atlas of finite groups.* Clarendon Press, Oxford, 1985.

[37] J. H. Conway, J. McKay, and A. Sebbar. On the discrete groups of Moonshine. *Proc. Amer. Math. Soc.*, 132:2233–2240, 2004.

[38] J. H. Conway and S. P. Norton. Monstrous Moonshine. *Bull. London Math. Soc.*, 11:308–339, 1979.

[39] J. H. Conway and N. J. A. Sloane. *Sphere Packings, Lattices and Groups.* Springer-Verlag, New York, 1988.

[40] A. Coste and T. Gannon. Remarks on Galois in rational conformal field theories. *Phys. Lett.*, B323:316–321, 1994.

[41] B. Craps, M. R. Gaberdiel, and J. A. Harvey. Monstrous branes. *Comm. Math. Phys.*, 234:229–251, 2003.

[42] C. J. Cummins. Congruence subgroups of groups commensurable with $PSL_2(\mathbb{Z})$ of genus 0 and 1. *Experiment. Math.*, 13:361–382, 2004.

[43] C. J. Cummins and T. Gannon. Modular equations and the genus zero property of moonshine functions. *Invent. Math.*, 129:413–443, 1997.

[44] C. J. Cummins and S. P. Norton. Rational Hauptmodul are replicable. *Canad. J. Math.*, 47:1201–1218, 1995.

[45] C. W. Curtis and I. Reiner. *Methods of representation theory with applications to finite groups and orders, I.* Wiley, New York, 1981.

[46] R. T. Curtis. A new combintorial approach to M_{24}. *Math. Proc. Cambridge Philos. Soc.*, 79:25–42, 1976.

[47] R. Dijkgraaf. The mathematics of fivebranes. In *Proc. Intern. Congr. Math. (Berlin, 1998), III*, pages 133–142, Bielefeld, 1998. Documenta Mathematica.

[48] L. Dixon, P. Ginsparg, and J. A. Harvey. Beauty and the beast: Superconformal symmetry in a monster module. *Comm. Math. Phys.*, 119:221–241, 1988.

[49] L. Dixon, J. A. Harvey, C. Vafa, and E. Witten. Strings on orbifolds. *Nucl. Phys.*, B261:678–686, 1985.

[50] L. Dolan, P. Goddard, and P. Montague. Conformal field theory of twisted vertex operators. *Nucl. Phys.*, B338:529–601, 1990.

[51] C. Dong, H. Li, and G. Mason. Modular invariance of trace functions in orbifold theory and generalized moonshine. *Comm. Math. Phys.*, 214:1–56, 2000.

[52] C. Dong and G. Mason. An orbifold theory of genus zero associated to the sporadic group M_{24}. *Comm. Math. Phys.*, 164:87–104, 1994.

[53] C. Dong and G. Mason. Nonabelian orbifolds and the boson-fermion correspondence. *Comm. Math. Phys.*, 163:523–559, 1994.

[54] C. Dong and G. Mason. Vertex operator algebras and moonshine: A survey. *Progress in Algebraic Combinatorics. Adv. Stud. Pure Math.*, 24:101–136, 1996.

[55] C. Dong and F. Xu. Conformal nets associated with lattices and their orbifolds. *arXiv:math.QA/041.*

[56] C. F. Doran. Picard-Fuchs uniformization and modularity of the mirror map. *Comm. Math. Phys.*, 212:625–647, 2000.

[57] V. G. Drinfeld. On quasitriangular quasi-Hopf algebras on a group that is closely related with $\mathrm{Gal}\,\mathbb{Q}/\mathbb{Q}$. *Leningrad. Math. J.*, 2:829–860, 1991.

[58] J. F. Duncan. Super-Moonshine for Conway's largest sporadic group. *arXiv:math.RT/0502267.*

[59] K. Erdmann and M. J. Wildon. *Introduction to Lie algebras.* Springer-Verlag, London, 2006.

[60] G. Faber. Über polynomische Entwicklungen. *Math. Ann.*, 57:389–408, 1903.

[61] W. Feit and J. Thompson. Solvability of groups of odd order. *Pacific J. Math.*, 13:775–1029, 1963.

[62] B. Fischer, D. Livingstone, and M. P. Thorne. *The characters of the monster simple group.* Birmingham, 1978.

[63] D. J. Ford and J. McKay. *Monstrous Moonshine - Problems Arising I, Tate Characters.* preprint, 2002.

[64] P. Di Francesco, P. Mathieu, and D. Sénéchal. *Conformal field theory.* Springer, New York, 1997.

[65] I. Frenkel. Representations of Kac-Moody algebras and dual resonance models. *Lect. Appl. Math. AMS*, 21:325–353, 1985.

[66] I. Frenkel, J. Lepowsky, and A. Meurman. A natural representation of the Fischer-Griess monster with the modular function J as character. *Proc. Natl. Acad. Sci. USA*, 81:3256–3260, 1984.

[67] I. Frenkel, J. Lepowsky, and A. Meurman. *Vertex Operator Algebras and the Monster.* Academic Press, New York, 1988.

[68] R. Fricke. *Die Elliptische Funktionen un Ihre Anwendungen.* Teubner, Leipzig, 1922.

[69] L. Fuchs. Über eine Klasse von Funktionen nehrerer Variabeln welche durch Umkehrun der Integrale von Lösungen der linearen Differenetialgleichungen mit Rationalen Coefficienten entstehen. *J. Reine Angew.*, 89:151–169, 1880.

[70] M. R. Gaberdiel and P. Goddard. Axiomatic conformal field theory. *Comm. Math. Phys.*, 209:549–594, 2000.

[71] T. Gannon. Monstrous Moonshine: the first twenty-five years. *Bull. London Math. Soc.*, 38:1–33, 2006.

[72] T. Gannon. *Moonshine Beyond the Monster.* Cambridge University Press, Cambridge, 2006.

[73] H. Garland and J. Lepowsky. Lie algebra homology and the Macdonald-Kac fornulas. *Invent. Math.*, 34:37–76, 1976.

[74] G. Glauberman and S. P. Norton. On McKay's connection between the affine E_8 diagram and the Monster. In *Proc. on Moonshine and related topics*, pages 37–42, 2001.

[75] P. Goddard and D. Olive. *Kac-Moody and Virasoro algebras: a reprint volume for physicists.* Scientific World, Singapore, 1988.

[76] P. Goddard and C. B. Thorn. Compatibility of the dual Pomeron with unitary and the absence of ghosts in the dual resonance model. *Phys. Lett. B*, 40:235–238, 1972.

[77] D. Gorenstein. *Finite simple groups: an introduction to their classification.* Plenum Press, New York, 1982.

[78] R. L. Griess. The structure of the 'Monster' simple group. In *Proc. of the Conference on Finite Groups*, pages 113–118, New York, 1976. Acad. Press.

[79] R. L. Griess. A construction of F_1 as automorphisms of a 196,993 dimensional algebra. *Proc. Natl. Acad. Sci.*, 78:689–691, 1981.

[80] R. L. Griess. The friendly giant. *Invent. Math.*, 69:1–102, 1982.

[81] R. L. Griess. Code Loops. *J. of Algebra*, 100:224–234, 1986.

[82] R. L. Griess. The Monster and its nonassociative algebra. *Contemp. Math.*, 45:121–157, 1986.

[83] R. L. Griess. Sporadic groups, code loops and nonvanishing cohomology. *J. of pure and appl. Algebra*, 44:191–214, 1987.

[84] R. L. Griess. A Moufang loop, the exceptional Jordan algebra and a cubic form in 27 variables. *J. of Algebra*, 131:281–293, 1990.

[85] R. L. Griess. Codes, loops and p-locals. In *Proc. of the Monster Bach Conference*, Columbus, 1993. Ohio U.

[86] R. L. Griess. *Twelve sporadic groups.* Springer-Verlag, Berlin, 1998.

[87] R. L. Griess, U. Meierfrankenfeld, and Y. Segev. A uniqueness proof for the Monster. *Annals of Math.*, 130:567–602, 1989.

[88] P. A. Grillet. *Abtract Algebra.* Springer-Verlag, GTM 242 series, 2007.

[89] V. A. Gritsenko and V. V. Nikulin. The arithmetic mirror symmetry and Calabi-Yau manifolds. *Comm. Math. Phys.*, 210:1–11, 2000.

[90] S. Gukov and C. Vafa. Rational conformal field theories and complex multiplication. *Comm. Math. Phys.*, 246:181–210, 2004.

[91] B. C. Hall. *Lie groups, Lie algebras and representations.* Springer-Verlag, GTM 222 series, 2003.

[92] J. A. Harvey and G. Moore. Algebras, BPS states, and strings. *Nucl. Phys.*, B463:315–368, 1996.

[93] G. Höhn. Generalized moonshine for the Baby Monster. *Prepint.*

[94] G. Höhn. The group of symmetries of the shorter Moonshine module. *arXiv:math.QA/0210076.*

[95] F. Hirzebruch, T. Berger, and R. Jung. *Manifolds and modular forms.* Friedr. Vieweg und Sohn, Berlin, 1991.

[96] N. Hitchin. Flat connections and geometric quantization. *Comm. Math. Phys.*, 131:347–380, 1990.

[97] K. Hoffman and R. Kunze. *Linear Algebra.* Prentice-Hall, 1971.

[98] T. Hsu. Quilts: Central extensions, braid actions, and finite groups. *Lecture Notes in Mathematics*, 1731, 2000.

[99] Y.-Z. Huang. *Two-dimensional conformal geometry and vertex operator algebras.* Birkhäuser, Boston, 1997.

[100] J. Humphreys. *Introduction to Lie algebras and representation theory.* Springer-Verlag, GTM 9 series, 1972.

[101] K. Ireland and M. Rosen. *A Classical introduction to Modern Number Theory.* Springer-Verlag, GTM 84 series, 1982.

[102] A. A. Ivanov. Geometric presentations of groups with an application to the Monster. In *Proc. Intern. Congr. Math. (Kyoto, 1990)*, pages 1443–1453, Hong Kong, 1990. Springer.

[103] A. A. Ivanov. Y-groups via transitive extension. *J. Algebra*, 218:412–435, 1999.

[104] R. Ivanov and M. P. Tuite. Some irrational generalized Moonshine from orbifolds. *Nucl. Phys.*, B635:473–491, 2002.

[105] J. McKay J. H. Conway and A. Sebbar. On the discrete groups of moonshine. *Proc. Amer. Math. Soc.*, 132:2233–2240, 2004.

[106] C. G. Jacobi. *Fundamenta nova theoriae functionum ellipticarum.* Königsberg, 1829.

[107] N. Jacobson. *Basic Algebra I.* W. H. Freeman and Company, San Francisco, 1974.

[108] N. Jacobson. *Basic Algebra II.* W. H. Freeman and Company, San Francisco, 1980.

[109] J. Jorgenson and S. Lang. The ubiquitous heat kernel. In *Mathematics unlimited - 2001 and beyond*, pages 655–683, Berlin, 2001. Springer-Verlag.

[110] V. Kac. Simple irreducible graded Lie algebras of finite growth. *Math. USSR-Izv.*, 2:1271–1311, 1968.

[111] V. Kac. An elucidation of: infinite-dimensional algebras, Dedekind's η-function, classical Möbius function and the very strange formula $E_8^{(1)}$ and the cube root of the modular invariant j. *Adv. Math.*, 35:264–273, 1980.

[112] V. Kac. *Infinite dimensional Lie algebras.* Cambridge U. Press, Cambridge, 1991.

[113] V. Kac. *Vertex algebras for Beginners.* American Mathematial Society, University Lecture Series 10, Providence, 1997.

[114] C. Kassel and V. Turaev. *Braid Groups.* Springer-Verlag, GTM 247 series, 2008.

[115] F. Klein. Zur Theorie der elliptischen Modulfunktionen. *Gesammente mathematische Abhandlungen*, 3:13–75, 1923.

[116] A. W. Knapp. *Lie Groups: Beyond an introduction.* Birkhauser, Berlin, 1996.

[117] N. Koblitz. *Introduction to Elliptic Curves and Modular Forms.* Springer-Verlag, GTM 97 series, 1984.

[118] M. Koike. On replication formulae and Hecke operators. *Nagoya University, preprint.*

[119] M. Koike. On McKay's conjecture. *Nagoya Math. J.*, 95:85–89, 1984.

[120] M. Koike. Mathieu group M_{24} and modular forms. *Nagoya Math. J.*, 99:147–157, 1985.

[121] M. Koike. Moonshine for $PSL_2(\mathbb{F}_7)$. *Adv. Stud. Pure Math.*, 7:103–111, 1985.

[122] M. Koike. Moonshines of $PSL_2(\mathbb{F}_q)$ and the automorphism group of Leech lattice. *Japan J. Math.*, 12:283–323, 1986.

[123] M. Koike. Modular forms and the automorphism group of Leech lattice. *Nagoya Math. J.*, 112:63–79, 1988.

[124] D. N. Kozlov. *On completely replicable functions and extremal poset theory.* MSc thesis, University of Lund, Sweden, 1994.

[125] R. Kultze. Elliptic genera and the moonshine module. *Math. Z.*, 223:463–471, 1996.

[126] M. L. Lang. On a question raised by Conway-Norton. *J. Math. Soc. Japan*, 41:263–284, 1989.

[127] S. Lang. *Algebra.* Benjamin/Cummings, California, 1984.

[128] J. M. Lee. *Introduction to Topological Manifolds.* Springer-Verlag, GTM 202 series, 2000.

[129] J. M. Lee. *Introduction to Smooth Manifolds.* Springer-Verlag, GTM 218 series, 2003.

[130] J. Leech. Some sphere packings in higher space. *Canadian J. Math*, 16:657–682, 1964.

[131] J. Leech. Notes on sphere packings. *Canadian J. Math.*, 19:251–267, 1967.

[132] J. Lepowsky. Euclidean Lie algebras and the modular function j. In *The Santa Cruz Conference on Finite Groups (Santa Cruz, 1979), Proc. Sympos. Pure Math.*, pages 567–570, Providence, 1980. American Mathematical Society.

[133] B. H. Lian and S. T. Yau. Arithmetic properties of mirror map and quantum coupling. *Comm. Math. Phys.*, 176:163–191, 1996.

[134] Elon Lages Lima. *Variedades Diferenciáveis.* Publicações matemáticas, Impa, Rio de Janeiro, 2008.

[135] K. Liu. On modular invariance and rigidity theorems. *J. Differential Geom.*, 41:343–396, 1995.

[136] F. J. MacWilliams and N. J. A. Sloane. *The theory of error correcting codes.* North-Holland, New York, 1977.

[137] M. Mahowald and M. Hopkins. The structure of 24 dimensional manifolds having normal bundles which lift to $BO[8]$. *Recent Progress in Homotopy Theory. Contemp. Math.*, 293:89–110, 2002.

[138] F. Malikov and V. Schechtman. Deformations of vertex algebras, quantum cohomology of toric varieties, and elliptic genus. *Comm. Math. Phys.*, 234:77–100, 2003.

[139] F. Malikov, V. Schechtman, and A. Vaintrob. Chiral de Rham complex. *Comm. Math. Phys.*, 204:439–473, 1999.

[140] G. Mason. On a system of elliptic modular forms attached to the large Mathieu group. *Nagoya Math. J.*, 118, 1990.

[141] E. Mathieu. Mémoire sur l'étude des functions de plusieures quantités, sur la manière de les formes et sur les substitutions qui les laissent invariables. *Crelle J.*, 6:241–323, 1861.

[142] J. McKay. Groups, singularities and finite groups. In *The Santa Cruz conferenceon Finite Groups*, pages 183–186, 1980.

[143] J. McKay, J. P. Harnad, and P. Winternitz. *Groups and Symmetries. From Neolithic Scots to John McKay.* CRM Proceedings and Lecture Notes, 47, Am. Math. Soc. Providence, 2009.

[144] J. McKay and A. Sebbar. *Proceedings on Moonshine and Related Topics.* CRM Proceedings and Lecture Notes, 30, Providence, 2001.

[145] J. McKay and A. Sebbar. Replicable functions: An introduction. In *Frontiers in Number Theory, Physics, and Geometry, II*, pages 373–386, Berlin, 2007. Springer-Verlag.

[146] J. McKay and D. Sevilla. Aplicación de la descomposición racional univariada a monstrous moonshine. *arXiv:0805.2311*, 2008.

[147] J. McKay and D. Sevilla. Decomposing replicable functions. *LMS J. Comput. Math.*, 11:146–171, 2008.

[148] J. McKay and H. Strauss. The q-series of monstrous moonshine and the decomposition of the head characters. *Commun. Alg.*, 18:253–278, 1990.

[149] R. V. Moody. Lie algebras associated with generalized Cartan matrices. *Bull. Amer. Math. Soc.*, 73:217–221, 1968.

[150] G. W. Moore. Les Houches lectures on strings and arithmetic. *arXiv: hepth/0401049.*

[151] D. R. Morrison. Picard-Fuchs equations and mirror maps for hypersurfaces. In *Essays on mirror manifolds*, pages 1241–264, Hong Kong, 1992. Intern. Press.

[152] H. Movasati. On differential modular forms and some analytic relations between Eisenstein series. *Ramanujan J.*, 17:53–76, 2008.

[153] H. V. Niemeier. Definite quadratische Formen der Dimension 24 und Diskriminante 1. *J. Number Theory*, 5:142–178, 1973.

[154] S. P. Norton. More on moonshine. In *Computational group theory*, pages 185–193, Cambridge, 1984. Cambridge University Press.

[155] S. P. Norton. Generalized moonshine. *Arcata Conference on Representations of Finite Groups. Proc. Sympos. Pure Math.*, 47:208–209, 1987.

[156] S. P. Norton. Constructing the Monster. In *Groups, Combinatorics and Geometry*, pages 63–76, Cambridge, 1992. Cambridge U. Press.

[157] S. P. Norton. From moonshine to the Monster. *Proc. on Moonshine and related topics*, Amer. Math. Soc, Providence:163–171, 2001.

[158] A. P. Ogg. Automorphismes des courbes modulaires. *Seminaire Delange-Pisot-Poitou, 16e annèe*, 7, 1975.

[159] L. Queen. Modular functions arising from some finite groups. *Math. Comp.*, 37:547–580, 1981.

[160] M. Ronan. *Symmetry and the Monster.* Oxford University Press, Oxford, 2006.

[161] A. J. E. Ryba. Modular moonshine? *Moonshine, the Monster, and related topics. Contemp. Math.*, 193:307–336, 1996.

[162] M. L. Lang S. P. Chan and C. H. Lim. Some modular functions associated to Lie algebra E_8. *Math. Z.*, 211:223–246, 1992.

[163] L. Schwartz. *Thèorie des distributions.* Hermann, Paris, 1973.

[164] G. Segal. Elliptic cohomology. *Seminaire Bourbaki 1987-88 no. 695*, 161-162:187–201, 1988.

[165] J.-P. Serre. *A course in arithmetic.* Springer-Verlag, GTM 7 series, 1973.

[166] J.-P. Serre. *Linear representations of finite groups.* Springer-Verlag, GTM 42 series, 1977.

[167] J. H. Silverman. *The Arithmetic of Elliptic Curves.* Springer-Verlag, GTM 106 series, 1986.

[168] J. Sotomayor. *Lições de equações diferenciais ordinárias.* Projeto Euclides, Impa, Rio de Janeiro, 1979.

[169] H. Tamanoi. Elliptic genera and vertex operator superalgebras. *Lecture Notes in Mathematics*, 1704, 1999.

[170] C. B. Thomas. *Elliptic cohomology.* Kluwer, New York, 1999.

[171] J. Thompson. Finite groups and modular functions. *Bull. London Math. Soc.*, 11:347–351, 1979.

[172] J. Thompson. Some numerology between the Fischer-Griess monster and modular functions. *Bull. London Math. Soc.*, 11:352–353, 1979.

[173] J. G. Thompson. A finiteness theorem for subgroups of $PSL_2(\mathbb{R})$ which are commensurable with $PSL_2(\mathbb{Z})$. In *Santa Cruz Conference on Finite Groups*, pages 533–555. Proc. Sympos. Pure Math. 37, 1980.

[174] J. Tits. Résumé de cours. *Annuaire du Collège de France, 1982-1983*, pages 89–102, 1983.

[175] J. Tits. On R. Griess' friendly giant. *Invent. Math.*, 78:491–499, 1984.

[176] J. Tits. Le monstre, Seminaire Bourbaki, exposé no. 620. *Astérisque*, 121-122:105–122, 1985.

[177] M. P. Tuite. On the relationship between Monstrous Moonshine and the uniqueness of the Moonshine module. *Comm. Math. Phys.*, 166:495–532, 1995.

[178] M. P. Tuite. Genus two meromorphic conformal field theories. In *Proc. on Moonshine and related topics*, pages 231–251, Providence, 2001. Amer. Math. Soc.

[179] S. Varadarajan. *Lie groups, Lie algebras and their representations.* Springer-Verlag, GTM 102 series, 1974.

[180] H. Verrill and N. Yui. Thompson series and the mirror maps of pencils of $K3$ surfaces. In *The arithmetic and geometry of algebraic cycles*, pages 399–432, Paris, 2000. Centre Res. Math. Proc. and Lecture Notes 24.

[181] M. Waldschmidt, P. Moussa, J. M. Luck, and C. Itzykson. *From Number Theory to Physics.* Springer Verlag, Berlin, 1992.

[182] Z.-X. Wan. *Introduction to Kac-Moody algebras.* World Scientific, Singapore, 1991.

[183] E. Witten. Elliptic genera and quantum field theory. *Comm. Math. Phys.*, 109, 1987.

[184] E. Witten. Physics and geometry. In *Proc. Intern. Congr. Math. (Berkeley, 1986)*, pages 267–301, Providence, 1987. Amer. Math. Soc.

[185] E. Witten. Quantum field theory, Grassmannians, and algebraic curves. *Comm. Math. Phys.*, 113:529–600, 1988.

[186] E. Witten. Three-Dimensional Gravity Reconsidered. *arXiv:0706.3359v1*, 2007.

[187] D. Zagier. Traces of singular moduli. In *Motives, Polylogarithms and Hodge Theory, Part I: Motives and Polylogarithms*, pages 211–244, Sommerville, 2002. International Press.

[188] Y. Zhu. Global vertex operators on Riemann surfaces. *Comm. Math. Phys.*, 165:485–531, 1994.

[189] Y. Zhu. Modular invariance of characters of vertex operator algebras. *J. Amer. Math. Soc.*, 9:237–302, 1996.

Appendix A

Proof of the Replication Formulae

This part is a continuation of Section 7.5. Our purpose is to give a detailed proof of the recursion formulas (7.19)-(7.22).

A.1 Faber polynomials

The Faber polynomials [60] originated in approximation theory in 1903 and are central to the theory of replicable functions. We define them in a formal way. The reader interested in analytical aspects of these polynomials can consult [60] or section 3 and 12 of [145]. Let $f : \mathbb{H} \to \mathbb{C}$ be a function having q-expansion

$$f(z) = \frac{1}{q} + \sum_{n \geq 0} H_n q^n,$$

where we take $q = e^{2\pi i z}$, for $z \in \mathbb{H}$, the upper half-plane. Throughout, we interpret derivatives of f with respect to q. We initially assume that the coefficients H_n of f are in $\mathbb{C}$ and we choose the constant term to be zero. For each $n \in \mathbb{Z}^+$, there exists a unique monic polynomial P_n in f, such that

$$P_n(f) = \frac{1}{q^n} + O(q) \quad \text{as } q \to 0,$$

(or equivalently, $P_n(f) \equiv q^{-n} \pmod{q\mathbb{Z}[q]}$). In fact, $P_n = P_n(f)$ depends on the coefficients of f, but we denote it simply by P_n when there is no confusion. The polynomial P_n is called the *n-th Faber polynomial* associated with f. It can be shown that the Faber polynomials are given by the generating series

$$\frac{qf'(q)}{z - f(q)} = \sum_{n \geq 0} P_n(z) q^n,$$

with $P_0(z) = 1$, $P_1(z) = z$, $P_2(z) = z^2 - 2H_1$, $P_3(z) = z^3 - 3H_1 - 3H_2$, and more generally:

$$P_n(z) = \det(zI - A_n),$$

where

$$A_n = \begin{pmatrix} H_0 & 1 & & & & \\ 2H_1 & H_0 & 1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ (n-2)H_{n-3} & H_{n-4} & H_{n-5} & \dots & 1 & \\ (n-1)H_{n-2} & H_{n-3} & H_{n-4} & \dots & H_0 & 1 \\ nH_{n-1} & H_{n-2} & H_{n-3} & \dots & H_1 & H_0 \end{pmatrix}.$$

It is useful to note that the Faber polynomials satisfy a Newton type recurrence relation of the form

$$P_{n+1}(z) = zP_n(z) - \sum_{k=1}^{n-1} H_{n-k}P_k(z) - (n+1)H_n, \tag{A.1}$$

for all $n \geq 1$.

Another useful way to see the Faber polynomials according to Norton [154] is the following. Consider f with the q-expansion

$$f(z) = \frac{1}{q} + \sum_{n\geq 0} H_n q^n, \quad q = e^{2\pi i z}.$$

We define some coefficients $H_{m,n}$ by the formula

$$F(y,z) = \log\big(f(y) - f(z)\big) = \log(p^{-1} + q^{-1}) - \sum_{m,n\geq 1} H_{m,n} q^m p^n,$$

(the *bivarial* transformation of f), where $p = e^{2\pi i y}$. Then $X_n(f) = \frac{1}{n}q^{-n} + \sum_m H_{m,n}q^m$ is the coefficient of p^n in the expansion

$$-\log p - \log\big(f(y) - f(z)\big),$$

so that it is a polynomial in f. In fact, $X_n(f) = \frac{1}{n}P_n(f)$, and we can see X_n as a (non-monic) n-th Faber polynomial associated to f.

Example A.1.1. If $f(z)$ is a modular form of weight $2k$ on $SL_2(\mathbb{Z})$, then the Hecke operators T_n, $n \geq 1$ act on f as

$$T_n(f)(z) = n^{k-1} \sum_{ad=n,\, 0\leq b<d} \frac{1}{d^k} f\Big(\frac{az+b}{d}\Big),$$

for all $n \geq 1$. See Chapter 6 and Section 7.1, or the first five chapter of [181] (especially Zagier's article) for background details. When $k = 0$ and $f(z)$ is the j-function, we have

$$T_n(j)(z) = \frac{1}{n} \sum_{ad=n,\, 0\leq b<d} j\Big(\frac{az+b}{d}\Big).$$

The generators of $SL_2(\mathbb{Z})$ permute the linear fractional transformations in the sum, hence $T_n(j)$ is invariant under $SL_2(\mathbb{Z})$. Since $T_n(j)$ has no poles in the upper half-plane $\mathbb{H}$, it follows that it is a polynomial in j (see Section 7.1). We find that,

$$T_n(f)(z) = \frac{1}{q^n} + O(q) \ \text{ as } q \to 0,$$

for all $n \geq 1$, and so $T_n(j) = \frac{1}{n} P_n(j)$. Thus the Hecke operator of the j-function are examples of Faber polynomials.

A.2 Replicable functions

Let f be a function having q-expansion

$$f(z) = \frac{1}{q} + \sum_{n \geq 0} H_n q^n,$$

where $q = e^{2\pi i z}$. Such a function is called *replicable* if there exist a sequence $\{f^{(s)}\}_{s \in \mathbb{Z}^+}$ of functions $f^{(s)} : \mathbb{H} \to \mathbb{C}$, such that for all $n \geq 1$, the expression

$$P_n(f) = \sum_{ad=n,\, 0 \leq b < d} f^{(a)}\Big(\frac{az+b}{d}\Big). \tag{A.2}$$

The function $f^{(s)}$ introduced above is called the *s-th replication power* of f.

Thus, we can think this polynomial $P_n(f)$ as the action of a generalized Hecke operator. Following Norton [154], let us consider the coefficients $\{H_{m,n}\}_{m,n \geq 1}$ introduced in previous section by

$$X_n(f) = \frac{1}{n} q^{-n} + \sum_{m \geq 1} H_{m,n} q^m, \tag{A.3}$$

for $n \geq 1$ (note in particular that $H_{n,1} = H_n$), and the slightly modified ones

$$h_{m,n} = (m+n) H_{m,n}.$$

It follows from equation (A.1) that this $h_{m,n}$ are given recursively by

$$h_{m,n} = (m+n) H_{m+n-1} + \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} H_{i+j-1} h_{m-i,n-j}.$$

A useful characterization for replicable functions proved by Norton in [154] is the following:

Theorem A.2.1. *The function f is replicable if, and only if, $H_{m,n} = H_{r,s}$ for all positive integers m, n, r, s satisfying $mn = rs$ and $(m, n) = (r, s)$.*

Let $H_n^{(s)}$ be the coefficient of q^n in the s-th replication power of f, that is

$$f^{(s)}(z) = \frac{1}{q} + \sum_{n\geq 0} H_n^{(s)} q^n,$$

for all $n, s \geq 1$ (in particular, $H_n^{(1)} = H_n$). According to Proposition 7.4.3 and Theorem 7.2.2, equation (A.2) can be written as

$$H_{m,n} = \sum_{s|(m,n)} \frac{1}{s} H^{(s)}_{mn/s^2}. \tag{A.4}$$

Applying the Möbius inversion formula to equation, then we have

$$H_n^{(s)} = s \sum_{d|s} \mu(d) H_{dsn,\frac{s}{d}}, \tag{A.5}$$

where μ is the Möbius function.

Another result obtained by Norton [154] is:

Theorem A.2.2. *Suppose that f is replicable. For any $k \in \mathbb{Z}^+$, as s ranges over divisors of k, the following are equivalent:*
1. $f^{(s)}$ is replicable.
2. The bivarial transform of $f^{(s)}$ is the generating function of $H^{(s)}_{m,n}$.
3. In addition to condition (1), the t-th replication power of $f^{(s)}$ is $f^{(st)}$.

Other results about replicable functions can be found in [124] and [147].

A.3 Deduction of the replication formulae

Let $f(z)$ be given by the series

$$f(z) = \frac{1}{q} + \sum_{n\geq 1} H_n q^n,$$

where $q = e^{2\pi i z}$. Define $X_n(f)$ as in (A.3) (observe in particular that $X_1(f) = \frac{1}{q} + \sum H_{n,1} q^n = f(z)$. We will write $X_1 = f$.

We have already mentioned that X_n is the coefficient of p^n in the expansion of $\log p^{-1} - \log\big(f(y) - f(z)\big)$, where $p = e^{2\pi i y}$. Writing

$$\log p^{-1} - \sum_{n\geq 1} X_n p^n = \log\big(f(y) - f(z)\big),$$

it follows that

$$p^{-1}\exp\Big(-\sum_{n\geq 1}X_np^n\Big)=p^{-1}+\sum_{n\geq 1}H_np^n-q^{-1}-\sum_{n\geq 1}H_nq^n,$$

so

$$\exp\Big(-\sum_{n\geq 1}X_np^n\Big)=1+p\Big(-q^{-1}-\sum_{n\geq 1}H_nq^n\Big)+\sum_{n\geq 1}H_np^{n+1}. \tag{A.6}$$

Using the Taylor expansion $\exp\big(-\sum X_np^n\big)=\sum_{k\geq 0}\frac{1}{k!}\big(-\sum X_np^n\big)^k$, and comparing the coefficients of p^2, p^3 and p^4 in (A.6), we obtain

$$\begin{aligned} H_1 &= \tfrac{1}{2}\Big(-2X_2+X_1^2\Big)=\tfrac{1}{2}\Big(-2X_2+f^2\Big); && \text{(A.7)}\\ H_2 &= \tfrac{1}{6}\Big(-6X_3+6X_2f-f^3\Big); && \text{(A.8)}\\ H_3 &= \tfrac{1}{24}\Big(-24X_4+24X_3f+12X_2^2-12X_2f^2+f^4\Big). && \text{(A.9)}\end{aligned}$$

Consider $f^{(s)}=\frac{1}{q}+\sum_{n\geq 1}H_n^{(s)}q^n$, the replication powers of f. Also, define the linear operator U_n, $n\geq 1$, such that for any series of the form $\sum_{\ell\in\mathbb{Z}}a_\ell q^\ell$, we have

$$\Big(\sum_{\ell\in\mathbb{Z}}a_\ell q^\ell\Big)\Big|_{U_n}=n\sum_{\ell\in\mathbb{Z}}a_{n\ell}q^\ell.$$

Then, Koike [118] proved the following formulas called 2-plication and 4-plication, respectively:

$$\begin{aligned} X_2(f) &= \tfrac{1}{2}\big(f|_{U_2}+f^{(2)}(2z)\big); && \text{(A.10)}\\ X_4(f) &= \tfrac{1}{4}\big(f|_{U_4}+f^{(2)}|_{U_2}(2z)+f^{(4)}(4z)\big). && \text{(A.11)}\end{aligned}$$

In fact, both these formulas can be obtained from the recursion equation (A.4) as follows. From (A.4) we have

$$H_{n,2}=\begin{cases} H_{2n}+\frac{1}{2}H_{n/2}^{(2)}, & n\equiv 0 \pmod 2;\\ H_{2n}, & n\equiv 1 \pmod 2;\end{cases}$$

and

$$H_{n,4}=\begin{cases} H_{4n}+\frac{1}{2}H_n^{(2)}+\frac{1}{4}H_{n/4}^{(4)}, & n\equiv 0 \pmod 4;\\ H_{4n}+\frac{1}{2}H_n^{(2)}, & n\equiv 2 \pmod 4;\\ H_{4n}, & n\equiv 1,3 \pmod 4).\end{cases}$$

Hence,

$$\begin{aligned}
\tfrac{1}{2}\big(f|_{U_2} + f^{(2)}(2z)\big) &= \sum_{n\geq 1} H_{2n}q^n + \frac{1}{2}\Big(\frac{1}{q^2} + \sum_{n\geq 1} H_n^{(2)} q^{2n}\Big) \\
&= \frac{1}{2}q^{-2} + \sum_{n\geq 1} H_{2n}q^n + \sum_{n\geq 1}\frac{1}{2}H_n^{(2)}q^{2n} \\
&= \frac{1}{2}q^{-2} + \sum_{n\geq 1} H_{n,2}q^n \\
&= X_2(f),
\end{aligned}$$

and

$$\begin{aligned}
\tfrac{1}{4}\big(f|_{U_4} + f^{(2)}|_{U_2}(2z) + f^{(4)}(4z)\big) &= \sum_{n\geq 1} H_{4n}q^n + \frac{1}{2}\sum_{n\geq 1} H_{2n}^{(2)}q^{2n}\Big) + \frac{1}{4}\Big(\frac{1}{q^4} + \sum_{n\geq 1} H_n^{(4)}q^{4n}\Big) \\
&= \frac{1}{4}q^{-4} + \sum_{n\geq 1} H_{4n}q^n + \sum_{n\geq 1}\frac{1}{2}H_{2n}^{(2)}q^{2n} + \sum_{n\geq 1}\frac{1}{4}H_n^{(4)}q^{4n} \\
&= \frac{1}{4}q^{-4} + \sum_{n\geq 1} H_{n,4}q^n \\
&= X_4(f).
\end{aligned}$$

Using these Koike's formulas, from (A.7) and (A.10) we obtain

$$H_1 = \tfrac{1}{2}\Big(f^2 - f|_{U_2} - f^{(2)}(2z)\Big); \tag{A.12}$$

and from (A.7)-(A.11):

$$\begin{aligned}
H_3 + \tfrac{1}{2}H_1^2 &= \Big(\tfrac{1}{24}f^4 - \tfrac{1}{2}X_2f^2 + \tfrac{1}{2}X_2^2 + X_3f - X_4\Big) + \tfrac{1}{8}\Big(f^2 - f|_{U_2} - f^{(2)}(2z)\Big)^2 \\
&= \Big(\tfrac{1}{24}f^4 - \tfrac{1}{4}f|_{U_2}f^2 - \tfrac{1}{4}f^{(2)}(2z)f^2 + \tfrac{1}{8}(f|_{U_2})^2 + \tfrac{1}{4}f|_{U_2}f^{(2)}(2z) + \\
&\quad + \tfrac{1}{8}(f^{(2)}(2z))^2 + X_3f - \tfrac{1}{4}f|_{U_4} - \tfrac{1}{4}f^{(2)}|_{U_2}(2z) - \tfrac{1}{4}f^{(4)}(4z)\Big) + \Big(\tfrac{1}{8}f^4 - \\
&\quad - \tfrac{1}{4}f|_{U_2}f^2 - \tfrac{1}{4}f^{(2)}(2z)f^2 + \tfrac{1}{8}(f|_{U_2})^2 + \tfrac{1}{4}f|_{U_2}f^{(2)}(2z) + \tfrac{1}{8}(f^{(2)}(2z))^2\Big) \\
&= \Big(\tfrac{1}{6}f^4 - X_2f^2 + X_3f\Big) - \tfrac{1}{4}f|_{U_4} - \tfrac{1}{4}f^{(2)}|_{U_2}(2z) - \tfrac{1}{4}f^{(4)}(4z) + \\
&\quad + \tfrac{1}{4}(f|_{U_2})^2 + \tfrac{1}{2}f|_{U_2}f^{(2)}(2z) + \tfrac{1}{4}(f^{(2)}(2z))^2 \\
&= -H_2f - \tfrac{1}{4}f|_{U_4} + \tfrac{1}{2}f|_{U_2}f^{(2)}(2z) + \tfrac{1}{4}(f|_{U_2})^2 + \\
&\quad + \tfrac{1}{4}\Big((f^{(2)}(2z))^2 - f^{(2)}|_{U_2}(2z) - f^{(4)}(4z)\Big) \\
&= -H_2f - \tfrac{1}{4}f|_{U_4} + \tfrac{1}{2}f|_{U_2}f^{(2)}(2z) + \tfrac{1}{4}(f|_{U_2})^2 + \tfrac{1}{2}H_1^{(2)},
\end{aligned}$$

thus, we have

$$H_3 + \tfrac{1}{2}H_1^2 - \tfrac{1}{2}H_1^{(2)} = \tfrac{1}{4}(f|_{U_2})^2 + \tfrac{1}{2}f|_{U_2}f^{(2)}(2z) - H_2 f - \tfrac{1}{4}f|_{U_4}. \tag{A.13}$$

If we compare the coefficients of q^{2k} and q^{2k+1} (for $k \geq 1$) of both sides in (A.12), and carry out some calculation, we find:

$$\begin{aligned}
H_1 &= \tfrac{1}{2}\Big(f^2 - f|_{U_2} - f^{(2)}(2z)\Big) \\
&= \frac{1}{2}\Big(\frac{1}{q} + \sum_{n\geq 1} H_n q^n\Big)^2 - \frac{1}{2}\Big(2\sum_{n\geq 1} H_{2n}q^n\Big) - \frac{1}{2}\Big(\frac{1}{q^2} + \sum_{n\geq 1} H_n^{(2)}q^{2n}\Big) \\
&= \frac{1}{2q^2} + \frac{1}{q}\sum_{n\geq 1} H_n q^n + \frac{1}{2}\sum_{n\geq 1}\sum_{j=1}^{n-1} H_j H_{n-j}q^n - \sum_{n\geq 1} H_{2n}q^n - \frac{1}{2q^2} - \frac{1}{2}\sum_{n\geq 1} H_n^{(2)}q^{2n} \\
&= H_1 + \sum_{n\geq 1}\Big(H_{n+1} + \frac{1}{2}\sum_{1\leq j<n} H_j H_{n-j} - H_{2n}\Big)q^n - \frac{1}{2}\sum_{n\geq 1} H_n^{(2)}q^{2n}.
\end{aligned}$$

Thus, for $n = 2k$ and $n = 2k+1$ we have the following relations

$$\begin{aligned}
n = 2k: &\qquad H_{2k+1} + \frac{1}{2}\sum_{1\leq j<2k} H_j H_{2k-j} - H_{4k} - \frac{1}{2}H_k^{(2)} = 0; \\
n = 2k+1: &\qquad H_{2k+2} + \frac{1}{2}\sum_{1\leq j\leq 2k} H_j H_{2k-j+1} - H_{4k+2} = 0.
\end{aligned}$$

Since $\sum_{1\leq j<2k} H_j H_{2k-j} = 2\sum_{1\leq j<k} H_j H_{2k-j} + H_k^2$ and $\sum_{1\leq j\leq 2k} H_j H_{2k-j+1} = 2\sum_{1\leq j\leq k} H_j H_{2k-j}$, in particular, we obtain our first two replicable formulas

$$H_{4k} = H_{2k+1} + \sum_{1\leq j<k} H_j H_{2k-j} + \frac{1}{2}\Big(H_k^2 - H_k^{(2)}\Big). \tag{A.14}$$

$$H_{4k+2} = H_{2k+2} + \sum_{1\leq j\leq k} H_j H_{2k-j+1}. \tag{A.15}$$

Develop the other two remaining replication formulas is a bit more demanding. Applying (A.14) to $2k, k+1$ and $2k+1$ in place of k we obtain

$$\begin{aligned}
H_{8k} &= H_{4k+1} + \sum_{1\leq j<2k} H_j H_{4k-j} + \frac{1}{2}\Big(H_{2k}^2 - H_{2k}^{(2)}\Big), &\text{(A.16)} \\
H_{4k+4} &= H_{2k+3} + \sum_{1\leq j\leq k} H_j H_{2k-j+2} + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big), &\text{(A.17)} \\
H_{8k+4} &= H_{4k+3} + \sum_{1\leq j\leq 2k} H_j H_{4k-j+2} + \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big); &\text{(A.18)}
\end{aligned}$$

and replacing $k+1$ for k in (A.15), we also have

$$H_{4k+6} = H_{2k+4} + \sum_{1\le j\le k} H_j H_{2k-j+3}. \tag{A.19}$$

Now, comparing the coefficients of q^{2k} and q^{2k+1} (for $k \ge 1$) of both sides in (A.13), and doing some calculations, we find:

$$\begin{aligned}
H_3 + \tfrac{1}{2}H_1^2 - \tfrac{1}{2}H_1^{(2)} &= \tfrac{1}{4}(f|_{U_2})^2 + \tfrac{1}{2}f|_{U_2}f^{(2)}(2z) - H_2 f - \tfrac{1}{4}f|_{U_4} \\
&= \frac{1}{4}\Big(2\sum_{n\ge1} H_{2n}q^n\Big)^2 + \frac{1}{2}\Big(2\sum_{n\ge1} H_{2n}q^n\Big)\Big(\frac{1}{q^2} + \sum_{n\ge1} H_n^{(2)}q^{2n}\Big) - \\
&\quad -H_2\Big(\frac{1}{q} + \sum_{n\ge1} H_n q^n\Big) - \frac{1}{4}\Big(4\sum_{n\ge1} H_{4n}q^n\Big) \\
&= \sum_{n\ge1}\sum_{j=1}^{n-1} H_{2j}H_{2n-2j}q^n + \frac{1}{q^2}\sum_{n\ge1} H_{2n}q^n + \Big(\sum_{k\ge1}\sum_{j=0}^{k-1} H_{4j+2}H_{k-j}^{(2)}q^{2k+1} + \\
&\quad + \sum_{k\ge1}\sum_{j=0}^{k-2} H_{4j+4}H_{k-j+1}^{(2)}q^{2k}\Big) - \frac{1}{q}H_2 - \sum_{n\ge1} H_2H_nq^n - \sum_{n\ge1} H_{4n}q^n \\
&= H_4 + \sum_{n\ge1}\Big(H_{2n+4} + \sum_{1\le j<n} H_{2j}H_{2n-2j} - H_2H_n - H_{4n}\Big)q^n + \\
&\quad + \Big(\sum_{k\ge1}\sum_{j=0}^{k-1} H_{4j+2}H_{k-j}^{(2)}q^{2k+1} + \sum_{k\ge1}\sum_{j=0}^{k-2} H_{4j+4}H_{k-j+1}^{(2)}q^{2k}\Big)
\end{aligned}$$

Thus, taking $n = 2k$ and $n = 2k+1$ above, we have the following relations

$$n = 2k: \qquad H_{4k+4} + \sum_{1\le j<2k} H_{2j}H_{4k-2j} + \sum_{j=0}^{k-2} H_{4j+4}H_{k-j+1}^{(2)} - H_2H_{2k} - H_{8k} = 0;$$

$$n = 2k+1: \qquad H_{4k+6} + \sum_{1\le j\le 2k} H_{2j}H_{4k-2j+2} + \sum_{k\ge1}\sum_{j=0}^{k-1} H_{4j+2}H_{k-j}^{(2)} - H_2H_{2k+1} - H_{8k+4} = 0.$$

Replacing (A.16) and (A.17) in the expression for $2k$ above, we have

$$\begin{aligned}
&H_{2k+3} + \sum_{1\le j\le k} H_jH_{2k-j+2} + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big) + \sum_{1\le j<2k} H_{2j}H_{4k-2j} + \\
&+ \sum_{j=0}^{k-2} H_{4j+4}H_{k-j+1}^{(2)} - H_2H_{2k} - H_{4k+1} - \sum_{1\le j<2k} H_jH_{4k-j} - \frac{1}{2}\Big(H_{2k}^2 - H_{2k}^{(2)}\Big) \;=\; 0.
\end{aligned}$$

Thus,

$$\begin{aligned}
H_{4k+1} &= H_{2k+3} - H_2 H_{2k} - \frac{1}{2}\Big(H_{2k}^2 - H_{2k}^{(2)}\Big) + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big) + \\
&\quad + H_{2k}^2 + 2\sum_{1\le j<k} H_{2j}H_{4k-2j} + \sum_{1\le j\le k} H_j H_{2k-j+2} + \sum_{1\le j<k} H_{4k-4j}H_j^{(2)} - \\
&\quad - \sum_{1\le j<2k} H_j H_{4k-j} \\
&= H_{2k+3} - H_2 H_{2k} + \frac{1}{2}\Big(H_{2k}^2 + H_{2k}^{(2)}\Big) + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big) + \\
&\quad \sum_{1\le j\le k} H_j H_{2k-j+2} + \sum_{1\le j<k} H_{4k-4j}H_j^{(2)} + \Big(2\sum_{1\le j<k} H_{2j}H_{4k-2j} - \sum_{1\le j<2k} H_j H_{4k-j}\Big) \\
&= H_{2k+3} - H_2 H_{2k} + \frac{1}{2}\Big(H_{2k}^2 + H_{2k}^{(2)}\Big) + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big) + \\
&\quad \sum_{1\le j\le k} H_j H_{2k-j+2} + \sum_{1\le j<k} H_{4k-4j}H_j^{(2)} + \sum_{1\le j<2k} (-1)^j H_j H_{4k-j}\Big).
\end{aligned}$$

Similarly, replacing (A.18) and (A.19) in the expression for $2k+1$ above, we obtain

$$\begin{aligned}
H_{2k+4} &+ \sum_{1\le j\le k+1} H_j H_{2k-j+3} + \sum_{1\le j\le 2k} H_{2j}H_{4k-2j+2} + \sum_{0\le j<k} H_{4j+2}H_{k-j}^{(2)} - \\
&- H_2 H_{2k+1} - H_{4k+3} - \sum_{1\le j\le 2k} H_j H_{4k-j+2} - \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big) \quad = \quad 0.
\end{aligned}$$

Thus,

$$\begin{aligned}
H_{4k+3} &= H_{2k+4} - H_2 H_{2k+1} - \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big) + \sum_{1\le j\le k+1} H_j H_{2k-j+3} + \\
&\quad + \sum_{0\le j<k} H_{4j+2}H_{k-j}^{(2)} + \sum_{1\le j\le 2k} H_{2j}H_{4k-2j+2} - \sum_{1\le j\le 2k} H_j H_{4k-j+2} \\
&= H_{2k+4} - H_2 H_{2k+1} - \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big) + \sum_{1\le j\le k+1} H_j H_{2k-j+3} + \\
&\quad + \sum_{1\le j\le k} H_{4k-4j+2}H_j^{(2)} + \Big(2\sum_{1\le j\le k} H_{2j}H_{4k-2j+2} - \sum_{1\le j\le 2k} H_j H_{4k-j+2}\Big) \\
&= H_{2k+4} - H_2 H_{2k+1} - \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big) + \sum_{1\le j\le k+1} H_j H_{2k-j+3} + \\
&\quad + \sum_{1\le j\le k} H_{4k-4j+2}H_j^{(2)} + \sum_{1\le j\le 2k} (-1)^j H_j H_{4k-j+2},
\end{aligned}$$

and we have

$$\begin{aligned} H_{4k+1} &= H_{2k+3} - H_2 H_{2k} + \frac{1}{2}\Big(H_{2k}^2 + H_{2k}^{(2)}\Big) + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big) + \\ &\quad + \sum_{1\le j\le k} H_j H_{2k-j+2} + \sum_{1\le j<k} H_{4k-4j} H_j^{(2)} + \sum_{1\le j<2k} (-1)^j H_j H_{4k-j}. \qquad \text{(A.20)} \end{aligned}$$

and

$$\begin{aligned} H_{4k+3} &= H_{2k+4} - H_2 H_{2k+1} - \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big) + \sum_{1\le j\le k+1} H_j H_{2k-j+3} + \\ &\quad + \sum_{1\le j\le k} H_{4k-4j+2} H_j^{(2)} + \sum_{1\le j\le 2k} (-1)^j H_j H_{4k-j+2}. \qquad \text{(A.21)} \end{aligned}$$

Hence, we have obtained the following four replication formulas (A.14), (A.20), (A.15), (A.21):

$$\begin{aligned} H_{4k} &= H_{2k+1} + \sum_{1\le j<k} H_j H_{2k-j} + \frac{1}{2}\Big(H_k^2 - H_k^{(2)}\Big); \\ H_{4k+1} &= H_{2k+3} - H_2 H_{2k} + \frac{1}{2}\Big(H_{2k}^2 + H_{2k}^{(2)}\Big) + \frac{1}{2}\Big(H_{k+1}^2 - H_{k+1}^{(2)}\Big) + \\ &\quad + \sum_{1\le j\le k} H_j H_{2k-j+2} + \sum_{1\le j<k} H_{4k-4j} H_j^{(2)} + \sum_{1\le j<2k} (-1)^j H_j H_{4k-j}; \\ H_{4k+2} &= H_{2k+2} + \sum_{1\le j\le k} H_j H_{2k-j+1}; \\ H_{4k+3} &= H_{2k+4} - H_2 H_{2k+1} - \frac{1}{2}\Big(H_{2k+1}^2 - H_{2k+1}^{(2)}\Big) + \sum_{1\le j\le k+1} H_j H_{2k-j+3} + \\ &\quad + \sum_{1\le j\le k} H_{4k-4j+2} H_j^{(2)} + \sum_{1\le j\le 2k} (-1)^j H_j H_{4k-j+2}. \end{aligned}$$

To obtain the set of equations (7.19)-(7.22), it is just a change of notation. Recall that we write $c_g(n)$ to be the coefficient of q^n in the McKay-Thompson series $T_g(z)'$ of $g \in \mathbb{M}$, that is $T_g(z)' = q^{-1} + \sum_{n\ge1} c_g(n) q^n$, with $q = e^{2\pi i z}$. Also, recall Conway and Norton already proved in [38] that the s-th replication associated to $T_g(z)'$ is exactly $T_{g^s}(z)'$, the McKay-Thompson series associated to g^s. Hence, taking $f = T_g'$, we have the following dictionary

$$H_n = c_g(n) \text{ and } H_n^{(2)} = c_{g^2}(n),$$

from where the replication formulae (7.19)-(7.22) directly appear.

Also, note that when comparing coefficients of q^{2k} and q^{2k+1} in (A.13), we have obtained in addition

$$H_4 = H_3 + \tfrac{1}{2}H_1^2 - \tfrac{1}{2}H_1^{(2)},$$

that is, equation (7.23), using notation $c_g(n)$.

Index