Examensarbete

Modular forms and converse theorems for Dirichlet series

Jonas Karlsson

LiTH - MAT - EX - - 2009 / 05 - - SE
Modular forms and converse theorems for Dirichlet series

Applied Mathematics, Linköpings Universitet

Jonas Karlsson

LiTH - MAT - EX - - 2009 / 05 - - SE

Examensarbete: 45 hp

Level: D

Supervisor: M. Izquierdo,
Applied Mathematics, Linköpings Universitet

Examiner: M. Izquierdo,
Applied Mathematics, Linköpings Universitet

Linköping: June 2009
This thesis makes a survey of converse theorems for Dirichlet series. A converse theorem gives sufficient conditions for a Dirichlet series to be the Dirichlet series attached to a modular form. Such Dirichlet series have special properties, such as a functional equation and an Euler product. Sometimes these properties characterize the modular form completely, i.e. they are sufficient to prove the proper transformation behaviour under some discrete group. The problem dates back to Hecke and Weil, and has more recently been treated by Conrey et.al. The articles surveyed are:

- “An extension of Hecke’s converse theorem”, by B. Conrey and D. Farmer
- “Converse theorems assuming a partial Euler product”, by D. Farmer and K. Wilson
- “A converse theorem for \( \Gamma_0(13) \)”, by B. Conrey, D. Farmer, B. Odgers and N. Snaith

The results and the proofs are described. The second article is found to contain an error. Finally an alternative proof strategy is proposed.
Abstract

This thesis makes a survey of converse theorems for Dirichlet series. A converse theorem gives sufficient conditions for a Dirichlet series to be the Dirichlet series attached to a modular form. Such Dirichlet series have special properties, such as a functional equation and an Euler product. Sometimes these properties characterize the modular form completely, i.e. they are sufficient to prove the proper transformation behaviour under some discrete group. The problem dates back to Hecke and Weil, and has more recently been treated by Conrey et.al. The articles surveyed are:

- “An extension of Hecke’s converse theorem”, by B. Conrey and D. Farmer
- “Converse theorems assuming a partial Euler product”, by D. Farmer and K. Wilson
- “A converse theorem for $\Gamma_0(13)$”, by B. Conrey, D. Farmer, B. Odgers and N. Snaith

The results and the proofs are described. The second article is found to contain an error. Finally an alternative proof strategy is proposed.

Keywords: Modular forms, Dirichlet series, converse theorems, Hecke groups, Euler products, elliptic curves
Acknowledgements

First and foremost, I want to thank my supervisor and examiner Milagros Izquierdo Barrios, who has been a great source of inspiration and support. Her invaluable help has by no means been limited to the writing of this thesis. Again, thank you!

I would also like to thank my opponent Johan Hemström, who has worked his way through several rough versions of this thesis and caught several errors. This thesis is written in \LaTeX, and I will take this opportunity to express my gratitude to Donald Knuth for his creation. Several figures, and the calculations on elliptic curves, have been done with the free mathematics software Sage, and I am most grateful to William Stein and all other contributors in the free software community - for this relief much thanks! I wish to thank my mother for encouraging me to do whatever I want. Thanks to Erik Aas for extensive computer support, intelligent questions, and much else besides. Finally, but no less importantly, $2^{691} + 1$ thanks to Karin Lundengård for distraction.
Nomenclature

Symbols

\( \mathbb{N} \) The natural numbers; 0 ∈ \( \mathbb{N} \)
\( \mathbb{R} \) The real numbers
\( \mathbb{C} \) The complex numbers
\( k^* \) The nonzero elements of the field \( k \)
\( \overline{k} \) An algebraic closure of \( k \)
\( \text{char}(k) \) The characteristic of \( k \)
\( \mathbb{R}_+ \) The positive real numbers
\( \mathbb{C}P^1 \) The complex projective line
\( \mathbb{Q}P^1 \) The rational projective line
\( \hat{\mathbb{C}} \) The extended complex plane (line)
\( \hat{\mathbb{R}} \) The extended real line
\( \mathcal{H} \) The upper half-plane
\( \mathbb{Z}_p \) p-adic integers
\( \mathbb{Q}_p \) p-adic numbers
\( \mathbb{Z}/N\mathbb{Z} \) The integers modulo \( N \)
\( \Re z \) The real part of the complex number \( z \)
\( \Im z \) The imaginary part of the complex number \( z \)
(\( \subseteq \)) Subset, not necessarily proper
(\( A \setminus B \)) Set difference
\( S^1 \) Circle (topological)
(\( (a:b) \)) Projective coordinates
(\( (m,n) \)) GCD of \( m \) and \( n \)
\( A_n \) The alternating group on \( n \) letters
\( C_n \) The cyclic group of order \( n \)
\( S_n \) The symmetric group on \( n \) letters
\( GL \) The general linear group
\( SL \) The special linear group
\( PGL \) The projective general linear group
\( PSL \) The projective special linear group
\( SU \) The special unitary group
\begin{itemize}
\item \textit{Aut}(X) \quad \text{The automorphism group of the space } X
\item \textit{Stab}(X) \quad \text{The stabilizer subgroup of } X \text{ under a group action}
\item \(g.x\) \quad \text{The image of } x \text{ under the action of } g
\item \(\tilde{X}\) \quad \text{The universal covering space of } X
\item \text{tr} \quad \text{The trace of an operator}
\item \(f|_{\gamma}\) \quad \text{Slash operator, see definition on page 63}
\item \(I\) \quad \text{Identity matrix; identity M"obius transformation}
\item \(\varphi\) \quad \text{Euler's totient function}
\end{itemize}
Contents

1 Introduction 1
  1.1 Background .................................. 1
  1.2 Chapter outline .............................. 2

2 Riemann surfaces 3
  2.1 The Riemann sphere .......................... 5
  2.2 Möbius transformations ...................... 6
    2.2.1 Transitivity properties .................. 8
    2.2.2 Stabilizer subgroups ...................... 8
    2.2.3 Classification of Möbius transformations 9
  2.3 Hyperbolic geometry .......................... 10
  2.4 Groups of Möbius transformations ............ 12
    2.4.1 Continuous groups ....................... 12
    2.4.2 Finite groups of Möbius transformations 13
    2.4.3 Infinite discrete groups of Möbius transformations 13
  2.5 Group actions, covering spaces and uniformization 15
    2.5.1 Group actions ............................ 15
    2.5.2 Covering spaces ........................... 16
    2.5.3 Quotient surfaces ........................ 18
    2.5.4 The Uniformization theorem ............... 19

3 Lattices and Elliptic Curves 23
  3.1 Lattices .................................... 23
    3.1.1 Eisenstein series ......................... 27
  3.2 Elliptic functions ........................... 27
  3.3 Elliptic curves ................................ 29
    3.3.1 History .................................. 30
    3.3.2 The most general form of the equation .... 30
    3.3.3 The group structure ....................... 31
    3.3.4 Algebra .................................. 32
    3.3.5 Elliptic curves over different fields ...... 33
    3.3.6 The L-series of an elliptic curve ........... 34
    3.3.7 Moduli spaces of elliptic curves .......... 35
    3.3.8 Complex multiplication ..................... 35
    3.3.9 Galois representations ..................... 36
    3.3.10 Hyperelliptic curves and Abelian varieties 36

# List of Figures

2.1 A torus ............................. 5
2.2 A loxodromic transformation. ................. 10
2.3 Hyperbolic geodesics ......................... 11
2.4 A hyperbolic triangle ....................... 11
2.5 An octahedral tesselation. ................... 17
2.6 An icosahedral tesselation. .................. 18
2.7 A fundamental domain for $\Gamma$. ............ 19
2.8 A tesselation induced by $\Gamma$. ............... 20
2.9 A fundamental domain for $\Gamma(2)$ .......... 20
2.10 Klein’s quartic. ........................... 21

3.1 A lattice and a sublattice of index 2 ............ 23
3.2 The square lattice .......................... 24
3.3 The hexagonal lattice ....................... 25
3.4 Some fundamental domains .................... 26
3.5 A Dirichlet fundamental domain and its tesselation .... 27
3.6 Addition on $y^2 = x^3 - x$ .................. 27
3.7 The curve $y^2 = x^3 - x$ over GF(101) ........... 34

5.1 The contour of integration. ................... 52
Chapter 1

Introduction

1.1 Background

In 1859, Riemann made his famous investigation of the zeta function which today bears his name. Among other things, he proved that it satisfies the functional equation

\[ \pi^{-s/2} \Gamma(\frac{s}{2}) \zeta(s) = \pi^{-(1-s)/2} \Gamma(\frac{1-s}{2}) \zeta(1-s), \]  

(1.1)

where \( \Gamma \) is the well-known gamma function from complex analysis. In 1921, Hamburger noted that (1.1), together with some growth assumptions, determines \( \zeta(s) \) up to a multiplicative constant [18]. Later, in 1936, Hecke [20] proved his converse theorem, which proves an analogous result for arbitrary functions satisfying a suitable functional equation. In modern vocabulary, he proved a converse theorem for the groups \( \Gamma \) and \( \Gamma_0(N) \) for \( N = 2, 3 \) and \( 4 \), where \( \Gamma \) now denotes the so-called modular group of Möbius transformations with integer coefficients and determinant equal to 1, and \( \Gamma_0(N) \) is the subgroup consisting of transformations

\[ z \mapsto \frac{az + b}{cz + d} \]  

(1.2)

where in addition \( N|c \). A converse theorem asserts that some function defined by a Dirichlet series arises from a modular form, which is a function transforming in a certain way under a group of Möbius transformations. Which group is determined by a functional equation satisfied by the Dirichlet series. The results obtained by Hecke were the best possible under the assumptions made (growth conditions, Dirichlet series, functional equation). In 1967, Weil studied the problem and managed to generalize the results by imposing additional assumptions [45]. More precisely, he assumed a functional equation not only for the Dirichlet series itself, but also for “twisted” Dirichlet series, which are constructed using multiplicative (Dirichlet) characters. In 1995, Conrey and Farmer made another generalization [5], this time by assuming that the Dirichlet series has an Euler product of a certain form. This allowed them to prove converse theorems for the groups \( \Gamma_0(N) \) for \( 5 \leq N \leq 12, 14 \leq N \leq 17, \) and \( N = 23 \). In 2005, this was further explored by Farmer and Wilson [15], who only assumed a partial Euler product. Finally, in 2006, Conrey, Farmer, Odgers, Karlsson, 2009.
and Snaith [6] proved a converse theorem for \( \Gamma_0(13) \), which was lacking from [5]. Again, the proof used the assumption of a partial Euler product.

This thesis surveys the theorems described, with emphasis on the last three articles. This is done in the last chapter. The first four chapters contain the necessary background material.

1.2 Chapter outline

The thesis is organized as follows:

Chapter 2: Riemann surfaces This chapter introduces Riemann surfaces and their automorphisms. For the so-called Riemann sphere, these are the Möbius transformations. Möbius transformations are isometries in a model of hyperbolic geometry. We classify Möbius transformations and study groups acting on surfaces, quotient surfaces, and uniformization. The main references for this chapter are [14], [23] and [27].

Chapter 3: Lattices and Elliptic Curves This chapter introduces discrete groups acting on the complex plane, so-called lattices, and the corresponding compact Riemann surfaces. These are called elliptic curves. They are not only geometric objects but have a group structure as well. We introduce the concept of moduli spaces of curves. The main references are [1], [13], [22], [24], [28], [37], [38], and [39].

Chapter 4: Dirichlet Series This chapter introduces Dirichlet series, the Mellin transform, and Euler products. The modern setting for these classical objects involves the notion of an adele, which is briefly described in the final section. The main references are [24], [36], [42] and [43].

Chapter 5: Modular Forms This chapter introduces modular forms, which are certain functions defined on the moduli space of elliptic curves. By the discussion in chapter 3, this is a set of lattices, parametrized by the upper half-plane. We introduce spaces of modular forms, determine their dimension, and describe the Hecke operators which act on these spaces. Finally, we mention some applications of modular forms. The main references are [1], [13], [36], [39].

Chapter 6: New Converse Theorems The final chapter surveys converse theorems for Dirichlet series. These theorems give sufficient conditions for Dirichlet series to arise from modular forms. This line of research was initiated by Hecke in 1936 [20] and continued by Weil in 1967 [45]. In 1995, Conrey and Farmer published the article “An extension of Hecke’s converse theorem” [5], which uses a different approach than Weil’s. In 2005 and 2006, the articles “Converse theorems assuming a partial Euler product” [15] by Farmer and Wilson, and “A converse theorem for \( \Gamma_0(13) \)” [6] by Conrey, Farmer, Odgers and Snaith further developed these ideas. The results by Hecke and Weil are mentioned, and the articles by Farmer et al. are described in some detail.
Chapter 2

Riemann surfaces

Riemann surfaces are of central importance in complex analysis and geometry. Intuitively, they are surfaces which locally look like the complex plane, allowing one to define analytic functions. For the general theory of Riemann surfaces, see [14]. We begin with some definitions.

Definition A surface $S$ is a (second countable) Hausdorff topological space, every point of which has a neighbourhood homeomorphic to an open disc in the complex plane $\mathbb{C}$. A pair $(\phi_i, U_i)$ of an open set $U_i \subseteq S$ and a homeomorphism $\phi_i : U_i \to \mathbb{C}$ is called a chart. A collection of charts, the open sets of which cover $S$, is called an atlas. If for any two charts $(\phi_i, U_i)$ and $(\phi_j, U_j)$, the so-called transition function

$$
\phi_i|_{U_i \cap U_j} \circ \phi_j^{-1}|_{U_i \cap U_j} : \phi_j(U_i \cap U_j) \to \phi_i(U_i \cap U_j)
$$

is analytic, the atlas is called analytic. Two analytic atlases are called compatible if their union is again an analytic atlas. Clearly this is an equivalence relation. An equivalence class of analytic atlases is called a complex structure. A Riemann surface, finally, is a surface equipped with a complex structure.

Note One often requires a Riemann surface to be connected as well. This will be assumed in the following.

The word “surface” indicates two-dimensionality, and indeed Riemann surfaces have dimension two over $\mathbb{R}$. It is, however, equally possible to regard them as one-dimensional complex manifolds, that is, as curves. In fact, this point of view is implicit in the definition above, where a surface was defined by a local identification with the complex “plane”. A more natural definition from the real perspective would be to define a Riemann surface as a two-dimensional real manifold with a conformal structure, that is, a notion of angles between curves. One can show that these definitions are equivalent.

The natural notion of equality between Riemann surfaces is biholomorphic (also called conformal) equivalence: a map $f : S_1 \to S_2$ between Riemann surfaces is called holomorphic if the corresponding map $f_*$ from $\mathbb{C}$ to itself is:

Where $f_*$ is $\phi_2 \circ f \circ \phi_1^{-1}$. If in addition $f$ is bijective, $S_1$ and $S_2$ are called biholomorphically equivalent; it turns out that the inverse is automatically holomorphic as well, so this is an equivalence relation. Next, we give some examples of Riemann surfaces.

**Example 1** The simplest example of a Riemann surface is $\mathbb{C}$ itself, which looks like $\mathbb{C}$ not only locally but globally as well. It is non-compact and simply connected (has trivial homotopy group).

**Example 2** Any open, simply connected, proper subset of $\mathbb{C}$ is a Riemann surface. By the Riemann mapping theorem, any such set is biholomorphically equivalent to the open unit disc $\{z \in \mathbb{C} : |z| < 1\}$.

**Example 3** A punctured open disc, such as $\{z \in \mathbb{C} : 0 < |z| < 1\}$, is a Riemann surface. This set is not simply connected; instead, the first homotopy group is isomorphic to the infinite cyclic group. Thus, it cannot be homeomorphic to $\mathbb{C}$ or the unit disc. *A fortiori*, it is not biholomorphically equivalent to either of them.

**Example 4** The quotient of $\mathbb{C}$ by a discrete group of translations inherits a complex structure. Thus, for example, the group generated by the translation $z \mapsto z + 1$ gives rise to an equivalence relation

$$z_1 \sim z_2 \Leftrightarrow (z_1 - z_2) \in \mathbb{Z}, \quad (2.1)$$

and the quotient $\mathbb{C}/\sim$ is a Riemann surface (topologically a cylinder, $S^1 \times \mathbb{R}$).

**Example 5** Taking instead the group generated by two translations, $z \mapsto z + 1$ and $z \mapsto z + c$ for some non-real $c$ gives a compact quotient: indeed, any point in $\mathbb{C}$ is equivalent to one in the parallelogram $\{s + tc : 0 \leq s, t < 1\}$, and opposite sides are identified, giving the quotient the topology of a torus ($S^1 \times S^1$). To emphasize the complex structure, one calls it a complex torus. A torus is shown in figure 2.1.

**Example 6** Affine curves are sets of the form

$$\mathcal{C} = \{(z, w) \in \mathbb{C}^2 : p(z, w) = 0\} \quad (2.2)$$

for some polynomial $p$. Locally, such curves look like $\mathbb{C}$ in the following sense: if $(z_0, w_0)$ is a point on the curve and if the partial derivatives

$$\frac{\partial p}{\partial z}|_{(z_0, w_0)} \quad \text{and} \quad \frac{\partial p}{\partial w}|_{(z_0, w_0)} \quad (2.3)$$

are not both zero, then by the implicit function theorem one of the variables is a holomorphic function of the other in some neighbourhood of $(z_0, w_0)$. When this holds at every point of the curve, it is called smooth or nonsingular. Projection onto one of the variables gives a chart, and the curve is a Riemann surface.
Example 7 Examples number 4 and 5 constructed Riemann surfaces from $\mathbb{C}$ by forming the quotient with a discrete group. It is also possible to start in the two-dimensional complex manifold $\mathbb{C}^2$ and quotient by a continuous group, reducing the complex dimension to 1. The complex projective line is the set $\mathbb{C}^2 \setminus (0,0)$ of pairs of complex numbers, not both zero, modulo the equivalence relation

$$(a, b) \sim (c, d) \quad \text{if} \quad (a, b) = (zc, zd) \quad (2.4)$$

for some non-zero $z \in \mathbb{C}$. The equivalence class of $(a, b)$ is written $(a : b)$. They can be thought of as one-dimensional subspaces of $\mathbb{C}^2$, namely the line through $(a, b)$ and the origin $(0,0)$. If $b \neq 0$, one can represent the class $(a : b) = (a/b : 1)$ by the single complex number $a/b$. Thus the projective line contains a copy of $\mathbb{C}$. In addition, it contains the point $(1 : 0)$ (note that if $b = 0$, $a$ is nonzero and can be normalized to 1). Now consider a sequence of points $\{(a_n : b_n)\}$, normalized so that $|a_n|^2 + |b_n|^2 = 1$ for every $n$, and such that $b_n$ tends to 0 as $n$ tends to infinity. Then $|a_n/b_n|$ tends to infinity, suggesting that $(1 : 0)$ plays the role of a “point at infinity”. This intuition can be made precise by means of the Riemann sphere.

Finally, we remark that for a compact Riemann surface $S$, the field of meromorphic functions defined on $S$ is a finite extension of the field $\mathbb{C}(z)$ of rational functions in a complex variable, so that any two meromorphic functions are algebraically dependent.

2.1 The Riemann sphere

We now construct the Riemann sphere, which is the Riemann surface hinted at in the last example above. For more on this, see [23]. Identify the complex number $z = x + iy$ with the point $(x,y,0)$ in $\mathbb{R}^3$, and consider the unit 2-sphere:

$$S^2 = \{(x,y,w) \in \mathbb{R}^3 : x^2 + y^2 + w^2 = 1\}. \quad (2.5)$$

By stereographic projection from the “north pole" $(0,0,1)$, the sphere (minus the point of projection) can be identified with the complex plane $\mathbb{C}$. Explicitly,
one has a projection
$$\pi_1 : S^2 \setminus \{(0,0,1)\} \rightarrow \mathbb{C}$$
$$\pi_1(x,y,w) = \frac{x + iy}{1 - w} \quad (2.6)$$

with an inverse given by
$$\pi_1^{-1}(z) = \left(\frac{2 \Re z}{|z|^2 + 1}, \frac{2 \Im z}{|z|^2 + 1}, \frac{|z|^2 - 1}{|z|^2 + 1}\right). \quad (2.7)$$

It is clear that there is nothing special with the point (0,0,1): any choice of projection point will do, and using instead (0,0,−1), one obtains another projection $\pi_2$:
$$\pi_2 : S^2 \setminus \{(0,0,−1)\} \rightarrow \mathbb{C}$$
$$\pi_2(x,y,w) = \frac{x - iy}{1 + w} \quad (2.8)$$

with inverse given by
$$\pi_2^{-1}(z) = \left(\frac{2 \Re z}{|z|^2 + 1}, -\frac{2 \Im z}{|z|^2 + 1}, \frac{1 - |z|^2}{|z|^2 + 1}\right).$$

The projections are homeomorphisms, the two charts $\{(\pi_1, S^2 \setminus \{(0,0,±1)\})$ cover the entire sphere, and the transition function
$$\pi_2 \circ \pi_1^{-1} : \mathbb{C}^* \rightarrow \mathbb{C}^*$$
$$\pi_2 \circ \pi_1^{-1}(z) = \frac{1}{z} \quad (2.9)$$

is holomorphic. Thus the sphere is a Riemann surface, called the Riemann sphere. The point (0,0,1) is the “point at infinity” referred to in the previous section. It is denoted $\{\infty\}$. Under the identification $(a : b) \leftrightarrow a/b$, it corresponds to the point (1:0). When thought of in this way, the Riemann sphere is the aforementioned complex projective line, often denoted $\mathbb{CP}^1$. An equivalent description is that of the extended complex plane, obtained by adjoining a single point $\{\infty\}$ to $\mathbb{C}$, declaring every set of the form $\{z \in \mathbb{C} : |z| > M\}$ to be a neighbourhood of $\{\infty\}. This embeds $\mathbb{C}$ in a compact space, and one speaks accordingly of a one-point compactification. The space obtained is denoted $\hat{\mathbb{C}}$. We will favour this notation. The Riemann sphere is the only compact Riemann surface of genus zero, i.e. a two-sphere admits only one complex structure. The next section introduces and classifies the self-mappings of the Riemann sphere.

### 2.2 Möbius transformations

Given any Riemann surface $S$, the set of biholomorphic mappings from $S$ to itself, so-called automorphisms, forms a group under composition, called the automorphism group and denoted $\text{Aut}(S)$. We now determine $\text{Aut}(\hat{\mathbb{C}})$. First, note that the concepts of holomorphic and meromorphic functions make sense on $\hat{\mathbb{C}}$. A function $f : \hat{\mathbb{C}} \rightarrow \hat{\mathbb{C}}$ is said to have a pole (zero) of order $n$ at $\{\infty\}$
2.2. Möbius transformations

if $f^{\frac{1}{n}}$ has a pole (zero) of order $n$ at $z = 0$. Liouville’s theorem (see for example [23], appendix 1) implies that two meromorphic functions with poles and zeros of the same order at the same places are proportional, because their quotient is a holomorphic function bounded in all of $\mathbb{C}$ (hence a constant). It follows that the meromorphic functions from the Riemann sphere to itself are precisely the rational functions. Bijectivity implies that neither nominator nor denominator can have degree greater than one. Thus the automorphisms of the Riemann sphere are precisely the Möbius transformations, which are functions of the form

$$z \mapsto \frac{az + b}{cz + d}, \quad \text{where} \quad ad - bc \neq 0. \quad (2.10)$$

The condition on the coefficients assures that the mapping is non-constant.

In projective coordinates, the action looks as follows:

$$(z : w) \mapsto (az + bw : cz + dw), \quad (2.11)$$

so

$\{\infty\} = (1 : 0) \mapsto (a : c) = \frac{a}{c} \quad \text{if} \quad c \neq 0,$

$\{\infty\} \mapsto \{\infty\} \quad \text{otherwise.} \quad (2.12)$

This is clearly consistent with what one obtains from limiting procedures, as is

$$-\frac{d}{c} = (d : -c) \mapsto (ad - bc : 0) = \{\infty\}. \quad (2.13)$$

It also shows that Möbius transformations can be conveniently represented by $2 \times 2$-matrices, i.e. one has a full (surjective) representation

$$\rho : \text{GL}(2, \mathbb{C}) \to \text{Aut}(\hat{\mathbb{C}}) \quad (2.14)$$

sending a nonsingular matrix $\left(\begin{array}{cc} a & b \\ c & d \end{array}\right)$ to the Möbius tranformation (2.10). Here, $\text{GL}(2, \mathbb{C})$ denotes the general linear group of nonsingular $2 \times 2$-matrices with entries in $\mathbb{C}$, under multiplication. Clearly this is a much larger group than $\text{Aut}(\hat{\mathbb{C}})$. Indeed, any scalar multiple of a matrix gives the same Möbius transformation:

$$awz + bw = az + b \quad \text{when} \quad w \neq 0, \quad (2.15)$$

so by division by $\sqrt{ad - bc}$, one may assume that the matrix has determinant equal to 1. The determinant is a homomorphism from matrices to complex numbers:

$$\det : \text{GL}(2, \mathbb{C}) \to (\mathbb{C}^*, \cdot), \quad (2.16)$$

and the group thus obtained is the kernel of this homomorphism. It follows that it is a normal subgroup. It is called the special linear group and is denoted $\text{SL}(2, \mathbb{C})$. The restriction of $\rho$ to $\text{SL}(2, \mathbb{C})$ is still not faithful (injective), because

$$\frac{az + b}{cz + d} \equiv z$$

and

$$az + b \equiv cz^2 + dz \Rightarrow b = c = 0, a = d,$$
so \( \text{ker } \varrho = \{ \pm I \} \), where \( I \) is the \( 2 \times 2 \)-identity matrix. The quotient \( \text{SL}(2, \mathbb{C})/\{ \pm I \} \) is called the \textit{projective special linear group} and is denoted \( \text{PSL}(2, \mathbb{C}) \). It is isomorphic to the group \( \text{Aut}(\hat{\mathbb{C}}) \) of Möbius transformations. One can also projectivize \( \text{GL}(2, \mathbb{C}) \) directly, forming the \textit{projective general linear group} \( \text{PGL}(2, \mathbb{C}) = \text{GL}(2, \mathbb{C})/\{ zI : z \in (\mathbb{C}^*) \} \). It is readily seen to be isomorphic to \( \text{PSL}(2, \mathbb{C}) \). Finally, note that a Möbius transformation can fix at most two different points, so that two Möbius transformations which agree in three points are identically equal.

2.2.1 Transitivity properties

Möbius transformations act 3-transitively on \( \hat{\mathbb{C}} \), that is, given any two sets \( \{ z_1, z_2, z_3 \} \) and \( \{ w_1, w_2, w_3 \} \) of nonequal points in \( \hat{\mathbb{C}} \), there is a Möbius transformation \( T \) such that \( T(z_i) = w_i, i = 1, 2, 3 \). To see this, note that

\[
T_1(z) = \frac{z - z_1}{z - z_3} / \frac{z_2 - z_1}{z_2 - z_3}
\]  

(2.17)

takes \( \{ z_1, z_2, z_3 \} \) to \( \{ 0, 1, \infty \} \). Let \( T_2 \) be the transformation taking \( \{ w_1, w_2, w_3 \} \) to \( \{ 0, 1, \infty \} \). Then \( T_2^{-1} \circ T_1 \) is the required Möbius transformation. The expression

\[
\frac{z - z_1}{z - z_3} / \frac{z_2 - z_1}{z_2 - z_3}
\]

(2.18)

is called the cross-ratio of \( z, z_1, z_2, z_3 \) and is denoted \( [z, z_1, z_2, z_3] \). Cross-ratios are preserved by Möbius transformations, i.e., for any transformation \( T \) one has \( [T(z), T(z_1), T(z_2), T(z_3)] = [z, z_1, z_2, z_3] \). One can show that the four points \( z, z_1, z_2, z_3 \) lie on a circle if and only if their cross-ratio is a real number (where circle is taken to mean circle on the Riemann sphere; these project to ordinary circles or lines in \( \mathbb{C} \), depending on whether or not the circle passes through \( \{ \infty \} \)). It follows that Möbius transformations take circles to circles, and moreover, the action is transitive, since a circle is determined by three points.

Example To determine a Möbius transformation mapping the real line to the unit circle in \( \mathbb{C} \), one can take \( T_1 \) to be the identity transformation (as \( 0, 1, \infty \) already belong to \( \mathbb{R} \)), and \( T_2 \) to be the transformation taking \( \{ -1, -i, 1 \} \) to \( \{ 0, 1, \infty \} \), i.e.

\[
T_2(z) = \frac{z + 1}{z - 1} / \frac{-i + 1}{-i - 1} = -i \frac{z + 1}{z - 1},
\]

and the inverse is given by

\[
T_2^{-1}(z) = \frac{z - i}{z + i}.
\]

One can show that this map sends the upper half-plane to the interior of the unit circle, for example by calculating that \( i \mapsto 0 \), or by noting that the mapping is orientation-preserving.

2.2.2 Stabilizer subgroups

Next, we determine some important stabilizer subgroups. For a subset \( S \) in \( \hat{\mathbb{C}} \), this is defined as the subgroup of \( \text{Aut}(\hat{\mathbb{C}}) \) mapping \( S \) into \( S \) (but not necessarily
2.2. Möbius transformations

We denote it \( \text{Stab}(S) \). By the discussion in the previous section, any circle (i.e., circle or line in \( \mathbb{C} \)) can be mapped onto the extended real line \( \hat{\mathbb{R}} = \mathbb{R} \cup \{\infty\} \). Thus the stabilizer subgroup of any circle is conjugate to the stabilizer of \( \hat{\mathbb{R}} \), which we now determine. Suppose that \( T \in \text{Stab}(\hat{\mathbb{R}}) \) and take three different points \( z_1, z_2, z_3 \), none of which maps to \( \{\infty\} \). Let \( T(z_i) = w_i \). Then all \( z_i, w_i \) are real numbers and the construction in the previous section shows that \( T \) is equal to a Möbius transformation with real coefficients. Conversely, it is easy to see that any such transformation fixes \( \hat{\mathbb{R}} \). Thus \( \text{Stab}(\hat{\mathbb{R}}) \) is isomorphic to the group \( \text{PGL}(2, \mathbb{R}) \) modulo the group \( \{zI : z \in \mathbb{R}^*\} \). This is not quite the special linear group, since the determinant equals \( \pm 1 \). Rather, \( \text{Stab}(\hat{\mathbb{R}}) \cong \text{PSL}(2, \mathbb{R}) \times \{\pm I\} \). The transformations with determinant \( -1 \) are orientation-reversing, i.e., they interchange the interior and exterior of a circle. It follows that the upper half-plane \( \{z : \Im z > 0\} \), denoted \( \mathcal{H} \), has stabilizer \( \text{PSL}(2, \mathbb{R}) \), and that the stabilizer of any disc (interior of a circle) is conjugate to this group.

Finally, \( \text{Stab}(\mathbb{C}) \) is equal to the group of affine transformations \( z \mapsto az + b \).

2.2.3 Classification of Möbius transformations

We now classify all Möbius transformations according to their conjugacy classes in \( \text{PSL}(2, \mathbb{C}) \). First, note that all elements of a conjugacy class must fix the same number of points. A nonconstant transformation

\[
T : z \mapsto \frac{az + b}{cz + d}
\]

fixes \( \{\infty\} \) if and only if \( c = 0 \). Solving

\[
\frac{az + b}{cz + d} = z
\]

gives \((2cz + (a - d))^2 = (a - d)^2 + 4bc = (a + d)^2 - 4\), (using \( ad - bc = 1 \)) so \( T \) has one fixed point if and only if \( (a + d)^2 = 4 \), and two otherwise. The quantity \( a + d \) is the trace of the matrix representing the transformation; it is determined up to a sign, so the squared trace \( \text{tr}^2(T) \) is a well-defined function of the Möbius transformation. Moreover, it depends only on the conjugacy class:

\[
\text{tr}^2(UTU^{-1}) = \text{tr}^2(U^{-1}UT) = \text{tr}^2(T).
\] (2.19)

Now suppose \( T \) has only one fixed point. By conjugation, it can be moved to \( \{\infty\} \), and any Möbius transformation fixing only \( \{\infty\} \) has the form \( z \mapsto az + b \). By another conjugation, the representative can be taken to be the translation \( z \mapsto z + 1 \). If \( T \) has two fixed points, they can be taken to be \( 0 \) and \( \{\infty\} \), giving a representative of the form \( z \mapsto az, a \neq 0, 1 \). It is now straightforward to verify that \( \text{tr}^2(T_1) = \text{tr}^2(T_2) \) is a sufficient condition for \( T_1 \) and \( T_2 \) to be conjugate, and we have already seen that it is necessary. Thus the squared trace completely classifies conjugacy classes of Möbius transformations. A Möbius transformation \( T \) is called

**Elliptic** if \( \text{tr}^2(T) \) is real and belongs to \([0,4]\); such a transformation has two fixed points and is conjugate to a rotation, \( z \mapsto e^{i\theta}z \),

**Parabolic** if \( \text{tr}^2(T) = 4 \); such a transformation has one fixed point and is conjugate to the translation \( z \mapsto z + 1 \),
Hyperbolic if $\text{tr}^2(T)$ is real and $\geq 4$; such a transformation has two fixed points and is conjugate to a dilatation, $z \mapsto rz$, $r \in \mathbb{R}$.

Loxodromic otherwise. Such a transformation has two fixed points and is conjugate to a transformation $z \mapsto cz$, $c$ nonreal and of absolute value $\neq 1$.

Sometimes hyperbolic transformations are counted as loxodromic as well. The name loxodromic means “slanting path” and describes how the flow lines of a loxodromic transformation on the Riemann sphere (fixing 0 and $\{\infty\}$) cross the meridians at constant angle. For an elementary (and visually pleasing) account of this classification, see [32]. Figure 2.2 shows the flow lines of a loxodromic transformation with two finite fixed points.

![Figure 2.2: A loxodromic transformation [27]](image)

2.3 Hyperbolic geometry

As noted by Poincaré, the upper half-plane $\mathcal{H}$, equipped with the $\text{PSL}(2,\mathbb{R})$-invariant metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2},$$

is a model of a hyperbolic plane. The geodesics (straight lines) are lines perpendicular to the real axis, and half-circles meeting the real axis orthogonally; see figure 2.3.

$\text{PSL}(2,\mathbb{R})$ acts on this plane as a group of isometries. It follows from the discussion in section 2.2.1 that its action on the set of geodesics is transitive. Furthermore, there exists a unique geodesic between any two points. When doing geometry in this model, it is convenient to adjoin the point at infinity and the boundary $\Im z = 0$. Any lines intersecting at these points meet at an angle zero. Figure 2.4 depicts a hyperbolic triangle.
In this example, all angles of the triangle are zero; that the sum is less than $\pi$ characterizes a space of negative curvature. The appropriate notion of area in this model is the Haar measure for $SL(2, \mathbb{R})$:

$$d\mu = \frac{dxdy}{y^2}. \quad (2.21)$$

Remarkably, the area of a triangle depends only on the sum of its angles. More precisely, it is given by the Gauß-Bonnet formula: let $\Delta$ be a triangle with angles $\alpha, \beta, \gamma$. Then

$$\mu(\Delta) = \pi - \alpha - \beta - \gamma. \quad (2.22)$$

For a proof, see [23], chapter 5. This immediately shows, for example, that the triangle in figure 2.4 has area equal to $\pi$. Finally, we remark that another common model of two-dimensional hyperbolic geometry uses the unit disc in $\mathbb{C}$ instead of the upper half-plane. The M"obius transformation constructed as an example in section 2.2.1 can be used to translate between the models. This alternative model is called the Poincaré disc model.
2.4 Groups of Möbius transformations

2.4.1 Continuous groups

We now study some groups of Möbius transformations, and their generators. We have already treated $\text{PSL}(2, \mathbb{C})$. It is generated by transformations of the following forms:

- Translations: $z \mapsto z + c$, $c \in \mathbb{C}$
- Rotations: $z \mapsto e^{i\theta}z$, $\theta \in \mathbb{R}$
- Dilatations: $z \mapsto rz$, $r \in \mathbb{R}$
- The inversion $z \mapsto \frac{1}{z}$.

This is easily seen by writing

\[
\begin{align*}
az + b &= \frac{a}{cz + d} + \frac{b}{cz + d}, \quad c = 0 \\
\frac{az + b}{cz + d} &= -\frac{c}{c(cz + d)}, \quad c \neq 0.
\end{align*}
\]

Equation (2.23) is valid for real Möbius transformations as well, and in this case, one can dispose with the rotations and take the translations to be real. This gives generators for $\text{PSL}(2, \mathbb{R})$.

A rotation of the Riemann sphere is clearly an automorphism. Thus it should be possible to interpret $\text{SO}(3, \mathbb{R})$, the special orthogonal group, which acts on $\mathbb{R}^3$ as rotations, as a group of Möbius transformations. It turns out to be easier to work not with $\text{SO}(3, \mathbb{R})$ directly but with the group $\text{PSU}(2, \mathbb{C})$, the projective special unitary group. It is well known that the special unitary group $\text{SU}(2, \mathbb{C})$ of unitary $2 \times 2$-matrices with determinant equal to one is a double cover of $\text{SO}(3, \mathbb{R})$; the quotient $\text{SU}(2, \mathbb{C})/\{\pm I\}$ is isomorphic to $\text{SO}(3, \mathbb{R})$. Noting that a Möbius transformation is a rotation if and only if it commutes with the antipodal mapping $z \mapsto -\frac{1}{z}$, one finds

\[
\begin{align*}
\frac{az + b}{cz + d} &= \frac{a}{cz + d} \left( -\frac{c}{cz + d} \right), \quad c \neq 0.
\end{align*}
\]

whence

\[
\begin{pmatrix} b & -a \\ d & -c \end{pmatrix} = \lambda \begin{pmatrix} -\sigma & -\bar{d} \\ \sigma & \bar{b} \end{pmatrix}
\]

for some nonzero $\lambda \in \mathbb{C}$. One must have $\lambda = \pm 1$ (take determinants), and using $ad - bc = 1$ one finds $b = -\sigma$, $d = \bar{\sigma}$. Furthermore,

\[
\begin{pmatrix} b & -a \\ \sigma & -\bar{b} \end{pmatrix} \begin{pmatrix} \bar{b} & a \\ -\sigma & b \end{pmatrix} = \begin{pmatrix} |b|^2 + |a|^2 & 0 \\ 0 & |a|^2 + |b|^2 \end{pmatrix},
\]

and $|a|^2 + |b|^2 = ad - bc = 1$. This gives the expected identification.

\footnote{Cf. quantum mechanics, where $\text{SU}(2, \mathbb{C})$ acts on spinors rather than vectors.}
2.4. Groups of Möbius transformations

2.4.2 Finite groups of Möbius transformations

Any finite group of Möbius transformations consists only of rotations. We now classify these groups, i.e., finite rotation groups of spheres, following [9] and [23]. Let \( G \) be such a group, and let \( |G| = n \). Any nonidentity element of \( G \) has exactly two (antipodal) fixed points, which are the points of intersection with the rotation axis. Such a point is called \( p \)-gonal if it belongs to a rotation of order \( p \) (and no higher order). Thus a \( p \)-gonal point is fixed by \( p - 1 \) nonidentity transformations. Now, count the pairs \((z, g)\) of nonidentity elements of \( G \) and their fixed points; there are clearly \( 2(n - 1) \) such pairs, and counting by points in the sphere, one finds

\[
2(n - 1) = \sum_z (|\text{Stab}(z)| - 1),
\]

(2.27)

since the identity of \( G \) has to be excluded from every stabilizer subgroup. The only nonzero terms are those from fixed points of \( G \), so the sum is finite. On the other hand, every point in an orbit containing a \( p \)-gonal point is \( p \)-gonal, and by the orbit/stabilizer theorem, the length of such an orbit is \( n/p \). Writing \( p_1, p_2, \ldots \) for the occurring \( p \):s (not necessarily distinct), one finds

\[
2(n - 1) = \sum_z (|\text{Stab}(z)| - 1) = \sum_i \frac{n}{p_i} (p_i - 1),
\]

(2.28)

whence

\[
2(1 - \frac{1}{n}) = \sum_{i=1}^\ell (1 - \frac{1}{p_i}),
\]

(2.29)

where \( \ell \) is the number of orbits, and the terms in the right-hand side are \( \leq \frac{1}{7} \), so there can be at most three terms. In the case \( \ell = 1, n = 1 \) also, so the group is trivial. In the case \( \ell = 2, G \) is cyclic. Finally, it can be checked that \( \ell = 3 \) gives the solutions \( \{p_1, p_2, p_3\} = \{2, 2, p\} \) for an integer \( p \) (in which case \( n = 2p \) and \( G \) is dihedral), \( \{p_1, p_2, p_3\} = \{2, 3, 3\}, \{2, 3, 4\} \) or \( \{2, 3, 5\} \). The corresponding orders of \( G \) are 12, 24 and 60, respectively, and the groups are \( A_4, S_4 \) and \( A_5 \), which are the symmetry groups of a tetrahedron, octahedron and icosahedron ([23], [27]). This completely classifies the finite groups of Möbius transformations.

2.4.3 Infinite discrete groups of Möbius transformations

Finally, we introduce infinite discrete groups. For the background, see [27]. [10] gives a general treatment of group presentations and generators. A discrete subgroup\(^2\) of \( \text{PSL}(2, \mathbb{C}) \) is called Kleinian; for example, the group \( \text{PSL}(2, \mathbb{Z} + i\mathbb{Z}) \) is Kleinian. A Kleinian group having an invariant disc (which by conjugation can be taken to be \( \mathcal{H} \)) is called Fuchsian. Thus a Fuchsian group is, up to conjugacy, a discrete subgroup of \( \text{PSL}(2, \mathbb{R}) \). The most important example is \( \text{PSL}(2, \mathbb{Z}) \), the Möbius transformations with integer coefficients and determinant equal to 1. This group is called the modular group and is traditionally denoted \( \Gamma \) or \( \Gamma(1) \); the second notation will be explained shortly, but first, we determine

\(^2\)In the metric topology on \( \text{PSL}(2, \mathbb{C}) \).
generators and a presentation for $\Gamma$. It is convenient to start with its double cover $\text{SL}(2, \mathbb{Z})$, and introduce the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

$$U = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}. \quad (2.30)$$

It is straightforward to verify that

$$T = S^{-1}V = S^{3}V, \quad U = S^{-1}V^{2} = S^{3}V^{2},$$

and hence

$$V = T^{-1}U, \quad S = T^{-1}UT^{-1},$$

so $T$ and $U$ generate $\text{SL}(2, \mathbb{Z})$ if and only if $S$ and $V$ do. To see that this is the case, let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{SL}(2, \mathbb{Z})$. Clearly

$$T^{\lambda}M = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + \lambda c & b + \lambda d \\ c & d \end{pmatrix}$$

and

$$U^{\lambda}M = \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ c + \lambda a & d + \lambda b \end{pmatrix},$$

so by application of the matrices $T$ and $U$, $M$ can be brought on one of the forms

$$\begin{pmatrix} a' & b' \\ 0 & d' \end{pmatrix} \text{ or } \begin{pmatrix} 0 & b' \\ c' & d' \end{pmatrix}. \quad (2.36)$$

In the first case, $a'd' = 1$ so the matrix equals $\pm T^{\lambda}$, and can be written $T^{\lambda}$ or $S^{2}T^{\lambda}$. In the second case, one has $-b'c' = 1$ so the matrix equals $ST^{\lambda}$ or $S^{3}T^{\lambda}$ for some $\lambda$. Thus $\text{SL}(2, \mathbb{Z})$ is generated by $S$ and $V$. They satisfy the relations

$$S^{2} = V^{3}, \quad S^{4} = I.$$  

(2.37)

Next, we show that there are no other relations. Suppose to the contrary that some word in $S$ and $V$ is equal to the identity. Clearly, $S^{2}$ commutes with everything, so we may assume that the relation has the form

$$S \cdot S^{3}V^{\alpha}S^{3}V^{\beta} \cdots S^{\omega} = I$$

or

$$S^{3}V^{\alpha}S^{3}V^{\beta} \cdots S^{\omega} = I,$$

where the $V$-exponents are equal to 1 or 2, and using $T = S^{3}V$ and $U = S^{3}V^{2}$, this becomes

$$S \cdot W(T, U) \cdot S^{\omega} = I$$

or

$$W(T, U) \cdot S^{\omega} = I.$$  

(2.40)
for some word \( W \) in \( T \) and \( U \). To see that this is impossible, we let the matrices act on some matrix \( A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) satisfying \( c > 0 \) and \( d > 0 \), according to
\[
M.A = M^TAM. \tag{2.42}
\]
Note that this really is an action, i.e., that \( N.(M.A) = (NM).A \). It is straightforward to verify that \( \text{tr}(S.A) = \text{tr}(A) \), \( \text{tr}(T.A) = c + \text{tr}(A) \) and \( \text{tr}(U.A) = d + \text{tr}(A) \), so the only word in \( T \) and \( U \) which is equal to the identity is the empty word. The relation then simplifies to \( S^2 = I \) for some \( \omega \), and clearly the only such relation is already implied by \( (2.37) \). Thus \( (2.37) \) defines \( \text{SL}(2, \mathbb{Z}) \), and since the image of \( S \) in \( \Gamma \) has order two rather than four, the defining relations for the modular group are:
\[
S^2 = I, \quad V^3 = I. \tag{2.43}
\]
Finally, we introduce so-called congruence subgroups (of \( \Gamma \)). For an integer \( N \), reduction modulo \( N \) is a ring homomorphism from \( \mathbb{Z} \) to \( \mathbb{Z}/N\mathbb{Z} \), which extends to a homomorphism from \( \Gamma \) to \( \text{PSL}(2, \mathbb{Z}/N\mathbb{Z}) \). The kernel is a normal subgroup in \( \Gamma \), called the principal congruence subgroup of level \( N \). It has finite index in \( \Gamma \). This explains the notation \( \Gamma(1) \) for \( \Gamma \). Thus \( \Gamma(N) \) consists of transformations
\[
z \mapsto \frac{az + b}{cz + d}, \quad a, b, c, d \text{ integers}, \tag{2.44}
\]
where \( a \equiv d \equiv \pm 1 \) (mod \( N \)) and \( b \equiv c \equiv 0 \) (mod \( N \)). A larger congruence subgroup is the set of Möbius transformations of the form \( (2.44) \) such that \( N | c \). This group is denoted \( \Gamma_0(N) \). One sometimes defines an analogous group \( \Gamma_0^3(N) \) of transformations satisfying \( N | b \), and \( \Gamma_0^3(N) \) where \( N \) divides both \( b \) and \( c \), see \[4\]. Furthermore, one denotes by \( \Gamma_1(N) \) the group of Möbius transformations with \( a \equiv d \equiv \pm 1 \) (mod \( N \)) and \( c \equiv 0 \) (mod \( N \)). Summarizing, one has
\[
\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma. \tag{2.45}
\]
In general, a congruence subgroup of \( \Gamma \) is a subgroup containing \( \Gamma(N) \) for some \( N \), and the smallest such \( N \) is called the level of the group. \(^3\) See \[24\], chapter 9.

Example The principal congruence subgroup of level 2, \( \Gamma(2) \), has index 6 in \( \Gamma \); the quotient \( \Gamma/\Gamma(2) \) is isomorphic to \( \text{PSL}(2, \mathbb{Z}/2\mathbb{Z}) \), which can be shown to be isomorphic to \( S_3 \), the symmetric group on three letters. This group can be identified as the group permuting the tree non-identity elements of the 2-torsion subgroup of an elliptic curve; see \[28\], chapter 4 for more on this. The group \( \Gamma(2) \) is generated by the transformations \( S : z \mapsto -1/z \) and \( T^2 : z \mapsto z+2 \).

2.5 Group actions, covering spaces and uniformization

2.5.1 Group actions

Group actions are ubiquitous in algebra and geometry and are described in any textbook on these subjects. In general, an action of a group \( G \) on a space \( X \) is

\(^3\)It is a nontrivial fact that not every subgroup of finite index in \( \Gamma \) is a congruence subgroup.
a pair \((G, X)\) together with a homomorphism \(G \rightarrow \text{Aut}(X)\). For example, \(X\) could be a set, and \(\text{Aut}(X)\) permutations of its elements. As another example, \(X\) could be a topological space, in which case the elements of \(\text{Aut}(X)\) are homeomorphisms from \(X\) to itself. This will be assumed in the following. One often identifies the elements of \(G\) with their images in \(\text{Aut}(X)\), and writes \(g.x\) for the action of \(g \in G\) on \(x \in X\). The set \(\{g.x : g \in G\}\) is called the orbit of \(x\) (under \(G\)). \(G\) is said to act freely on \(X\) if every point of \(X\) has a neighbourhood \(U\) such that no two of its translates \(g.U\) overlap. If at most finitely many translates overlap in each orbit, the action is said to be properly discontinous. This is clearly a (strictly) weaker assumption.

The orbits of \(G\) partition \(X\), because “belonging to the same orbit” is an equivalence relation, as is readily verified. The set of orbits is denoted \(X/G\), and it inherits a topology from \(X\), as follows: let \(\pi\) denote the canonical projection \(X \rightarrow X/G\), which takes a point in \(X\) to its orbit under \(G\). Define a set \(S \subseteq X/G\) to be open if and only if \(\pi^{-1}(S)\) is open in \(X\). The topology thus induced is called the quotient topology. It is the coarsest topology in which \(\pi\) is continuous.

### 2.5.2 Covering spaces

Again, let \(X\) be a topological space. A covering space of \(X\) is a topological space \(Y\) together with a map \(\pi : Y \rightarrow X\) such that

1. Every point \(x \in X\) has a neighbourhood \(V\) such that \(\pi^{-1}(V) = \bigcup \alpha U_\alpha\) is the union of pairwise disjonts sets \(U_\alpha\) in \(Y\); and
2. the restriction of \(\pi\) to any such neighbourhood is a homeomorphism.

The preimage \(\pi^{-1}(x)\) of a point is called a fiber. Clearly, if \(X\) is connected, every fiber has the same cardinality. The significance of free group actions is the following: the canonical projection \(X \rightarrow X/G\), which takes a point in \(X\) to its orbit under \(G\). Define a set \(S \subseteq X/G\) to be open if and only if \(\pi^{-1}(S)\) is open in \(X\). The topology thus induced is called the quotient topology. It is the coarsest topology in which \(\pi\) is contious.

One can show (see, for example, [36], [39]) that a fundamental domain for the action of \(\Gamma(1)\) on \(\mathcal{H}\) is given by the set

\[
\left\{ z : -1/2 \leq z \leq 0, |z| \geq 1 \right\} \cup \left\{ z : 0 < z < 1/2, |z| > 1 \right\},
\]

see figure 2.7. The translation \(\tau \mapsto \tau + 1\) identifies the left and right boundary lines, and the inversion \(\tau \mapsto -1/\tau\) identifies the arcs on the unit circle from

---

4Actually, what we have defined is called a left action and the quotient ought to be written \(G \backslash X\); however, we will not deal with double cosets, so no confusion should occur.

5i.e., a global section of the fibration \(\pi : X \rightarrow X/G\)
Figure 2.5: An octahedral tesselation \[27]\n
\[
(-1 + i\sqrt{3})/2 \text{ to } i \text{ and from } (1 + i\sqrt{3})/2 \text{ to } i, \text{ respectively.}
\]

Figure 2.8 shows the tesselation induced by this choice of fundamental domain, that is, all translates of it by elements of $\Gamma$. The action of $\Gamma$ is not free: the mapping $\tau \mapsto -1/\tau$ fixes $i$ and $\tau \mapsto -1/(\tau + 1)$ fixes $(-1 \pm i\sqrt{3})/2$, and these points are identified by the translation $\tau \mapsto \tau + 1$. The corresponding stabilizer groups are isomorphic to $C_2$ and $C_3$, respectively (the cyclic groups of order 2 and 3). One can show that $\Gamma$ possesses no other fixed points in the fundamental domain (2.46), see for example [36], chapter 7. These points are actually an obstacle when constructing the quotient. Intuitively, the fundamental domain only sees one half point at $i$ and one sixth of a point at each of $(-1 \pm i\sqrt{3})/2$, for a total of one third of a point. We will gloss over this subtlety in the following.

Continuing the final example in section 2.4.3, figure 2.9 shows a fundamental domain for $\Gamma(2)$, which is given by the union of the fundamental domain (2.46) for $\Gamma(1)$ and its image under the translation $T : z \mapsto z + 1$.

There is a useful analogy between covering spaces and Galois theory. To describe it, we need the concept of a deck transformation. Suppose that $Y$ covers $X$ by the projection $\pi$, and consider (invertible) functions $f : Y \to Y$ such that $\pi \circ f = \pi$; that is, functions acting on $Y$ in such a way that the image in $X$ is unchanged. Such a function $f$ is called a deck transformation. Deck transformations evidently form a group under composition. This should be compared to the Galois group $\text{Gal}(k'/k)$ of a field extension $k' \geq k$, consisting of morphisms $k' \to k'$ which fix $k$. The algebraic closure of $k$, denoted by $\overline{k}$, in which every every polynomial in $k[x]$ splits completely, corresponds to the so-called universal covering space of $X$. We denote it here by $\tilde{X}$. Is is simply connected, so the homotopy group $\pi_1(\tilde{X}, x_0)$ is trivial (here $x_0 \in \tilde{X}$ is some base point; this choice is immaterial, as different choices give the same group up to conjugacy for connected spaces; for more general spaces, on the other
hand, one cannot define homotopy groups at all, but only groupoids). It is also
unique up to isomorphism, due essentially to its definition as initial object in the
category of spaces covering \( X \). For details, see any book on algebraic topology,
such as [30]. There is a bijective correspondence between covering spaces of \( X \)
and conjugacy classes of subgroups of \( \pi_1(X, x_0) \), and the cardinality of a fiber
of such a covering equals the index of the corresponding homotopy group in
\( \pi_1(X, x_0) \).

2.5.3 Quotient surfaces

Continuing the discussion in section 2.5, the quotient of a Riemann surface by
a freely acting, discrete group is again a Riemann surface. Is is compact if and
only if the group contains no parabolic elements. Even if the group does not act
freely, the quotient is almost a Riemann surface. The more general notion of
orbifold can handle the milder types of singularities which occur, for example,
with the modular group. See [40] for more on orbifolds (and much else).

The quotient of \( \mathcal{H} \) by \( \Gamma(N) \) is denoted \( Y(N) \); it is non-compact, because
\( \Gamma(N) \) contains the parabolic element

\[
\tau \mapsto \tau + N,
\]

which fixes only \( \{\infty\} \). The compactification can be obtained by letting \( \Gamma(N) \)
act on \( \mathcal{H} \cup \bar{Q} = \mathcal{H} \cup \bar{Q} \cup \{\infty\} \) instead; it is denoted \( X(N) \) and is called a modular
curve. The points \( X(N) \setminus Y(N) \) are called the cusps of \( X(N) \) (and by extension
of \( \Gamma(N) \)); there are only finitely many. For example, \( \Gamma(1) \) has only one cusp,
\( \{\infty\} \), which is equivalent to all real rational numbers at the boundary of \( \mathcal{H} \): to
see this, let \( a/c \in \mathbb{Q} \) be arbitrary. We may suppose that \( (a, c) = 1 \), in which
case there are integers \( b \) and \( d \) such that \( ad - bc = 1 \), and hence a M"obius
transformation

\[
\tau \mapsto \frac{a\tau + b}{c\tau + d}
\]
2.5. Group actions, covering spaces and uniformization

Figure 2.7: A fundamental domain for $\Gamma$

in $\Gamma$ taking $\{\infty\}$ to $a/c$. Thus $\Gamma$ identifies $\{\infty\}$ and all rational numbers at the boundary of $\mathcal{H}$. Examining the fundamental domain in figure 2.7 and its identifications shows that $Y(1)$ is a sphere with one point missing, so that $X(1)$ is a sphere. It is also a Riemann surface, so by the discussion in 2.1, there should be a biholomorphic function from $X(1)$ to $\hat{\mathbb{C}}$. Such a function does indeed exist; it will be introduced in section 5.2.2. One can similarly construct surfaces $X_0(N)$ and $X_1(N)$, which are the compactifications of the quotients $\mathcal{H}/\Gamma_0(N)$ and $\mathcal{H}/\Gamma_1(N)$. In section 3.3.7, we will interpret these surfaces as moduli spaces of elliptic curves. Finally, note that the groups $\Gamma(N)$ are normal in $\Gamma$, so the action of $\Gamma$ descends to an action on $\mathcal{H}/\Gamma(N)$. This symmetry group is isomorphic to $\Gamma/\Gamma(N) \cong \text{PSL}(2, \mathbb{Z}/N\mathbb{Z})$. We give an example (of considerable historical interest):

**Klein’s quartic** The quotient $\mathcal{H}/\Gamma(7)$ is known as the *Klein quartic*. In projective coordinates, it has equation $X^3Y + Y^3Z + Z^3X = 0$. The group $\text{PSL}(2, \mathbb{Z}/7\mathbb{Z})$, also written $\text{PSL}(2, 7)$, acts as symmetries of the surface. This group has 168 elements. A theorem by Hurwitz shows that a Riemann surface of genus $g \geq 2$ has at most $84(g - 1)$ automorphisms. The Klein quartic has genus 3, and $168 = 84 \cdot 2$, so the maximum is attained. For more on Klein’s quartic, see [26].

2.5.4 The Uniformization theorem

For the background for this section, see [14], [23]. We have seen that the Riemann sphere $\hat{\mathbb{C}}$, the complex plane $\mathbb{C}$ and the upper half-plane $\mathcal{H}$ are Riemann surfaces, as are quotients of the last two by discrete groups. The uniformization theorem by Poincaré shows that, in fact, all Riemann surfaces arise in this way. It states that the universal covering space of any Riemann surface is one of $\hat{\mathbb{C}}$, $\mathbb{C}$ or $\mathcal{H}$.
A different formulation is that any Riemann surface admits a metric, giving the surface constant curvature, which can be taken to be $-1, 0$ or $1$. Surfaces are accordingly classified as elliptic, parabolic or hyperbolic. The only compact elliptic Riemann surface is $\mathbb{C}$ itself, and the only compact parabolic Riemann surfaces are the complex tori, $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$. All other compact Riemann surfaces are uniformized by $\mathcal{H}$, and so are intrinsically hyperbolic.
Figure 2.10: Klein’s quartic [27]


Chapter 3

Lattices and Elliptic Curves

3.1 Lattices

In this chapter, we introduce the compact Riemann surfaces uniformized by \( \mathbb{C} \). The material is standard and is described in for example [1], [22], [23], [24]. Consider the additive group \((\mathbb{C}, +)\), i.e., the complex numbers under addition. By a lattice in \( \mathbb{C} \), we shall mean a free subgroup of rank two. Any such group \( \Lambda \) has a two-element basis \((\lambda_1, \lambda_2)\), such that the elements in the lattice are precisely the complex numbers of the form \( \lambda = m\lambda_1 + n\lambda_2 \), where \( m \) and \( n \) are integers. Thus, a lattice is also a free \( \mathbb{Z} \)-module on two generators. To avoid degenerate cases, we will assume that \( \lambda_2/\lambda_1 \) is non-real. When we wish to emphasize the basis, we write \( \Lambda = \Lambda(\lambda_1, \lambda_2) \). The lattice also acts on \( \mathbb{C} \) as a group of translations in the obvious way: \( \lambda \) corresponds to the translation \( z \mapsto z + \lambda \). Figure 3.1 depicts a lattice and a sublattice of index 2 (in bigger dots). The arrows show a possible basis.

![Figure 3.1: A lattice and a sublattice of index 2](image)

The symmetry group of the lattice (seen as a point set in a Euclidean plane)

always contains a translation group generated by the translations \( \lambda \mapsto \lambda + \lambda_1 \) and \( \lambda \mapsto \lambda + \lambda_2 \); it is isomorphic to \( \mathbb{Z}^2 \). It also contains the central inversion \( \lambda \mapsto -\lambda \). Sometimes the symmetry group is larger, containing rotations as well.

Consider a rotation around a lattice point, which may be taken to be the origin. It corresponds to multiplication by a complex number of magnitude 1: \( \lambda \mapsto c\lambda \), and since the group is finite (the point set being discrete), \( c \) is a root of unity. It is easy to show that the number of points with a minimal distance to the origin must be 4 or 6; hence \( c \) equals \( i \) or \( \rho = e^{2\pi i/3} = \frac{-1 + \sqrt{3}i}{2} \). Figure 3.2 and 3.3 show these lattices, together with a basis for each; they consist of, respectively, the Gaussian and Eisensteinian integers: these are \( \mathbb{Z} + i\mathbb{Z} \) and \( \mathbb{Z} + \rho\mathbb{Z} \). They are Euclidean domains, with unique factorization (up to units). They are also called the square and hexagonal lattices. The hexagonal lattice gives the densest packing of spheres (i.e. circles) in two dimensions. Note that the parameters for these lattices are precisely the fixed points of \( \Gamma \), and that the order of the stabilizer groups at these points is reflected in the rotation groups of the lattices. Lattice geometry is described in [9].

A central concept is that of a fundamental domain for the lattice (also called fundamental region or cell). By this, one means a fundamental domain for the action of the lattice on \( \mathbb{C} \) as a translation group, see 2.5. We repeat the definition for convenience: a fundamental domain is defined as a domain (simply connected set) \( \Omega \subseteq \mathbb{C} \), containing precisely one element from each orbit of \( \mathbb{C} \) under the action of \( \Lambda \). (Care must be taken at the boundary; the domain can be neither open nor closed. It is not uncommon to ignore this, working instead with the closure or interior of a domain). The translates of a fundamental domain form a tesselation of \( \mathbb{C} \). It follows in particular that each copy of the domain contains precisely one lattice point. This requirement does not determine a fundamental domain uniquely, but it fixes its area. Figure 3.4 shows some possible fundamental domains for a lattice. A fundamental domain which is also a polygon is called a fundamental polygon.

For any basis \( \langle \lambda_1, \lambda_2 \rangle \), the set \( \{ s\lambda_1 + t\lambda_2 : 0 \leq s, t < 1 \} \) is a fundamental domain, shaped like a parallelogram. The domains in figure 3.4 are of this kind. Another alternative, known as the Dirichlet fundamental domain, is to let
\[ \Omega \) (the interior of \( \Omega \)) be the set of points closer to a fixed lattice point than to any other. In other contexts, this is known as a Voronoi tessellation. It is shown in figure 3.5. In both cases, the fundamental domain is a convex polygon, which furthermore is centrally symmetric because the lattice is invariant under the central inversion \( \lambda \mapsto -\lambda \). It is easy to show that a Dirichlet fundamental polygon always has four or six sides. Less obvious, but nonetheless true, is that the same restriction holds for any convex fundamental polygon, see [9], chapter 4.

Clearly, the basis is not uniquely determined by the lattice. Suppose that \( \Lambda(\lambda_1, \lambda_2) = \Lambda(\lambda'_1, \lambda'_2) \). Then, since \( \{\lambda_1, \lambda_2\} \) is a basis, there are integers \( a, b, c, d \) such that

\[
\begin{pmatrix} \lambda'_1 \\ \lambda'_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix},
\]

(3.1)

and since \( \{\lambda'_1, \lambda'_2\} \) is also a basis, there are integers \( a', b', c', d' \) such that

\[
\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \begin{pmatrix} \lambda'_1 \\ \lambda'_2 \end{pmatrix}.
\]

(3.2)

Putting these together and taking determinants, one finds

\[
\det \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \cdot \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = 1,
\]

(3.3)

so the determinants, being integers, must be equal to \( \pm 1 \). On the other hand, any such matrix has an inverse with integer entries and hence is a valid change of basis. Thus the automorphism group of the lattice, seen as a \( \mathbb{Z} \)-module, is equal to \( SL(2, \mathbb{Z}) \times \{1, -1\} \). The matrices with determinant equal to 1 are orientation-preserving, while those with determinant equal to \(-1\) change the handedness of the basis. In either case, the area of the fundamental region is unchanged. The factor \( \{1, -1\} \) arises from regarding the basis as an ordered pair, rather than a set.

With the assumption above, the quotient \( \mathbb{C}/\Lambda \) is a torus, and all such tori are topologically equivalent. They are, however, not in general equivalent as

\begin{figure}[ht]
\centering
\includegraphics[width=\textwidth]{hexagonal_lattice}
\caption{The hexagonal lattice}
\end{figure}
Riemann surfaces (with the complex structure inherited from \( \mathbb{C} \)). For Riemann surfaces, the natural concept of equality is that of conformal equivalence. The corresponding equivalence relation for lattices is called homothety (or similarity): two lattices related by multiplication by a non-zero complex number (geometrically, dilatations and rotations around the origin) are called homothetic. In the following, “lattice” will be taken to mean “equivalence class of homothetic lattices”. One may therefore take one basis element to equal 1, and by suitably numbering the basis elements, one may suppose that \( \lambda_2 / \lambda_1 \) lies in the upper half-plane \( \mathcal{H} \). The quotient \( \lambda_2 / \lambda_1 \) is called the modulus and will be denoted \( \tau \). The corresponding lattice is written \( \Lambda_\tau \).

By the discussion above, different moduli \( \tau = \lambda_2 / \lambda_1 \) and \( \tau' = \lambda'_2 / \lambda'_1 \) give rise to equivalent tori if they correspond to different bases for the same lattice, i.e. if there are integers \( a, b, c, d \) such that \( ad - bc = \pm 1 \) and

\[
\begin{pmatrix} \lambda'_2 \\ \lambda'_1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda_2 \\ \lambda_1 \end{pmatrix},
\]

whence

\[
\tau' = \frac{\lambda'_2}{\lambda'_1} = \frac{a \lambda_2 + b \lambda_1}{c \lambda_2 + d \lambda_1} = \frac{a \tau + b}{c \tau + d}.
\]

Thus moduli of different tori correspond to orbits of elements in \( \mathcal{H} \) under the action of \( \Gamma \); this is what motivates the name modular group. The quotient \( X(1) \) (see section 2.5.3) is called the *moduli space* of elliptic curves\(^1\). We will return to this in section 3.3.7, having introduced elliptic curves.

\(^1\)Technically, a coarse moduli space
3.1.1 Eisenstein series

Given a lattice, one defines the Eisenstein series of weight $2k$, $k \geq 2$ an integer, as the sum

$$G_{2k}(\Lambda) = \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{\lambda^{2k}}, \quad (3.6)$$

or, as a function of a complex variable $\tau \in \mathcal{H}$, $G_{2k}(\tau) = G_{2k}(\Lambda \tau)$. Note that for odd exponents, the sum is zero by the central inversion symmetry of the lattice; $k \geq 2$ assures uniform convergence. See [1], [36],[39].

3.2 Elliptic functions

We now introduce elliptic functions. See [23], chapter 3 for a more complete account. Lattices, as defined above, are period modules of doubly periodic functions, generalizing the ordinary periodic functions. We will consider only meromorphic functions periodic with respect to some lattice $\Lambda(\omega_1, \omega_2)$. Such functions are called elliptic with respect to $\Lambda$. The reason for this is historical: elliptic functions first arose as inverses of elliptic integrals, which in turn arose in attempts to calculate the arc-length of an ellipse. Another similar problem asks the same question about lemniscates, which are curves with equations of the form $(x^2+y^2)^2 = 2c^2(x^2-y^2)$. The problem was studied by Jakob Bernoulli, but in fact the lemniscate is a special case of the Cassini ovals. Other contributors are Fagnano, Euler and Gauß. The special case of the lemniscate leads to integrals of the form

$$\int \frac{dx}{\sqrt{1-x^4}}, \quad (3.7)$$

and more generally, one considers integrals

$$\int R(x, y)dx, \quad (3.8)$$
where $R$ is a rational function and $y^2 = p(x)$ for some polynomial $p$ of degree three or four without repeated roots. By suitable changes of variables, such integrals can be reduced to a small canonical set. Several conventions exist, such as those of Jacobi, Legendre, Weierstraß and more recently Carlson. That of Weierstraß is the cleanest and will be described here. Apart from those aspects of elliptic functions relevant here, the theory has numerous applications in geometry, mechanics and astronomy. Indeed, Jacobi commented on the remarkable unity of mathematics evidenced by the occurrence of elliptic functions both in the study of both number theory and the pendulum.

The poles of a meromorphic function are isolated, hence countable. Thus, we may assume that the boundary of the fundamental polygon contains no pole. Applying Cauchy’s residue theorem to the path encircling the boundary, one obtains zero, the contributions from opposing sides cancelling by periodicity. It follows that the residues sum to zero. The sum of the orders of the poles in the fundamental polygon is called the order of the elliptic function. An elliptic function of order zero is bounded everywhere and must be constant by Liouville’s theorem. Elliptic functions of order one would have simple poles in the fundamental polygons and hence the sum of the residues would be nonzero. Thus elliptic functions of order one do not exist. Weierstraß constructed an elliptic function of order two, as follows:

$$
\wp(z) = \wp(z, \Lambda) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda \setminus \{0\}} \left( \frac{1}{(z - \lambda)^2} - \frac{1}{\lambda^2} \right). 
$$

(3.9)

$\wp$ is an even elliptic function of order two, and its derivative $\wp'$ is odd and has order three. The easiest way to verify the $\Lambda$-periodicity is to note that the derivative

$$
\wp'(z) = -2 \sum_{\lambda \in \Lambda} \frac{1}{(z - \lambda)^3}
$$

(3.10)

is periodic (termwise differentiation being allowed by uniform convergence). Next, take a $\lambda \in \Lambda$ and integrate

$$
\wp'(z + \lambda) - \wp'(z) = 0 
$$

(3.11)

to obtain

$$
\wp(z + \lambda) - \wp(z) = C(\lambda), 
$$

(3.12)

where $C(\lambda)$ is independent of $z$ but may depend on $\lambda$. Finally, take $z = -\lambda/2$:

$$
\wp(\lambda/2) - \wp(-\lambda/2) = C(\lambda) = 0 
$$

(3.13)

since $\wp$ is even. Clearly, the elliptic functions with respect to some fixed lattice form a field, and since all constant functions are trivially elliptic, this field can be seen as an extension of $\mathbb{C}$. It is in fact a finite extension, because two elliptic functions with poles and zeroes of the same order and at the same places are necessarily proportional (their quotient being an elliptic function without poles). This can be used to express any given even elliptic function with prescribed zeroes and poles as a rational function of $\wp$, so the subfield of even elliptic functions equals $\mathbb{C}(\wp)$. Splitting an arbitrary elliptic function in an even and an odd part and using that $\wp'$ is odd, one obtains an expression which is a rational function of $\wp$ and $\wp'$. Thus the field of elliptic functions equals $\mathbb{C}(\wp, \wp')$. 
Next, we expand \( \wp \) in a Laurent series:

\[
\wp(z) = \frac{1}{z^2} + 3G_4z^2 + 5G_6z^4 + 7G_8z^6 + \cdots, \tag{3.14}
\]
giving

\[
\begin{align*}
\wp'(z) &= -\frac{2}{z^3} + 6G_4z + 20G_6z^3 + 42G_8z^5 + \cdots, \\
\wp'(z)^2 &= \frac{4}{z^6} - \frac{24G_4}{z^2} - 80G_6 + (36G_4^2 - 168G_8)z^2 + \cdots, \\
\wp(z)^3 &= \frac{1}{z^3} + \frac{9G_4}{z^2} + 15G_6 + (21G_8 + 27G_4^2)z^2 + \cdots, \\
\wp'(z)^2 - 4\wp(z)^3 &= -\frac{60G_4}{z^2} - 140G_6 - (72G_4^2 + 252G_8)z^2 + \cdots, \\
\wp'(z)^2 - 4\wp(z)^3 + 60G_4\wp(z) + 140G_6 &= (108G_4^2 - 252G_8)z^2 + \cdots.
\end{align*}
\]

Note that the last expression is an elliptic function without poles, hence a constant, and letting \( z = 0 \) shows the constant to be zero. Thus Weierstrass' elliptic function satisfies the differential equation:

\[
(\wp')^2 = 4\wp^3 - g_2\wp - g_3, \tag{3.15}
\]

where

\[
g_2 = 60 \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{\lambda^4} = 60G_4 \tag{3.16}
\]

and

\[
g_3 = 140 \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{\lambda^6} = 140G_6 \tag{3.17}
\]

are called the elliptic invariants. We note in passing that the vanishing of the coefficients in the right-hand side gives relations between the Eisenstein series:

\[
G_8 = \frac{3}{7}G_4^2, \\
G_{10} = \frac{5}{11}G_4G_6, \ldots
\]

These identities will later be interpreted as equalities between modular forms.

### 3.3 Elliptic curves

We now introduce elliptic curves. References for this section are [13], [22], [24], [33], [37], [38], and [39]. An elliptic curve is a curve of genus 1, which is simultaneously a group; over a field with characteristic \( \neq 2, 3 \), it is given by an equation of the form \( y^2 = p(x) \) where \( p \) is a polynomial of degree three or four without repeated roots. As a Riemann surface, this is a torus, which is uniformized by \( \mathbb{C} \), and the group structure of \( (\mathbb{C},+) \) descends to a group structure on the torus. However, over other fields, the group structure can be described both algebraically and geometrically. First, we give some historical background.
3.3.1 History

We give two examples of problems giving rise to elliptic curves, one ancient and one more recent.

Example 1 A congruent number is a number which is the area of some right triangle with rational side-lengths. The problem of determining whether or not a given number is congruent has been studied at least since antiquity by Arabian and Greek mathematicians; Fibonacci treated the problem as well. For a more detailed account, see [44]. Denote by \( h \) the hypotenuse of the triangle; then the integer \( D \) is congruent if and only if \((h/2)^2 \pm D)\) are both squares (of rational numbers), i.e. if and only if the system

\[
\begin{align*}
    u^2 + Dv^2 &= a^2 \\
    u^2 - Dv^2 &= z^2
\end{align*}
\]

is solvable in rationals. A change of variables gives the equation \( y^2 = x^3 - D^2x \), which is an elliptic curve. One can show that the original system is solvable precisely when there are infinitely many rational points on the elliptic curve. In 1983, Tunnell made a connection with the theory of modular forms and obtained a sufficient condition for an integer to be non-congruent; the other direction relies on the Birch/Swinnerton-Dyer conjecture and remains unsolved.

Example 2 In 1982, Frey realized that a non-trivial solution \((a, b, c)\) to the Fermat equation \(x^n + y^n = z^n\) gives rise to an elliptic curve \(y^2 = x(x - a^n)(x + b^n)\), which can be shown to be non-modular. This would contradict the conjecture by Taniyama and Shimura that every elliptic curve is modular; thus Taniyama-Shimura (actually the weaker special case of semistable curves) implies “Fermat’s last theorem”. For more on this, see [41].

3.3.2 The most general form of the equation

Consider a cubic polynomial in two variables over some field \(k\), say \(p(x, y)\). It defines a curve, namely the locus of points with \(p(x, y) = 0\). This curve is an affine variety; by homogenization, it gives rise to a projective curve. Assume that the projective curve has a \(k\)-rational point \(O\); take the tangent at that point as \(Z\)-axis in projective coordinates. Take as \(X\)-axis a point of intersection between the \(Z\)-axis and the cubic, and as \(Y\)-axis any other line through \(O\). Dehomogenizing with respect to \(Z\) \((x = X/Z, y = Y/Z)\) gives an equation of the form

\[
xy^2 + axy + by = cx^2 + dx + e,
\]

and multiplication by \(x\) and the change of variables \(y = xy\) gives

\[
y^2 + (ax + b)y = cx^3 + dx^2 + ex.
\]

If \(\text{char}(k)\), the characteristic of \(k\), does not equal 2, the square in the left hand side can be completed, and if in addition \(\text{char}(k) \neq 3\), the cubic in the right hand side can be depressed. The most general form of a cubic in this case is

\[
y^2 = cx^3 - g_2x - g_3,
\]
and finally taking \( c = 4 \), one obtains the Weierstraß normal form. An equation \( y^2 = p(x) \) with \( p \) of degree 4 can be transformed to this form via rational changes of variables, sending one of the roots of \( p \) to infinity. An alternative standard form is

\[
E_\lambda : y^2 = x(x - 1)(x - \lambda),
\]

which is called the Legendre normal form. From this form, it is apparent that the parameter \( \lambda \) is not uniquely determined by the curve, because the roots of the original cubic can be permuted before being sent to \( \{0, 1, \lambda\} \). Thus, one expects a group isomorphic to \( S_3 \), the symmetric group on 3 letters, to act on \( \lambda \). Indeed, suppose that \( E_\lambda \) and \( E_\mu \) are isomorphic elliptic curves. Then the change of variables

\[
x = u^2 x' + r, \quad y = u^3 y'
\]

(which is the most general possible if the form of the equation and its intersection with the line at infinity are to be preserved, see [38]) gives

\[
x(x - 1)(x - \lambda) = \left(x + \frac{r}{u^2}\right) \left(x + \frac{r - 1}{u^2}\right) \left(x + \frac{r - \lambda}{u^2}\right),
\]

so \( \mu \) is one of

\[
\{\lambda, 1 - \lambda, 1/\lambda, 1/1 - \lambda, \lambda, 1/\lambda - 1, 1 - \lambda, \lambda - 1\},
\]

and this set of Möbius transformations is a group permuting \( \{0, 1, \infty\} \), hence isomorphic to \( S_3 \).

### 3.3.3 The group structure

As will be described below, one may consider one and the same equation over several different fields. We write \( E(k) \) for the solution set in the field \( k \), and calls it a curve over \( k \). For curves over \( \mathbb{R} \), the group structure can be described geometrically. First, note that for \( p \) of degree three, a chord through two points on the curve generically intersects it in a third point as well. By adjoining a point at infinity, vertical lines also have three point of intersection (ignoring, for the moment, tangent lines)\(^2\). If \( P \) and \( Q \) are points on the curve, denote by \( P \ast Q \) the third point of intersection. The binary operation \( \ast \) does not make the curve a group, as it lacks an identity element. Any point on the curve can be chosen as the identity; suppose the choice has been made, and denote the point \( \mathcal{O} \). Next, define an operation \( + \) by \( P + Q = (P \ast Q) \ast \mathcal{O} \). Clearly this makes \( \mathcal{O} \) an identity: \( P + \mathcal{O} = (P \ast \mathcal{O}) \ast \mathcal{O} = P \) for any \( P \), and any \( P \) has an inverse, denoted by \(-P\) and given by \(-P = P \ast \mathcal{O}\). The operation \( \ast \) is commutative, hence so is \( + \). The only thing remaining to check is associativity, which however is somewhat involved to do geometrically. It is common practice to take the point at infinity as the identity; the \("\ast \mathcal{O}\"-operation then corresponds to a reflection in the \( x \)-axis. Finally, a line tangent to the curve at the point \( P \) can be thought of as the limit of the family of lines through \( P \) and \( Q \) as \( Q \to P \); this allows one to add points to themselves. To summarize, the operation \("+\) defined above makes the points of \( E(\mathbb{R}) \) an abelian group. This is in fact the same structure as the addition in \( \mathbb{C} \), described intrinsically. The next section

---

\(^2\)In projective coordinates, Bézout’s theorem can be applied directly
describes the operation algebraically. Figure 3.6 shows addition of the points $P_1 = (-\frac{1}{\sqrt{2}}, -\sqrt[4]{\frac{4}{27}})$ and $P_2 = (0, 0)$ on the curve $y^2 = x^3 - x$.

![Figure 3.6: Addition on the elliptic curve $y^2 = x^3 - x$](image)

### 3.3.4 Algebra

We now derive formulæ for the group operation. Consider the curve $y^2 = x^3 + ax^2 + bx + c$, and let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be two points on it. The line through $P_1$ and $P_2$ has equation

$$y = \lambda x + \nu, \quad \lambda = \frac{y_2 - y_1}{x_2 - x_1}, \quad \nu = y_1 - \lambda x_1,$$

and substitution in the equation gives, after rearrangement,

$$x^3 + (a - \lambda^2)x^2 + (b - 2\lambda\nu)x + (c - \nu^2) = 0.$$  

The roots of this equation are the $x$-coordinates of $P_1$, $P_2$ and $P_1 + P_2$; call this last coordinate $x_3$. The left-hand side factors as

$$(x - x_1)(x - x_2)(x - x_3) = x^3 - (x_1 + x_2 + x_3)x^2 + \cdots,$$
and comparing \( x^2 \)-coefficients gives

\[
x_3 = \lambda^2 - a - x_1 - x_2 = \left( \frac{y_2 - y_1}{x_2 - x_1} \right)^2 - a - x_1 - x_2.
\]

This is the \( x \)-coordinate of \( P_3 = P_1 + P_2 \). The \( y \)-coordinate equals

\[
y_3 = -(\lambda x_3 + \nu),
\]

where the minus sign accounts for the reflection in the \( x \)-axis. To add a point to itself, the tangent is used instead:

\[
\lambda = \frac{3x_1^2 + 2ax + b}{2y_1}.
\]

The inverse, finally, is given simply by \( -(x, y) = (x, -y) \). This construction is easily extended to more general equations (i.e. arbitrary fields), and it remains true that the group operations are expressible as rational functions of the coordinates.

**Example** The curve \( y^2 = x^3 + 17 \) contains the rational points \( P_1 = (2, 5) \) and \( P_2 = (-2, 3) \). Using the formulæ above, one finds:

\[
2P_1 = \left( \frac{-64}{25}, \frac{59}{125} \right),
\]

\[
P_1 + P_2 = \left( \frac{1}{4}, -\frac{33}{8} \right),
\]

\[
2P_1 - P_2 = (43, 282),
\]

\[
-P_1 + 3P_2 = (52, 375).
\]

All linear combinations of these points have rational coordinates, but the numerators and denominators grow large quickly: for example,

\[
5P_1 = \left( \frac{279124379042}{111229587121}, \frac{212464088270704525}{37096290830311831} \right).
\]

## 3.3.5 Elliptic curves over different fields

Over \( \mathbb{C} \), an elliptic curve is simply a torus with a choice of origin for the group operation. The torus is the quotient of \( \mathbb{C} \) by a lattice \( \Lambda \), and the group operation arises from addition in \( \mathbb{C} \). The lattice gives rise to Eisenstein series and hence to an equation \( y^2 = 4x^3 - g_2x - g_3 \), and conversely, for any such nonsingular equation, a corresponding lattice can be found. Then the Weierstraß elliptic functions parametrize the curve. Over other fields, the picture is more complicated. The cases of curves over \( \mathbb{R} \) and \( \mathbb{Q} \) have already been discussed. One can also consider curves over number fields (finite-dimensional extensions of \( \mathbb{Q} \)) and over finite fields. As noted above, for some fixed equation, one denotes by \( E(k) \) the elliptic curve defined over the field \( k \).

In the early 1900s, Poincaré pointed out that the group \( E(k') \) is a subgroup of \( E(k) \) when \( k' \) is a subfield of \( k \), and raised the question whether the group \( E(\mathbb{Q}) \) is finitely generated. That this is indeed the case was proved by Mordell in 1922. Later, in his doctoral dissertation, Weil extended the result to arbitrary abelian
varieties over number fields (finite field extensions of $\mathbb{Q}$). Abelian varieties are briefly discussed in section 3.3.10 below.

The group $E(\mathbb{Q})$, being a finitely generated abelian group, is isomorphic to the direct product of its torsion subgroup $E(\mathbb{Q})_{\text{tor}}$ and $\mathbb{Z}^r$ for some integer $r$; $r$ is called the rank of the group, and, by extension, of the curve:

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tor}} \times \mathbb{Z}^r.$$  \hspace{1cm} (3.20)

Work by Mazur (1977) shows that $E(\mathbb{Q})_{\text{tor}}$ is isomorphic to the cyclic group $C_N$ with $1 \leq N \leq 10$ or $N = 12$, or the direct product $C_2 \times C_2N$ with $1 \leq N \leq 4$.

The rank is much more difficult to understand. It is not even known (but widely believed to be true) that there exists curves with arbitrarily high rank. The elliptic curve with the (currently) highest known rank is the following, found by Elkies in 2006:

$$y^2 + xy + y = x^3 - x^2 - 2006776241557526585033208299338542750930230312178956502x + 344816117950305564670329856903907203748559443593191803\ldots$$

$$\ldots 61266008296291939448732243429.$$  \hspace{1cm} (3.21)

It has trivial torsion and rank at least 28. The Birch and Swinnerton-Dyer conjecture asserts that the rank is equal to the order of vanishing of the associated $L$-function $L(E, s)$ at $s = 1$ (see the next section). The conjecture is supported by numerical evidence, but partial proofs exist only for rank $\leq 1$.

### 3.3.6 The $L$-series of an elliptic curve

Let $E$ be an elliptic curve over $\mathbb{Q}$ and $p$ a prime. If reducing the equation modulo $p$ gives a non-singular elliptic curve, $E$ is said to have good reduction at $p$. Let $N_p$ be the number of points on $E$ over $\mathbb{F}_p$ (excluding the point at infinity) and let
$a_p = p - N_p$. One defines the local $L$-series as $L_p(T) = 1 - a_pT + T^2$ for primes with good reduction; for the (finitely many) primes with bad reduction, another local series is used. The local data is then assembled into a global $L$-series:

$$L(E, s) = \prod_p \frac{1}{L_p(p^{-s})}.$$  

The product has analytic continuation to the entire complex plane: this follows from the modularity theorem. Define a function

$$G(s) = \left( \frac{\sqrt{N}}{2\pi} \right)^s \Gamma(s)L(E, s), \quad (3.22)$$

where $N$ is an integer called the conductor of the curve; its prime factors are the primes of bad reduction. It satisfies a functional equation:

$$G(s) = \pm G(2 - s). \quad (3.23)$$

The Birch/Swinnerton-Dyer conjecture asserts that the rank of the curve $E$ equals the order of vanishing of $L(E, s)$ at $s = 1$. This is a specific example illustrating a general principle: algebraic properties of an object imply analytic properties of a suitable associated $L$-series.

### 3.3.7 Moduli spaces of elliptic curves

Recall from 2.5.3 that any point in $X(1)$ corresponds to a lattice, and conversely, any lattice corresponds to precisely one point in $X(1)$. Moreover, the geometry of $X(1)$ reflects the properties of sets of lattices: lattices that are “close” correspond to “close” points in the moduli space. Since the (equivalence classes of) lattices are in correspondence with elliptic curves, this construction actually parametrizes elliptic curves. Thus, one calls $X(1)$ the moduli space of elliptic curves. More generally, the modular curve $X(N)$ parametrizes elliptic curves with a preferred basis for their $N$-torsion subgroup. Similarly, the surface $X_0(N)$ parametrizes elliptic curves with a chosen cyclic subgroup of order $N$, and $X_1(N)$ parametrizes elliptic curves with a chosen point of order $N$; see [22]. Note that it is intuitive that elliptic curves with extra structure have larger moduli spaces, which are obtained by forming quotients by smaller groups.

### 3.3.8 Complex multiplication

Let $E$ be an elliptic curve. The map from $E$ to itself, given by $P \mapsto nP$ is clearly a endomorphism for any (real) integer $n$. An elliptic curve with endomorphisms not arising in this way is said to admit complex multiplication. Here, we will only deal with elliptic curves over $\mathbb{C}$. Consider an morphism between elliptic curves $\phi : E \rightarrow E'$. It lifts to an endomorphism $\hat{\phi}$ of the group $(\mathbb{C}, +)$, which fixes 0 and hence must be of the form $\hat{\phi}(z) = cz$ for some $c \in \mathbb{C}$:

$$\begin{array}{ccc}
\mathbb{C} & \xrightarrow{\hat{\phi}} & \mathbb{C} \\
\pi \downarrow & & \pi \\
\mathbb{C}/\Lambda & \xrightarrow{\phi} & \mathbb{C}/\Lambda'
\end{array}$$
For the diagram to commute, one must have \( c\Lambda \subseteq \Lambda' \); equality corresponds to a surjective map (automorphism). The only real \( c \) for which the condition is satisfied are real integers (the surjective ones being \( c = \pm 1 \), corresponding to the identity mapping and the map sending an element to its inverse, which is an automorphism since the group is Abelian). These are the inner endomorphisms. All other endomorphisms arise from non-real \( c \), motivating the name complex multiplication. For example, the curve \( y^2 = x^3 - x \) is the quotient of \( \mathbb{C} \) by the square lattice \( \mathbb{Z} + i\mathbb{Z} \) and has an automorphism arising from the map \( z \mapsto iz \).

In terms of coordinates, it is \((x, y)\mapsto(-x, iy)\), see [43], chapter 3.

As for any abelian group, the elements of a fixed order \( n \) in a curve \( E \) over some field \( k \) form a subgroup, called the \( n \)-torsion group, or in this context the \( n \)-division points. This group is denoted \( E[n] \). Addition being expressible by rational functions, the equation \( nP = O \) has solutions with coordinates which are algebraic over \( k \). Considering the group structure in \( \mathbb{C} \) shows that \( E[n] \cong (\mathbb{Z}/n\mathbb{Z})^2 \).

### 3.3.9 Galois representations

Galois representations are homomorphisms from some Galois group of interest to \( \text{GL}(n, R) \) for some integer \( n \) and ring \( R \). The most interesting and general questions concern the “absolute” Galois group \( G = \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \) for some algebraic closure \( \overline{\mathbb{Q}} \) of \( \mathbb{Q} \). This object is however very large and only defined up to inner automorphisms; this is what necessitates the study of its representations instead.

We now describe how to get Galois representations from elliptic curves. The construction uses the so-called Tate module. Let \( E \) be an elliptic curve over a field \( k \) and \( n \geq 2 \) an integer coprime to \( \text{char}(k) \) (if finite). The equation \( nP = O \) is an algebraic equation, so the elements of \( \text{Gal}(\overline{k}/k) \) map the group \( E[n] \) of \( n \)-division points to itself, respecting the group structure. Hence there is a homomorphism \( \rho : \text{Gal}(\overline{k}/k) \to \text{Aut}(E[n]) \cong \text{GL}(2, \mathbb{Z}/n\mathbb{Z}) \). This already captures some information about the Galois group but is not entirely satisfactory since \( \mathbb{Z}/n\mathbb{Z} \) has finite characteristic. It is natural, then, to study representations on \( \text{GL}(2, \mathbb{Z}_{\ell^n}) \) for primes \( \ell \) and then pass to the inverse limit \((n \to \infty)\), obtaining a representation on \( \text{GL}(2, \mathbb{Z}_{\ell}) \) instead. One has a system of projections

\[ E[\ell^{n+1}] \rightarrow \pi_{\ell^n} E[\ell^n], \]

given by \( P \mapsto \ell P = P + P + \ldots + P \), where there are \( \ell \) terms. The Tate module is the inverse limit:

\[ T_\ell(E) = \lim_n E[\ell^n], \]

which is a \( \mathbb{Z}_\ell \)-module with a natural topology. The action of \( \text{Gal}(\overline{k}/k) \) is continuous and commutes with the maps \( \pi_{\ell^n} \), giving a continuous action on \( T_\ell(E) \) as well; if \( \ell \neq \text{char}(k) \), \( T_\ell(E) \cong \mathbb{Z}_\ell^2 \), so \( \text{Aut}(\mathbb{Z}_\ell^2) \cong \text{GL}(2, \mathbb{Z}_\ell) \).

### 3.3.10 Hyperelliptic curves and Abelian varieties

Allowing the polynomial in \( y^2 = p(x) \) to have higher degree than four (while still requiring the roots to be distinct) gives a so-called hyperelliptic curve. For \( p \) of degree \( n \), the genus of the curve equals \( \left\lfloor \frac{n-1}{2} \right\rfloor \).
Another generalization of elliptic curves is the concept of an Abelian variety: for an arbitrary integer \( g \geq 1 \), form the quotient of \( \mathbb{C}^n \) by a lattice of rank \( 2g \). The resulting space is torus of complex dimension \( g \), with a group structure inherited from \( \mathbb{C}^n \). If in addition the torus is a projective variety over \( \mathbb{C} \), it is called an Abelian variety. For \( g = 1 \), any complex torus is an Abelian variety, but this does not hold for higher genera.

To any non-singular complex curve, one associates an Abelian variety, called the Jacobian of the curve, containing the curve as a subvariety. The Jacobian, as a group, is generated by the curve. See [43], chapter 2.
Chapter 4

Dirichlet series

A Dirichlet series is a series of the form

\[ F(s) = \sum_{n=1}^{\infty} a_n e^{-\lambda_n s}, \quad (4.1) \]

where \( a_n \) are complex numbers and \( \lambda_n \) are real numbers tending to infinity with \( n \). The variable \( s \) is allowed to take complex values. For the general theory of such series, see [19]. Taking \( \lambda_n = n \) gives a power series in \( e^{-s} \). Often, one takes instead \( \lambda_n = \ln n \) (so-called normal Dirichlet series), which have the form

\[ F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}. \quad (4.2) \]

Such series were studied by Dirichlet; see, for example, [25]. If \( |a_n| = O(n^c) \) for some \( c \), the series is absolutely convergent in the half-plane \( \Re s > c + 1 \). Similarly, the convergence is uniform in some half-plane. These half-planes need not coincide.

Examples include \( a_n = 1 \Rightarrow F(s) = \zeta(s) \) (the Riemann zeta function). The next example requires the notion of a multiplicative character. Fix an integer \( N \) and consider a group homomorphism \( \chi \) from the additive group \( \mathbb{Z}/N\mathbb{Z} \) to the multiplicative group \( (\mathbb{C}^*, \cdot) \) (nonzero complex numbers; note that, \( \mathbb{Z}/N\mathbb{Z} \) being finite, the image must actually consist of roots of unity). Extend \( \chi \) periodically to a function \( \chi : \mathbb{Z} \to \mathbb{C}^* \). Such a function is called a (multiplicative) character modulo \( N \). A series of the form

\[ L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}, \quad (4.3) \]

is called a Dirichlet \( L \)-series. They were used by Dirichlet to prove that every arithmetic progression \( \{an + b : n \in \mathbb{N}, (a, b) = 1\} \) contains infinitely many primes.

Dirichlet series satisfy

\[ \left( \sum_{n=1}^{\infty} \frac{f(n)}{n^s} \right) \left( \sum_{n=1}^{\infty} \frac{g(n)}{n^s} \right) = \sum_{n=1}^{\infty} \frac{(f \ast g)(n)}{n^s}, \quad (4.4) \]

where the sums are absolutely convergent, and where
\[ (f * g)(n) = \sum_{d \mid n} f(d) g\left(\frac{n}{d}\right) \]  
(4.5)

is a Dirichlet convolution. This implies, for example, identities such as
\[ \frac{1}{\zeta(s)} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}, \]  
(4.6)

where \( \mu \) is the Möbius function, and
\[ \zeta^2(s) = \sum_{n=1}^{\infty} \frac{d(n)}{n^s}, \]  
(4.7)

where \( d(n) \) is the number of divisors of \( n. \) More generally, it suggests that a Dirichlet series should be thought of as some sort of transform of the sequence \( \{a_n\}; \) cf. the convolution theorem for Fourier transforms. Indeed, a Dirichlet series is a special case of the Laplace-Stieltjes transform, defined by
\[ \int_{0}^{\infty} e^{-st} d\alpha(t), \]  
(4.8)

which for \( d\alpha(t) = \alpha'(t)dt \) gives the ordinary Laplace transform of \( \alpha' \), and for \( \alpha \) a step function with jumps \( \alpha(n) \) at \( \lambda_n \) gives the general Dirichlet series. This will be developed further in the next section.

### 4.1 The Mellin transform

The Mellin transform of a function \( f \) is defined as
\[ M\{f\}(s) = \int_{0}^{\infty} f(t) t^s \frac{dt}{t}, \]  
(4.9)

so that, for example, the \( \Gamma \) function is the Mellin transform of the function \( e^{-t} \):
\[ \Gamma(s) = \int_{0}^{\infty} e^{-t} t^{s} \frac{dt}{t}. \]  
(4.10)

The Mellin transform is intimately connected to the Fourier transform, defined (essentially) by
\[ \mathcal{F}\{f\}(\xi) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \xi t} dt. \]  
(4.11)

To see this, note that
\[ \mathcal{F}\{e^{-t}\}(\xi) = \int_{-\infty}^{\infty} f(t) e^{-2\pi i \xi t} dt = \int_{-\infty}^{\infty} f(y) e^{2\pi i \xi \frac{dy}{y}} \]  
(4.12)

after the change of variable \( e^{-t} = y. \) The measure \( \frac{dt}{t} \) is invariant under dilations instead of translations, i.e., it is the Haar measure for the multiplicative
4.1. The Mellin transform

(\mathbb{R}_+ \cdot \cdot). Note that $t^s$ are the characters of this group. A key example is the following. Define the theta function

$$\theta(z) = \sum_{n \in \mathbb{Z}} e^{\pi in^2z}. \quad (4.13)$$

Then

$$\pi^{-s} \Gamma(s) \zeta(2s) = \int_0^\infty \frac{1}{2} \left( \theta(it) - 1 \right) t^s \frac{dt}{t}, \quad (4.14)$$

because

$$\frac{1}{2} \left( \theta(it) - 1 \right) = \sum_{1}^\infty e^{-\pi n^2 it}, \quad (4.15)$$

giving

$$\sum_{1}^\infty \int_{-\infty}^\infty e^{-\pi n^2 t^s} t^s \frac{dt}{t} = \sum_{1}^\infty \frac{1}{(\pi n^2)^s} \int_{-\infty}^\infty e^{-\pi n^2t^s} dt = \pi^{-s} \Gamma(s) \sum_{1}^\infty \frac{1}{n^{2s}}. \quad (4.16)$$

This works for any Fourier series, sending it to a Dirichlet series times a gamma factor: if

$$f(\tau) = \sum_{n=1}^\infty a_n e^{2\pi in\tau}, \quad (4.17)$$

(note the omitted constant term), one has

$$Mf(i\tau) = (2\pi)^{-s} \Gamma(s) \sum_{n=1}^\infty \frac{a_n}{n^s} \quad (4.18)$$

after an analogous calculation (see, for example [43], chapter 4). The theta function above satisfies

$$\left(\frac{i}{t}\right) = -\frac{1}{i\theta}, \quad (4.19)$$

which implies a functional equation for $\zeta$. First, we prove this so-called theta inversion formula. Consider the gaussian

$$g(x) = e^{-\pi tx^2}, \quad t > 0, \quad (4.20)$$

and its Fourier transform

$$\mathcal{F}\{g\}(\xi) = \hat{g}(\xi) = \frac{1}{\sqrt{t}} e^{-\pi \xi^2/t}. \quad (4.21)$$

Poisson's summation formula,

$$\sum_{n \in \mathbb{Z}} g(n) = \sum_{n \in \mathbb{Z}} \hat{g}(n), \quad (4.22)$$

gives in this case

$$\theta(it) = \sum_{n \in \mathbb{Z}} e^{-\pi tn^2} = \frac{1}{\sqrt{t}} \sum_{n \in \mathbb{Z}} e^{-\pi n^2/t} = \frac{1}{\sqrt{t}} \theta(i/t), \quad (4.23)$$
proving the assertion. Next, introduce the auxiliary function

\[ G(s) = \pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \int_0^\infty \frac{1}{2} \left( \theta(it) - 1 \right) t^{s/2} \frac{dt}{t}, \]  

(4.24)

and split the integral as \( \int_0^\infty = \int_0^1 + \int_1^\infty \). The interval \([0, 1]\) is sent to \([1, \infty]\) by the substitution \( u = 1/t \):

\[
\begin{align*}
\int_0^1 \frac{1}{2} \left( \theta(it) - 1 \right) t^{s/2} \frac{dt}{t} &= \frac{1}{2} \int_0^1 \theta(it) t^{s/2} \frac{dt}{t} - \frac{1}{s} = \\
\frac{1}{2} \int_0^1 \theta(-\frac{1}{it}) t^{(s-1)/2} \frac{dt}{t} - \frac{1}{s} &= \frac{1}{2} \int_1^\infty \theta(itu) u^{(1-s)/2} \frac{du}{u} - \frac{1}{s} = \\
\int_1^\infty \frac{1}{2} \left( \theta(itu) - 1 \right) u^{(1-s)/2} \frac{du}{u} - \left( \frac{1}{s} + \frac{1}{1 - s} \right),
\end{align*}
\]

giving the representation

\[ G(s) = \int_1^\infty \frac{1}{2} \left( \theta(it) - 1 \right) (t^{s/2} + t^{(1-s)/2}) \frac{dt}{t} - \left( \frac{1}{s} + \frac{1}{1 - s} \right). \]  

(4.25)

The function \( \theta(it) \) tends to 0 rapidly as \( t \to \infty \), so the integral defines a function which is analytic everywhere. Thus \( G(s) \) is meromorphic everywhere, with simple poles at \( s = 0, s = 1 \) and no other places. The integral representation makes it obvious that \( G(s) = G(1-s) \), which is the functional equation for the Riemann \( \zeta \) function. Explicitly,

\[
\pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \zeta(s) = \pi^{-(1-s)/2} \Gamma\left(\frac{1-s}{2}\right) \zeta(1-s).
\]  

(4.26)

Finally, we note that the factor \( \pi^{-s/2} \Gamma\left(\frac{s}{2}\right) \) is nowadays interpreted as corresponding to a “prime at infinity”, rather than something extrinsic to the zeta function.

4.2 Euler products

For the history of Euler products, see [42], chapter 6. The archetypal Euler product is Euler’s formula for the classical zeta function, found in 1737:

\[ \zeta(s) = \prod_p \frac{1}{1 - \frac{1}{p^s}}, \]  

(4.27)

where \( p \) runs over all primes. The formula can be thought of as expressing the fundamental theorem of arithmetic: expand the factors in geometric series:

\[
\frac{1}{1 - \frac{1}{p^s}} = 1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots,
\]  

(4.28)

and notice that in the product, every positive integer will occur as denominator exactly once. Euler then notes that the harmonic series diverges, so

\[
\prod_{p \leq N} \left(1 - \frac{1}{p}\right) \to \infty \quad \text{as} \quad N \to \infty.
\]  

(4.29)
Thus the sum
\[ \sum_p \frac{1}{p} \]  
(4.30)
diverges as well. This was arguably the first time an arithmetic fact was proved by analytic methods, and the calculation marks the birth of analytic number theory. Euler went further, considering products of the form
\[ \prod_p \frac{1}{1 - \frac{f(p)}{p^s}}, \]  
(4.31)
where the function \( f \) is fully multiplicative, i.e.
\[ f(mn) = f(m)f(n) \]  
(4.32)
for all integers \( m, n \). A similar argument now gives
\[ \sum_{n=1}^{\infty} \frac{f(p)}{p^{ns}} = \prod_p \left( 1 - \frac{f(p)}{p^s} \right)^{-1}, \]  
(4.33)
which can be used to prove analogous results for primes belonging to prescribed residue classes modulo fixed integers. A later generalization of the classical Euler products is to consider products of the form
\[ \prod_p P(p^{-s})^{-1}, \]  
(4.34)
where \( P \) is a polynomial with constant term equal to 1. The degree of the polynomial is called the degree of the product. Many such products (which, too, are called \( L \)-functions) have been constructed in arithmetic and geometry. An example is the \( L \)-function of an elliptic curve, to be defined in the next chapter. Another example is the \( L \)-function of a modular form (see 5.4). These objects encode arithmetic and geometric data as analytic properties of the \( L \)-function, allowing one to translate problems from one area to another. The classical reciprocity laws can be understood in this way as equalities between \( L \)-functions. The corresponding Euler products are of degree one. The current challenge is to obtain a similar theory for products of higher degree. The recent proof of the modularity theorem (chapter 5.5.2) is a partial result in this direction. Another description of this line of research is as search for a “nonabelian class field theory”. The classical class field theory of Kronecker and Weber shows that any field extension \( k \) of \( \mathbb{Q} \) with abelian Galois group is contained in a cyclotomic extension, say \( \mathbb{Q}(\zeta_{2N}) \). The Langlands program is a set of conjectures extending this to higher-dimensional representations (the non-abelian case). See [17] for an introduction.

4.3 Adeles

The topics in this chapter are classical; much of the material has been reformulated in more modern terms during the 20th century. The notion of an adele, in particular, is central in current thinking. This section gives a brief introduction.
In arithmetic, a natural notion of size should reflect the multiplicative behaviour of primes. This lead Kurt Hensel (in 1897) to introduce \( p \)-adic numbers. First, given a prime \( p \), define the order of a rational number \( r \), denoted by \( v_p(r) \), as the unique integer \( n \) such that
\[
    r = p^n s t,
\]
where \( s \) and \( t \) are integers not divisible by \( p \); in addition, let \( v_p(0) = \infty \) for every \( p \). Clearly \( v_p \) satisfies
\[
    v_p(qr) = v_p(q) + v_p(r),
    v_p(q + r) \geq \min(v_p(q), v_p(r)),
    v_p(q) = \infty \quad \text{if and only if} \quad q = 0.
\]
Any such map from a field to \( \mathbb{R} \cup \{ \infty \} \) is called a valuation. Next, define the \( p \)-adic norm (or absolute value) as
\[
    |r|_p = p^{-v_p(r)}.
\]
This is indeed a norm, satisfying
\[
    |0|_p = 0,
    |1|_p = 1,
    |qr|_p = |q|_p|r|_p,
    |q + r|_p \leq |q|_p + |r|_p.
\]
One considers equivalence classes of norm, where two norms are considered equivalent if one is a power of the other; the equivalence classes are called places. In fact, by virtue of the inequality satisfied by a valuation, the \( p \)-adic norm satisfies a stronger condition, namely
\[
    |q + r|_p \leq \max(|q|_p, |r|_p). \tag{4.36}
\]
Such a norm gives rise to a so-called ultrametric. Conversely, it is easy to see that any ultrametric norm arises from a valuation. The corresponding places are called finite (or nonarchimedean). By contrast, the usual absolute value is called infinite (archimedean). It is accordingly written \( |.|_\infty \). This explains the remarks on a “prime at infinity” made in section 4.1.

The \( p \)-adic numbers, denoted \( \mathbb{Q}_p \), are the completion of \( \mathbb{Q} \) with respect to the \( p \)-adic norm. A general \( p \)-adic number can be written
\[
    \sum_{n=m}^{\infty} a_n p^n,
\]
where \( m \) is an integer and \( a_n \) belongs to \( \{0, 1, \ldots, p - 1\} \); the sum converges with respect to the norm \( |.|_p \). The \( p \)-adic numbers for which \( m \geq 0 \) form a subring, called \( p \)-adic integers and denoted \( \mathbb{Z}_p \). Note that several conventions for the representation of \( p \)-adics exist; Teichmüller used instead the \( p \)th roots of unity as coefficients \( a_n \). The notion of Witt vector encompasses both of these conventions.
The $p$-adics are a natural framework for the local study of equations. Indeed, if an equation has a solution in integers (or rationals), the solution descends to a solution modulo every prime; counterpositively, if some equation lacks solutions locally (“at one place”), this “local obstruction” shows that no solution exists globally. Unfortunately, the converse does not hold in general.

**Example** As shown by Thue (see [37]), the equation $3x^3 + 4y^3 + 5z^3 = 0$ has solutions modulo every prime but no solution over $\mathbb{Z}$.

In any case, one wants to complete $\mathbb{Q}$ with respect to all places simultaneously; this motivates the study of the profinite completion:

$$\hat{\mathbb{Z}} = \varprojlim \mathbb{Z}/n\mathbb{Z},$$

which by the Chinese remainder theorem factors as

$$\hat{\mathbb{Z}} = \prod_p \mathbb{Z}_p.$$
Chapter 5

Modular forms

5.1 Definitions

The theory of modular forms is treated in [1], [11], [13], [36], [39] and [43], chapter 4. Modular forms can be approached in several different ways. We begin by defining them as functions on lattices.

First definition A modular form $f$ is a complex-valued function on lattices, satisfying the following conditions:

1. For a fixed $\lambda_1$, $f(\Lambda(\lambda_1, \lambda_2))$ is a holomorphic function of $\lambda_2$.
2. There is a constant $k$, such that $f(z\Lambda) = f(\Lambda(z\lambda_1, z\lambda_2)) = z^{-k}f(\Lambda)$ for any nonzero $z$.
3. $f$ is bounded from above if the smallest nonzero element of $\Lambda$ is bounded from below.

The constant $k$ is called the weight of the form. Next, we define modular forms as functions on $\mathcal{H}$, by putting $f(\tau) = f(\Lambda_{\tau})$. Suppose that $\left(\begin{array}{cc}a & b \\ c & d \end{array}\right) \in \Gamma$. Then one has:

$$f(\frac{a\tau + b}{c\tau + d}) = f(\Lambda(1, \frac{a\tau + b}{c\tau + d})) = (c\tau + d)^k f(\Lambda(c\tau + d, a\tau + b)) = (c\tau + d)^k f(\tau),$$

motivating the following definition:

Second definition A modular form for $\Gamma$ is a holomorphic function $f : \mathcal{H} \rightarrow \mathbb{C}$, bounded as $\tau \rightarrow i\infty$, satisfying

$$f(\gamma \tau) = f(\frac{a\tau + b}{c\tau + d}) = (c\tau + d)^k f(\tau) \quad (5.1)$$

for all $\gamma \in \Gamma$.

Modular forms for subgroups of $\Gamma$ are defined analogously, with the requirement that $f$ be holomorphic at the cusps (recall the definition of cusps from section Karlsson, 2009.)
2.5.3). If in addition \( f \) vanishes at the cusps, it is called a \textit{cusp form}. Since the translation \( \tau \mapsto \tau + 1 \) is contained in \( \Gamma \), modular forms are 1-periodic:
\[
f(\tau + 1) = f(\tau),
\]
(5.2)
and can be written as functions of \( q = e^{2\pi i \tau} \), defined in the punctured open unit disc \( \{ z : 0 < |z| < 1 \} \). The point \( q = 0 \) corresponds to \( \tau \to i\infty \); boundedness and the Riemann removability theorem then implies that \( f \) is holomorphic at \( q = 0 \) as well, and can be expanded in a Taylor series:
\[
f(q) = \sum_{n=0}^{\infty} a_n q^n = \sum_{n=0}^{\infty} a_n e^{2\pi i n \tau}.
\]
(5.3)

### 5.2 Examples

#### 5.2.1 Eisenstein series

The Eisenstein series, defined in section 3.1, are modular forms; the proper transformation under the action of \( \Gamma \) is readily verified, and the growth condition can be checked by inspection of their Fourier series, which we now determine. Start with the identity
\[
\pi \cot \pi \tau = \frac{1}{\tau} + \sum_{n=1}^{\infty} \left( \frac{1}{\tau + n} + \frac{1}{\tau - n} \right)
\]
(5.4)
and note that
\[
\pi \cot \pi \tau = \frac{\pi i}{q - 1} = \pi i \left( 1 - 2 \sum_{n=0}^{\infty} q^n \right).
\]
(5.5)
Equate the right-hand sides and differentiate \( k - 1 \) times with respect to \( \tau \). This gives
\[
\sum_{m \in \mathbb{Z}} \frac{1}{(\tau + m)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=0}^{\infty} n^{k-1} q^n.
\]
(5.6)
Next, rewrite (note that \( k \) is even) as
\[
G_k(\tau) = \sum_{(m,n):m \neq 0,0} \frac{1}{(n\tau + m)^k} = 2 \sum_{m=1}^{\infty} \frac{1}{m^k} + 2 \sum_{n=1}^{\infty} \sum_{m \in \mathbb{Z}} \frac{1}{(n\tau + m)^k},
\]
(5.7)
which, using (5.6), becomes
\[
G_k(\tau) = 2\zeta(k) + 2 \frac{(-2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n,
\]
(5.8)
where \( \sigma_k(n) \) denotes the sum of the \( k \)th powers of the divisors of \( n \):
\[
\sigma_k(n) = \sum_{d|n} d^k.
\]
(5.9)
Finally, using Euler’s formula relating values of the \( \zeta \) function for even integer arguments to the Bernoulli numbers (see appendix A),
\[
\zeta(2k) = -\frac{(2\pi i)^{2k}}{2(2k)!} B_{2k},
\]
(5.10)
one obtains

\[ G_{2k}(\tau) = 2\zeta(2k) \left( 1 - \frac{4k}{\pi} \sum_{n=1}^{\infty} \sigma_{2k-1}(n)q^n \right). \]  (5.11)

It is sometimes useful to normalize the constant term to one, putting

\[ E_{2k}(\tau) = \frac{G_{2k}(\tau)}{2\zeta(2k)}. \]  (5.12)

Thus, for example,

\[ E_4(\tau) = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n)q^n, \]

\[ E_6(\tau) = 1 - 504 \sum_{n=1}^{\infty} \sigma_5(n)q^n. \]  (5.13)

### Theta functions

Historically, a source of modular forms are the theta functions. These are not in general invariant under the action of \( \Gamma \), but transform in such a way that certain combinations of them are modular forms. The literature on theta functions is vast, and we will only give a short example. More on theta functions can be found in [8], [11] and [31]. The Jacobi theta function is defined as follows:

\[ \vartheta(z, \tau) = \sum_{n \in \mathbb{Z}} e^{\pi i n^2 \tau + 2\pi i nz}, \]  (5.14)

for \( z \in \mathbb{C} \) and \( \tau \in \mathcal{H} \). The sum can be shown to converge absolutely and uniformly on compact sets ([31]). One can show ([31],[11]) that \( \theta(\tau) = \vartheta_{00}(0, \tau)^k \) satisfies the equations

\[ \theta(-1/\tau) = (-i\tau)^{k/2}\theta(\tau), \quad \theta(\tau + 2) = \theta(\tau), \]  (5.15)

making it a modular form for \( \Gamma(2) \) of weight \( k/2 \).

#### 5.2.2 The modular invariant

The discriminant of the cubic polynomial \( 4x^3 - g_2x - g_3 \) is, up to constant factors, \( \Delta = g_2^3 - 27g_3^2 \); by construction, it vanishes when the roots of the polynomial are not all distinct. Since \( g_2 \) has weight 4 and \( g_3 \) has weight 6, the discriminant is a modular form of weight 12. It is related to the Dedekind eta function, given by:

\[ \eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n). \]  (5.16)

Namely, one has

\[ \Delta = (2\pi)^{12} \eta^{24}, \]  (5.17)

giving

\[ \Delta = (2\pi)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24}. \]  (5.18)
It is evidently a cusp form, as expected from its definition. The Fourier coefficients of $\eta^{24}$ are integers; one defines the Ramanujan function $\tau(n)$ by

$$q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum_{n=1}^{\infty} \tau(n)q^n.$$  \hspace{1cm} (5.19)

It has been widely studied and satisfies numerous identities, such as

$$\tau(m)\tau(n) = \sum_{d\mid(m,n)} d^{11}\tau\left(\frac{mn}{d^2}\right),$$  \hspace{1cm} (5.20)

$$\tau(n) \equiv \sigma_{11}(n) \pmod{691}, \text{ and}$$  \hspace{1cm} (5.21)

$$n^{610}\tau(n) \equiv \sigma_{1231}(n) \pmod{3^6} \text{ when } 3 \nmid n.$$  \hspace{1cm} (5.22)

Note that both $\Delta$ and $g_3^2$ are modular forms of weight 12. Thus, their quotient $g_3^2/\Delta$ is invariant under the modular group. It is not a modular form because it fails to be bounded at the cusp $i\infty$; instead, it is called a modular function. It is convenient to introduce a normalizing factor, defining

$$j = 12^{3/2} \frac{g_3^2}{\Delta}$$  \hspace{1cm} (5.23)

It has a simple pole at infinity. All its Fourier coefficients are integers. The expansion begins:

$$j(\tau) = \frac{1}{q} + 744 + 196884q + 21493760q^2 + \ldots.$$  \hspace{1cm} (5.24)

Liouville’s theorem shows that any modular function, i.e., meromorphic function invariant under the modular group, is a rational function of $j$; hence, it is called the modular function. The Fourier coefficients have remarkable properties, related to the so-called monster group. More precisely, the coefficients are (simple) linear combinations of the dimensions of irreducible representations of the monster, viz. 196884 = 1 + 196883. See [16].

Another remarkable fact is that $j\left(1+\sqrt{-D}\right)$, where $D$ is a squarefree natural number, is an integer when the ring $\mathbb{Z}\left[1+\sqrt{-D}\right]$ is a unique factorization domain; the largest $D$ for which this happens is 163, giving that

$$-e^{\pi\sqrt{163}} + 744 - 196884e^{-\pi\sqrt{163}} + \ldots$$

is an integer; $e^{\pi\sqrt{163}}$ is large, so the series is well approximated by the first two terms. Thus $e^{\pi\sqrt{163}}$ should be close to an integer, and indeed

$$e^{\pi\sqrt{163}} \approx 262537412640768743.99999999999925 \ldots$$

This is described in, for example, [13].

The modular function can also be expressed directly in terms of the coefficients of the equation of the elliptic curve. For the Legendre curve $y^2 = x(x-1)(x-\lambda)$, one has:

$$j(E) = 2^8\frac{(\lambda^2 - \lambda + 1)^3}{\lambda^2(\lambda-1)^2},$$  \hspace{1cm} (5.25)
5.3 Spaces of modular forms

and a similar but more complicated formula exists for elliptic curves over fields
with characteristic equal to 2 or 3. $j(E)$ is a complete invariant in the following
sense: if $E$ and $E'$ are elliptic curves, then $j(E) = j(E')$ if and only if $E$ and $E'$
are equivalent. We now show this (following [13]). By the discussion in section
3.3.2, it suffices to show that $j(\lambda) = j(\lambda')$ precisely when $\lambda'$ is one of
$\lambda, 1 - \lambda, 1/\lambda, 1/(1 - \lambda), \lambda/(\lambda - 1)$ or $(\lambda - 1)/\lambda$. To see that this holds, consider the
polynomial

$$p(x, \lambda) = (x - \lambda)(x - (1 - \lambda)) \left( x - \frac{1}{\lambda} \right) \left( x - \frac{\lambda}{\lambda - 1} \right) \left( x - \frac{\lambda - 1}{\lambda} \right),$$

and put $k = j/2^7$ for convenience. A straightforward calculation shows that

$$p(x, \lambda) = x^6 - 3x^5 - (k - 6)x^4 + (k - 7)x^3 - (k - 6)x^2 - 3x + 1,$$  \hspace{1cm} (5.26)

so $k$, hence also $j$, is invariant under the Möbius transformations in question.
For the other direction, suppose that $j(\lambda) = j(\lambda')$. Then $p(x, \lambda) = p(x, \lambda')$
holds identically, so that in particular $p(\lambda, \lambda') = p(\lambda, \lambda) = 0$. It follows that
$\lambda$ and $\lambda'$ are related by a Möbius transformation of the desired kind, proving
the assertion. It is also possible to construct elliptic curves with the desired
$j$-invariant: one finds that the curve $E$ given by

$$y^2 = x^3 - \frac{1}{2}x^2 - \frac{36}{j - 1728}x - \frac{1}{j - 1728}$$  \hspace{1cm} (5.27)

satisfies $j(E) = j$; also $y^2 = x^3 - 1$ has $j$-invariant 0 and $y^2 = x^3 - x$ has
$j$-invariant 1728.

5.3 Spaces of modular forms

Modular forms of a fixed weight $k$ obviously form a vector space over $\mathbb{C}$. In
fact, these spaces are all finite-dimensional, a consequence of the requirement
of boundedness at the cusps. Acting on these spaces is an algebra of operators,
introduced by Hecke. First, we determine the dimensions. This exposition
mostly follows [36].

Let $f(\tau)$ be a modular form of weight $k$. Denote by $v_p(f)$ the order of
vanishing of $f$ at a point $p$, i.e., the unique integer $n$ such that

$$f(\tau) = (\tau - p)^ng(\tau)$$  \hspace{1cm} (5.28)

for some analytic function $g$, which is nonzero in $p$. Define $v_\infty(f(\tau))$ as $v_0(f(q))$.
By assumption, $f$ is holomorphic, so there are no poles. Also, the zeroes are
isolated, so there is some neighbourhood of $q = 0$ containing no other zero. In
the fundamental region in the upper half-plane, this neighbourhood corresponds
to some number $M$ such that all zeroes except the one at $\{i\infty\}$ have imaginary
part $\leq M$. The intersection of the closure of the fundamental domain with the
set $\{ \tau : \Im \tau \leq M \}$ is compact, and hence contains only finitely many zeroes. Let
$N(f)$ be this number, counted with multiplicity. Enclose this region by arcs as
in figure 5.1, and suppose that $f$ has no zeroes on the arcs. Recall from 2.5.2
that the points $i$ and $(-1 + i\sqrt{3})/2$ are the fixed points of $\Gamma$ in the canonical
fundamental domain. The stabilizer subgroups have order 2 and 3, respectively.
Then, $N(f)$ is given by
\[ N(f) = \frac{1}{2\pi i} \int_A \frac{df}{f}, \]  

where \( A \) is the contour of integration. The integrals along \( A_1 \) and \( A_4 \) cancel (note that \( f(\tau + 1) = f(\tau) \)). Letting the radii of the circles enclosing \( \rho, i \) and \( \infty \) tend to zero, one obtains

\[ \frac{1}{2\pi i} \int_{A_5} \to -v_{\infty}(f), \]  
and similarly

\[ \frac{1}{2\pi i} \int_{C_1} \to -\frac{1}{6}v_{\rho}(f), \]  
\[ \frac{1}{2\pi i} \int_{C_2} \to -\frac{1}{6}v_{\rho}(f), \]  
\[ \frac{1}{2\pi i} \int_{C_3} \to -\frac{1}{2}v_i(f). \]  

Finally, the involution \( S : \tau \mapsto -\frac{1}{\tau} \) takes the arc \( A_3 \) to \( A_2 \), but \( f \) is not invariant under \( S \). Rather, \( f(S\tau) = z^k f(\tau) \), giving

\[ \frac{df(S(\tau))}{f(S(\tau))} = k \frac{d\tau}{\tau} + \frac{df}{f}, \]  
so

\[ \frac{1}{2\pi i} \int_{A_2 \cup A_3} \frac{df}{f} = \frac{1}{2\pi i} \int_{A_2} \left( \frac{df}{f} - k \frac{d\tau}{\tau} \right) \to \frac{k}{12}. \]
5.3. Spaces of modular forms

Taken together, this gives

\[ N(f) + v_\infty(f) + \frac{1}{3}v_\rho(f) + \frac{1}{2}v_i(f) = \frac{k}{12}. \]  

(5.36)

In the case that \( f \) has zeroes on the contour, indentions have to be made. It turns out that formula (5.36) is valid in these cases as well. All the variables on the left-hand side are nonnegative integers, so \( k \) must be even. The fraction \( 2/12 = 1/6 \) is not expressible in this way, so \( k \neq 2 \). On the other hand, the Eisenstein series give modular forms for every even integer \( \geq 4 \). Next, note that for a cusp form, \( v_\infty(f) \geq 1 \), giving \( k \geq 12 \). The modular discriminant \( \Delta \) gives an example of a cusp form of weight 12. Now take \( f = E_4 \):

\[ N(E_4) + \frac{1}{3}v_\rho(E_4) + \frac{1}{2}v_i(E_4) = \frac{1}{3}, \]  

(5.37)

giving \( v_\rho(E_4) = 1, v_i(E_4) = 0 \) and \( N(E_4) = 0 \). Similarly one finds \( v_\rho(E_6) = 0, v_i(E_6) = 1 \) and \( N(E_6) = 0 \). Furthermore,

\[ N(\Delta) + v_\infty(\Delta) + \frac{1}{3}v_\rho(\Delta) + \frac{1}{2}v_i(\Delta) = 1, \]  

(5.38)

and \( v_\infty(\Delta) \geq 1 \) forces \( N(\Delta) = v_\rho(\Delta) = v_i(\Delta) = 0, v_\infty(\Delta) = 1 \). Denote by \( \mathcal{M}_k \) the \( \mathbb{C} \)-vector space of level one modular forms of weight \( k \), and by \( \mathcal{S}_k \) the subspace of cusp forms. Further, put \( m_k = \dim_\mathbb{C} \mathcal{M}_k \) and \( s_k = \dim_\mathbb{C} \mathcal{S}_k \) and consider the evaluation homomorphism \( \phi : f \mapsto f(i\infty) \). Then, whenever there exists non-cusp forms of weight \( k \), i.e. when \( k \geq 4 \), the following sequence is exact, where the second arrow is inclusion:

\[ 0 \longrightarrow \mathcal{S}_k \longrightarrow \mathcal{M}_k \xrightarrow{\phi} \mathbb{C} \longrightarrow 0. \]

It follows that in these cases, \( m_k = s_k + 1 \). The Eisenstein series \( E_k \) belongs to \( \mathcal{M}_k \setminus \mathcal{S}_k \), so for \( k \geq 4 \),

\[ \mathcal{M}_k = \mathcal{S}_k \oplus \mathbb{C}E_k. \]  

(5.39)

Also, multiplication by the discriminant \( \Delta \in \mathcal{S}_{12} \) clearly gives an isomorphism from \( \mathcal{M}_k \) to \( \mathcal{S}_{k+12} \) for every \( k \), invertible because \( \Delta \) only has a simple zero at infinity. Thus \( \mathcal{M}_k \cong \mathcal{S}_{k+12} \) and \( m_k = s_{k+12} \). Knowing that \( s_k = 0 \) for \( k < 12 \) gives \( m_k \leq 1 \), and the Eisenstein series show that in fact \( m_k = 1 \) for \( 4 \leq k \leq 10 \), \( k \) even. This determines the dimensions for every \( k \). The following table shows some results:

<table>
<thead>
<tr>
<th>( k )</th>
<th>( m_k )</th>
<th>( s_k )</th>
<th>Basis for ( \mathcal{M}_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>( - )</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>( E_4 )</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>( E_6 )</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>0</td>
<td>( E_4^2 )</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>0</td>
<td>( E_4E_6 )</td>
</tr>
<tr>
<td>12</td>
<td>2</td>
<td>1</td>
<td>( E_6^2 ) ( \Delta )</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>( E_4^2E_6 )</td>
</tr>
</tbody>
</table>

All basis elements shown are polynomials in \( E_4 \) and \( E_6 \) (Recall that \( \Delta = g_3^2 - 27g_2^3 \)). This holds in general:
Theorem Every element of $\mathcal{M}_k$ is expressible as
\[ \sum_{4m+6n=k} a_{m,n}E_4^m E_6^n, \]
where the coefficients are complex numbers and the indices are nonnegative integers.

Proof For $k \leq 12$, this is clear from the table. If $f \in \mathcal{M}_k$ for some $k > 12$, subtract some form $f(i\infty)E_4^m E_6^n$ of the correct weight, giving a cusp form. Then $(f - f(i\infty)E_4^m E_6^n)/\Delta$ is a modular form of weight $k - 12$. The assertion follows by induction.

Theorem All monomials $E_4^m E_6^n$ of equal weight are algebraically independent.

Proof Assume they are not. Dividing the equation with one of the terms would give a polynomial equation in $E_4^m E_6^n$; thus this function would be constant. This is not the case, and the assertion follows.

Theorem The dimension $m_k$ is given by
\[ m_k = \begin{cases} \left\lfloor \frac{k}{12} \right\rfloor & \text{if } k \equiv 2 \pmod{12}, \\ \left\lfloor \frac{k}{12} \right\rfloor + 1 & \text{if } k \not\equiv 2 \pmod{12}. \end{cases} \]

Proof One way is to check the formula for $k \leq 12$, and then note that equality must continue to hold because the formulae give $m_{k+12} = m_k + 1$. Another way is to count the number of nonnegative integers satisfying $4m+6n = k$, which by our earlier theorems is equal to the dimension.

The fact that all the spaces $\mathcal{M}_k$ are finite-dimensional, implies that a set of more than $m_k$ forms is linearly dependent. This in turn gives relations between the Fourier coefficients of the forms. For example, $\mathcal{M}_8$ has dimension one, and both $E_8$ and $E_4^2$ belong to this space. Since the constant terms are equal, one has $E_8 = E_4^2$, and equating coefficients yields the relation:
\[ \sigma_7(n) = \sigma_3(n) + 120 \sum_{k=1}^{n-1} \sigma_3(k)\sigma_3(n-k), \quad (5.40) \]
which would be rather more difficult to discover and prove using more elementary methods.

5.3.1 Hecke operators

Hecke introduced a set of linear operators acting on modular forms. They can be thought of as periodizing the forms. In terms of lattices, the $n$th Hecke operator, acting on $\mathcal{M}_k$, is defined by (see [24], [39])
\[ T_k(n)f(\Lambda) = \sum_{[\Lambda:\Lambda']=n} f(\Lambda'), \quad (5.41) \]
where the sum is taken over all sublattices of index $n$. For example, when $n = 2$, there are three such sublattices. If the lattice $\Lambda$ is spanned by $\lambda_1$ and $\lambda_2$, these
sublattices are spanned by \{2\lambda_1, \lambda_2\}, \{\lambda_1, 2\lambda_2\} and \{\lambda_1 + \lambda_2, \lambda_1 - \lambda_2\}. The last sublattice is the one depicted in figure 3.1. It is readily seen that \(T_k(n)\) maps \(\mathcal{M}_k\) into itself for every \(n\) and \(k\). It is also clear that for \(m\) and \(n\) relatively prime, one has
\[
T_k(m)T_k(n) = T_k(mn),
\]
(5.42)
there being only one lattice \(\Lambda'\) satisfying \([\Lambda : \Lambda'] = m\) and \([\Lambda' : \Lambda''] = n\). Less clear, but true, is that for any \(m, n\):
\[
T_k(m)T_k(n) = \sum_{d|(m,n)} d^{k-1}T_k\left(\frac{mn}{d^2}\right).
\]
(5.43)
For functions \(f(\tau) = f(\Lambda_\tau)\), one defines
\[
T_k(n)f(\tau) = n^{k-1} \sum_{\substack{ad = n, a \geq 0 \\ 0 \leq b < d}} d^{-k}f\left(\frac{a\tau + b}{d}\right),
\]
(5.44)
which corresponds to the previous definition except for the factor \(n^{k-1}\). The action on the Fourier coefficients looks as follows:
\[
T_k(n)\left(\sum_{m=0}^{\infty} a(m)q^m\right) = \sum_{m=0}^{\infty} \left(\sum_{d|(m,n)} d^{k-1}a\left(\frac{mn}{d^2}\right)\right)q^m,
\]
(5.45)
where \((0,n)\) is interpreted as \(n\). Requiring
\[
a(m)a(n) = \sum_{d|(m,n)} d^{k-1}a\left(\frac{mn}{d^2}\right)
\]
(5.46)
amounts to asking that
\[
T_k(n)f = a(n)f,
\]
(5.47)
i.e. that \(f\) is an eigenform for the operator \(T_k(n)\). Clearly the Hecke operators take cusp forms to cusp forms, so all cusp forms of weight 12, 16, 18, 20, 22 and 26 are automatically eigenforms for all Hecke operators (these spaces being one-dimensional). In particular it holds for \(\Delta\), proving the equation
\[
\tau(m)\tau(n) = \sum_{d|(m,n)} d^{11}\tau\left(\frac{mn}{d^2}\right)
\]
(5.48)
satisfied by Ramanujan’s function. In general, multiplicativity implies
\[
\sum_{m=1}^{\infty} \frac{a(m)}{m^s} = \prod_p \left(1 + \sum_{n=1}^{\infty} a(p^n)p^{-ns}\right),
\]
(5.49)
and Hecke showed that
\[
a(p)a(p^n) = a(p^{n+1}) + p^{k-1}a(p^{n-1}),
\]
(5.50)
for primes \(p\), whence
\[
\sum_{n=1}^{\infty} \frac{a(n)}{n^s} = \prod_p \frac{1}{1 - a(n)p^{-s} + p^{k-1-2s}}.
\]
(5.51)
Thus the factors in the Euler product arise from \( f \) being an eigenform for Hecke operators; a form which is a simultaneous eigenform for all \( T(p) \) (which generate the algebra of Hecke operators) will have a full Euler product. The Hecke operators have been generalized by Atkin and Lehner: see [2]. These operators will be used in the final chapter. Finally, note that for forms \( f, g \) of weight \( k \), the measure

\[
f(z)g(z)y^k \frac{dx\,dy}{y^2}
\]

is invariant and hence passes to a well-defined measure on \( \mathcal{H}/\Gamma \). One defines the (Petersson) inner product on \( S_k \) by

\[
\langle f, g \rangle = \int\int_{\mathcal{H}/\Gamma} f(z)g(z)y^k \frac{dx\,dy}{y^2}.
\]

The integral converges, \( f \) and \( g \) being cusp forms. This inner product makes \( S_k \) a Hilbert space. The Hecke operators are Hermitian with respect to this inner product, and since they commute, they can be simultaneously diagonalized. Thus \( S_k \) has an orthonormal basis of eigenforms.

### 5.4 The \( L \)-series of a modular form

Analogously to the \( L \)-series of an elliptic curve (recall section 3.3.6), one defines the \( L \)-series of a modular form

\[
f(\tau) = \sum_{n=0}^{\infty} a_n q^n
\]

as

\[
L(f, s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.
\]

As in section 4.1, the mapping from \( f \) to \( L(f, s) \) need not be treated purely algebraically, as it is essentially the Mellin transform. See [43], chapter 4.

### 5.5 Applications

#### 5.5.1 Additive number theory

With hindsight, the applications of modular forms to problems in additive number theory arguably goes back as far as to Jacobi, who in 1834 determined the number of ways to express an arbitrary integer as a sum of squares (although naturally, the solution was not formulated in that way). Let \( r_k(n) \) denote the number of ways to express \( n \) as a sum of \( k \) squares of integers, where two representations are considered different if the order differs (i.e., \( r_k(n) \) is the number of \( k \)-tuples \( (x_1, x_2, \ldots, x_k) \in \mathbb{Z}^k \) such that \( \sum_{i=1}^{k} x_i^2 = n \)). For example, one has \( r_2(25) = 12 \). Diophantus had claimed that every natural number is expressible as a sum of four squares, i.e., that \( r_4(n) > 0 \) for all \( n \in \mathbb{N} \). This was proved by Lagrange in 1770. Jacobi succeeded in determining \( r_4(n) \) explicitly. The most
natural solution uses theta functions, so will only state the result. Note that
\[ \vartheta_3(0, q)^4 = \left( \sum_{n \in \mathbb{Z}} q^n \right)^4 = \sum_{n=0}^{\infty} r_4(n)q^n. \] (5.56)

One can show that
\[ \vartheta_3(0, q)^4 = 1 + 8 \sum_{n=1}^{\infty} \left( \sum_{\substack{d|n \atop 4 \nmid d}} d \right) q^n, \] (5.57)

so \( r_4(n) \) equals eight times the sum of the divisors of \( n \) which are not divisible by four. The partition function also has connections to modular forms. The partition function \( p(n) \) is defined as the number of ways to express \( n \) as a sum of positive integers (regardless of the order). As noted already by Euler, the generating function factors as
\[ \sum_{m=0}^{\infty} p(m)q^m = \prod_{n=1}^{\infty} \frac{1}{1 - q^n}. \] (5.58)

Recall from section 5.2.2 that Dedekind’s eta function is given by
\[ \eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n). \] (5.59)

and that
\[ \Delta = (2\pi)^{12} \eta^{24} \] (5.60)

where \( \Delta \) is the modular discriminant, which is a modular form. Expressing the product in (5.58) in terms of this form and using its transformation properties, Rademacher obtained an explicit formula for \( p(n) \); for this calculation, see [1], chapter 5. Modular forms also arise as weight enumerators for lattices and linear codes; for these topics, see [8], [12]. For other applications, such as the construction of Ramanujan graphs, see [34].

### 5.5.2 The modularity theorem

The modularity theorem has its roots in conjectures made by Taniyama and Shimura in 1955-1957; it has been referred to as the Taniyama-Shimura-Weil conjecture and several variations thereof. Work by Frey, Serre and Ribet in the 1980s showed that the conjecture implies Fermat’s “last theorem” (rather, conjecture), prompting Wiles to attack the problem. With help from Taylor, the semistable case was settled in 1995, proving Fermat’s conjecture. The full modularity conjecture was proved by Breuil, Conrad, Diamond and Taylor in 1999. It is now seen as part of the even more general Langlands program.

The modularity theorem asserts that all elliptic curves over the rational numbers are modular; that is, parametrized by the modular curve \( X_0(N) \) (for some \( N \), assumed to be minimal) via a rational map with integer coefficients:
\[ X_0(N) \to E(\mathbb{Q}). \] This implies that the \( L \)-series attached to \( E \) equals the \( L \)-series of a (new-)form of weight two of level \( N \).
Example The equation \( y^2 + y = x^3 - x^2 \) is parametrized by the modular form

\[
\eta(\tau)^2 \eta(11\tau)^2 = q \prod_{n=1}^{\infty} (1 - q^n)^2 (1 - q^{11n})^2 = \\
q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - \\
2q^{10} + q^{11} - 2q^{12} + 4q^{13} - q^{15} - 4q^{16} - \\
2q^{17} + 4q^{18} + 2q^{20} + 2q^{21} - 2q^{22} - \ldots 
\]

(5.61)

and, for example, the equation has \( 1 + 7 - (-2) = 10 \) solutions over the field with 7 elements, as predicted (counting a solution at infinity). As in section 3.3.6, we write \( N_p \) for the number of points on the curve over the finite field with \( p \) elements. The following table shows \( p, N_p \) and \( a_p \) for some small \( p \):

<table>
<thead>
<tr>
<th>( p )</th>
<th>2</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>11</th>
<th>13</th>
<th>17</th>
<th>19</th>
<th>23</th>
<th>29</th>
<th>97</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N_p )</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>9</td>
<td>19</td>
<td>19</td>
<td>24</td>
<td>29</td>
<td>104</td>
</tr>
<tr>
<td>( a_p )</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-2</td>
<td>1</td>
<td>4</td>
<td>-2</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-7</td>
</tr>
</tbody>
</table>

5.6 Generalizations

Several generalizations of modular forms exist; one can consider forms of several variables, for more general groups, \( p \)-adic forms etc. The forms studied in this chapter are called elliptic modular forms, because of their connection to the moduli space of elliptic curves. We mention also that the term modular is sometimes reserved for forms for the full modular group, \( \Gamma(1) \). Forms for other groups are then called automorphic.
Chapter 6

New converse theorems

6.1 Background and motivation

As we have seen in section 5.4, modular forms give rise to Dirichlet series, and it is natural to ask when a given Dirichlet series arises in this way. That is, starting with the series

\[ F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}, \]  

(6.1)

under what conditions is the function

\[ f(\tau) = \sum_{n=1}^{\infty} a_n e^{2\pi i n \tau}, \]  

(6.2)

a modular form for some Fuchsian group? Recall that \( f \) is said to be modular with respect to the group \( G \) if it holomorphic, remains bounded at the cusps of \( \mathbb{H}/G \) and satisfies

\[ f\left( a\tau + b \atop c\tau + d \right) = (c\tau + d)^k f(\tau) \]  

(6.3)

for every transformation \( \tau \mapsto \frac{a\tau + b}{c\tau + d} \) in \( G \), and for some fixed integer \( k \), called the weight of the form. The Mellin transform, defined in 4.1, is linear and sends forms to Dirichlet series; explicitly, the function

\[ e^{-2\pi n \tau} \]

is sent to

\[ (2\pi)^{-s} \Gamma(s) \frac{1}{n^s}. \]

The transformation behaviour of the form \( f \) under the group \( G \) translates to special properties of the function \( F \). A converse theorem gives sufficient conditions on \( F \) for \( f \) to be a modular form. This line of research can be said to have started with Hamburger’s proof [18] that the functional equation satisfied by Riemann’s zeta function, together with some growth conditions, does in fact determine the zeta function up to constant factors. The Riemann zeta function is the Mellin transform of a well-known theta function related to a cusp form.
of weight 4 for the principal congruence subgroup \( \Gamma(2) \). The conditions on the theta function expressing the proper transformation under \( \Gamma(2) \), namely
\[
\theta\left(-\frac{1}{i\tau}\right) = \sqrt{\tau}\theta(i\tau)
\]
\[
\theta(\tau + 2) = \theta(\tau)
\] (6.4)

imply the functional equation
\[
\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).
\] (6.5)

In [35], Selberg defines what is now known as the Selberg class of Dirichlet series: the functions in this class have Euler products and satisfy certain functional equations. This class includes the classical zeta and \( L \)-functions, as well as those from modular forms. Selberg conjectured that all these functions satisfy a generalized Riemann hypothesis, and they are all believed to emerge from modular forms. See [7] for more on this. These conjectures motivate the current work on converse theorems. They are also loosely related to the Langlands program. Recall from section 4.2 that the degree\(^1\) of the polynomial in the local factors of an Euler product is called the degree of the product, and, by extension, of the \( L \)-function. The modern work in this area, to be described in sections 6.4 - 6.6, can be seen as an attempt to extend known results concerning degree 1 to degree 2.

### 6.2 Hecke’s original converse theorem

In his article "Über die Bestimmung Dirichletscher Reihen durch ihre Functionalgleichungen" [20], published in Matematische Annalen in 1936, Hecke proves a converse theorem for Dirichlet series. First, we need some definitions. Let \( k \), \( N \) and \( \gamma \) be constants, let \( F \) be an analytic function and let
\[
G(s) = \left(\frac{N}{2\pi}\right)^s \Gamma(s)F(s).
\] (6.6)

Hecke asks the question when \( F \) is defined by a Dirichlet series
\[
F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},
\] (6.7)

satisfies a functional equation of the form
\[
G(s) = \gamma G(k - s)
\] (6.8)

and furthermore is such that \((s - k)F(s)\) is an entire function. In Hecke’s terminology, a function \( F \) satisfying these requirements is of type \( \{N, k, \gamma\} \). One wishes the determine for which \( \{N, k, \gamma\} \) such functions exist, and when they do, how many linearly independent \( F \) exist. By a Mellin transformation,

---

\(^1\)More precisely, the maximal degree of all polynomials, which may be different at different primes.
this translates to a question about the existence of functions automorphic under
the group generated by
\[ \tau \mapsto \tau + N \quad (6.9) \]
and
\[ \tau \mapsto -1/\tau. \quad (6.10) \]
The equation expressing the transformation behaviour under the transformation (6.10) is equivalent to the functional equation (after a Mellin transformation). Hecke shows that the group generated by (6.9) with \( N \) real, and (6.10), is discrete precisely when

1. \( N \geq 2 \), or
2. \( N = 2 \cos(\pi/q) \), where \( q \geq 3 \) is an integer.

These groups are called **Hecke groups**. \( N = 1 \) gives the modular group \( \Gamma \). Hecke goes on to determine the dimensions of the spaces of functions in question. It is infinite-dimensional precisely when \( N > 2 \). Hecke also recovers Hamburger’s theorem. Hecke groups are studied in their own right, see for example [3], [4]. Hecke’s result, as far as we are concerned, reads as follows (\( \Gamma(s) \) is the gamma function):**

**Hecke’s converse theorem** Suppose that the function \( F \), defined as in (6.7), is such that \( \Gamma(s)F(s) \) is an entire function of a complex variable which is bounded in vertical stripes, and that \( F \) satisfies a functional equation of the form (6.8). Then, if \( N = 1, 2, 3 \) or \( 4 \), \( F \) is the \( L \)-function of a modular form for the group \( \Gamma_0(N) \).

Here, \( \Gamma_0(1) \) is simply the full modular group \( \Gamma \). The proof will be described as part of the proof in section 6.4 below.

### 6.3 Weil’s converse theorem

In 1967, Weil published the article "Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen" [45]. As will be described in 6.4 below, Hecke’s theorem does not generalize to higher congruence subgroups without additional assumptions. Weil introduces the idea to study not only the ordinary Dirichlet series
\[ F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}, \quad (6.11) \]
but also so-called **twisted** Dirichlet series:
\[ F_{\chi}(s) = \sum_{n=1}^{\infty} \frac{a_n \chi(n)}{n^s}, \quad (6.12) \]
where \( \chi(n) \) is a multiplicative character (recall the definition from chapter 4). The auxiliary function (6.6) is modified accordingly:
\[ G_{\chi}(s) = \left( \frac{N}{2\pi} \right)^s \Gamma(s)F_{\chi}(s). \quad (6.13) \]
Thus $G(s)$ corresponds to the trivial character. Weil also needs Gauss sums of the form

$$g(\chi) = \sum_r \chi(r)e^{2\pi ir/m}, \quad (6.14)$$

where $m$ is a fixed integer and $r$ runs over the residue classes modulo $m$. Weil’s result is the following:

**Weil’s converse theorem** Suppose that $G_\chi(s)$ is entire, bounded in vertical stripes and satisfies a functional equation of the form

$$G_\chi(s) = \pm\chi(-N)\frac{g(\chi)}{g(\overline{\chi})}N^{k/2-s}G_\chi(k-s) \quad (6.15)$$

for every $m$ such that $(m, N) = 1$ and every character $\chi$ modulo $m$. Then $F(s)$ is a modular form of weight $k$ for the group $\Gamma_0(N)$, and if the Dirichlet series converges absolutely, it is a cusp form.

Actually, an additional character has been hidden in the $\pm$-sign. Weil ends the article by commenting on the Taniyama - Shimura conjecture (now theorem), which asserts that all elliptic curves over $\mathbb{Q}$ are parametrized by modular forms. By the time of writing, this was already known for some special cases. Weil remarks that the full conjecture appears “problematic at present”, and suggests it as an exercise for the interested reader.

### 6.4 An extension of Hecke’s converse theorem

#### 6.4.1 Introduction and results

In 1995, Brian Conrey and David Farmer published the article “An extension of Hecke’s converse theorem” [5]. In this article, Conrey and Farmer extend Hecke’s results to some congruence subgroups of higher level, namely $\Gamma_0(N)$ for $5 \leq N \leq 12$, $14 \leq N \leq 17$, and $N = 23$. In the case considered by Hecke, the corresponding spaces of functions have infinite dimension, so additional assumptions are required to get a converse theorem. Unlike Weil, Conrey and Farmer assume that the Dirichlet series has an Euler product

$$F(s) = \prod_p \frac{1}{L_p(p^{-s})} \quad (6.16)$$

with local factors of the form

$$L_p(p^{-s}) = \begin{cases} 1 - a_p p^{-s} + p^{k-1-2s}, & p \nmid N \\ 1 - p^{k/2-1-s}, & p|N \text{ but } p^2 \nmid N \\ 1, & p^2|N, \end{cases} \quad (6.17)$$

in addition to a functional equation and a growth assumption. Namely, put

$$F(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s} \quad (6.20)$$
6.4. An extension of Hecke’s converse theorem

and

\[ f(\tau) = \sum_{n=1}^{\infty} a_n e^{2\pi in\tau}, \quad (6.21) \]

and introduce the auxiliary function

\[ G(s) = \left( \frac{\sqrt{N}}{2\pi} \right)^s \Gamma(s) F(s). \quad (6.22) \]

The functional equation of level \( N \) and weight \( k \) takes the form

\[ G(s) = \pm (-1)^{k/2} G(k - s). \quad (6.23) \]

In the cases treated by Hecke, the growth condition (\( F \) in an entire function, bounded in vertical strips) and the functional equation alone force \( f \) to be automorphic under \( \Gamma_0(N) \). This no longer holds when \( N > 4 \), so additional information coming from the Euler product has to be used. Of course, it is still sufficient to check that \( f \) transforms correctly on a generating set for \( \Gamma_0(N) \).

### 6.4.2 The proof

First, note that the functional equation (6.23) is equivalent to

\[ f\left( -\frac{1}{N\tau} \right) = \pm N^{k/2} \tau^k f(\tau), \quad (6.24) \]

and the Euler product is equivalent to \( f \) being an eigenfunction of the Hecke operators \( T(p) \) when \( p \nmid N \), and the Atkin-Lehner operators \( U(p) \) when \( p|N \). It is convenient, at this point, to work with matrices and introduce the notation

\[ (f|k\gamma)(z) = (\det \gamma)^{k/2} (cz + d)^{-k} f\left( \frac{az + b}{cz + d} \right) \quad (6.25) \]

for a matrix \( \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) in \( GL(2, \mathbb{R}) \). If the matrix has determinant equal to 1, the factor \( (\det \gamma)^{k/2} \) is clearly redundant; it is included only to allow for writing, for example, the transformation\(^2\)

\[ F(N) : \tau \mapsto -\frac{1}{N\tau} \quad (6.27) \]

as the matrix

\[ \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \quad (6.28) \]

rather than

\[ \begin{pmatrix} 0 & -1/\sqrt{N} \\ \sqrt{N} & 0 \end{pmatrix} \quad (6.29) \]

as would otherwise have been necessary. Thus the “\( k \)”-operator can be used for any Möbius transformation with positive determinant. The group of such

\(^2\)It is actually the well-known Fricke involution, hence the notation; we will not deal with its other properties here.
transformations, denoted $\text{GL}^+ (2, \mathbb{R})$, is a normal subgroup in $\text{GL}(2, \mathbb{R})$, of index 2. The proper transformation behaviour under a Möbius transformation $\gamma$ is then expressed as $f|_k = \pm f$. Note also that the transformation (6.27) does not belong to $\Gamma_0(N)$, but writing $T$ for the translation $\tau \mapsto \tau + 1$, one has

$$F(N)(F(N)T)^{-1} : \tau \mapsto \frac{1}{N\tau + 1},$$

(6.30)

and this transformation does belong to $\Gamma_0(N)$. In fact, together with $T$, it generates $\Gamma_0(N)$ when $N \leq 4$. The “$|_k$”-operator is extended linearly to the group ring $\mathbb{C}[\text{GL}^+ (2, \mathbb{R})]:$

$$f| \sum_i c_i \gamma_i = \sum_i c_i f| \gamma_i,$$

(6.31)

where we suppress the weight, and the $c_i$ are complex constants. Clearly the set of Möbius transformations annihilating $f$ are a (right) ideal. It will be denoted $\Omega_f$. All congruences between Möbius transformations (or their matrices) in the following will be modulo this ideal. This is natural, as the potential form $f$ does not “see the difference” between congruent transformations.

In terms of matrices, the Hecke operators (recall section 5.3.1) become

$$T(p) = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} + \sum_{i=0}^{p-1} \begin{pmatrix} 1 & i \\ 0 & p \end{pmatrix},$$

(6.32)

and the Atkin-Lehner operators ([2], [5])

$$U(p) = \sum_{i=0}^{p-1} \begin{pmatrix} p & i \\ 0 & p \end{pmatrix}$$

(6.33)

respectively. The assumptions on $F$ translate to

$$f|_k T(p) = a_p f, \quad p \nmid N,$$

(6.34)

$$f|_k U(p) = f|_k \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}, \quad p|N \quad \text{but} \quad p^2 \nmid N$$

(6.35)

$$f|_k U(p) = 0, \quad p^2|N.$$  

(6.36)

The proof breaks down in several cases: first, for $N = 5, 7, 9$, one has generators

$$\Gamma_0(5) = \langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 5 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \rangle$$

(6.37)

$$\Gamma_0(7) = \langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 7 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 7 & 4 \end{pmatrix} \rangle$$

(6.38)

$$\Gamma_0(9) = \langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 9 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 9 & 5 \end{pmatrix} \rangle$$

(6.39)

and the first two in each case express 1-periodicity (cf (6.21)) and the functional equation. It remains to check the last one, which has the general form

$$M_2(N) = \begin{pmatrix} 2 & 1 \\ N & \frac{N+1}{2} \end{pmatrix}.$$  

(6.40)
Modulo $\Omega_f$, one has from (6.32)

$$T(2) = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \equiv a_2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$ (6.41)

and pre- and postmultiplication by $\begin{pmatrix} 0 \\ N \end{pmatrix}$ gives

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ N & 1 \end{pmatrix} \equiv a_2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$ (6.42)

whence

$$\begin{pmatrix} 2 \\ N \\ 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix},$$ (6.43)

and postmultiplication by $\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$ gives

$$\begin{pmatrix} 2 & 1 \\ N & N+1 \\ 2 \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$ (6.44)

which proves the theorem for $N = 5, 7, 9$. Next, for $N = 6$, one has generators

$$\Gamma_0(6) = \left\{ \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 6 & 1 \end{pmatrix}, \begin{pmatrix} 7 & -3 \\ 12 & -5 \end{pmatrix} \right\},$$ (6.45)

where again the first two are free and the third one has the form $A(6)^{-1} \begin{pmatrix} 1 & 0 \\ 6 & 1 \end{pmatrix} A(6)$,

with

$$A(N) = \begin{pmatrix} -2 & 1 \\ N & -N+2 \end{pmatrix} \equiv \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}.\quad (6.46)$$

The congruence in (6.46) follows from the Atkin-Lehner operator $U(2)$, which must be used since $2|6$. One has, from (6.33) and (6.35),

$$U(2) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \equiv \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},$$ (6.47)

and

$$\begin{pmatrix} 0 & -1 \\ 2N & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \equiv \pm \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}.\quad (6.48)$$

It follows that

$$\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 2N & 0 \end{pmatrix} \equiv \pm \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}.\quad (6.49)$$

Cancelling and premultiplying by $\begin{pmatrix} 0 \\ N \end{pmatrix}$ gives (6.46), proving the theorem for $N = 6$. The case $N = 10$ follows as well, as all matrices in a generating set can be constructed from matrices already examined, and the case $2^k|N$, finally,
gives a proof for $N = 8, 12$ and $16$. Namely, writing

\[ F(N) = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}, \]  
\[ W(N) = \begin{pmatrix} 1 & 0 \\ N & 1 \end{pmatrix}, \]  
\[ B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \]  
\[ T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \]  

one has

\[ \Gamma_0(10) = \langle T, W(10), (W(10)A(10))^2, F(10)(W(10)A(10))^2F(10), A(10)^{-1}W(10)^{-1}A(10)T^{-1} \rangle \]  
\[ \Gamma_0(8) = \langle T, W(8), (W(10)A(10))^2, B^{-1}W(8)B \rangle \]  
\[ \Gamma_0(12) = \langle T, W(12), BW(12)^{-1}B, F(12)B^{-1}W(12)B^{-1}F(12), BF(12)BW(12)^{-1}BF(12)B \rangle \]  
\[ \Gamma_0(16) = \langle T, W(16), BW(16)^{-1}B, (BF(16))^4, (B^{-1}F(16))^4 \rangle. \]  

Clearly

\[ B = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \equiv \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \]  

when $U(2) \equiv 0$, i.e. when $4|N$. This proves the theorem for $N = 8, 10, 12$ and $16$. The remaining cases, $N = 11, 14, 15, 17$ and $23$, require more work. For $N = 11$ and $17$, the authors use some general properties of Hecke operators, but the proof method is feasible only when a certain parameter $n$ satisfies $\varphi(n) = 2$, so only $n = 3, 4$ and $6$ can be used. This limits the applicability of the general method to the aforementioned cases $N = 11$ and $17$. The cases $N = 14, 15$ and $23$ are outright \textit{ad hoc}. Finally, it is comparatively easy to verify that $f$ must vanish at the cusps of $\Gamma_0(N)$ in all cases considered.

### 6.5 Converse theorems assuming a partial Euler product

#### 6.5.1 Introduction and results

The article [5] assumes that $F(s)$ has a full Euler product:

\[ F(s) = \prod_p \frac{1}{L_p(p^{-s})}, \]  

where $p$ runs over all primes. However, the proof for the cases $N = 5, 6, 7, 8, 9, 10, 12$ and $16$ focuses on the local factor at $p = 2$. It seems unlikely that any one prime plays a special role, and indeed the information extracted from this local

factor does not suffice to prove a converse theorem for all the groups $\Gamma_0(N)$. There are two obvious ways to proceed. One is to treat all factors symmetrically, hoping to obtain general results. Another is to relax the assumption of an Euler product at all primes. The 2005 article “Converse theorems assuming a partial Euler product” by David Farmer and Kevin Wilson [15] takes the second route. More precisely, the authors assume that the Dirichlet series factors as

$$F(s) = (1 - a_2 2^{-s} + 2^{k-1-2s})^{-1} \sum_{2 \nmid n} \frac{a_n}{n^s}.$$  (6.60)

This is called a partial Euler product. Recall from 5.3.1 the connection between Euler products and Hecke operators. Under certain reasonable (but unproved) assumptions, the authors show that this implies that $f$ is automorphic under the group

$$\Gamma_{0,2}(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a \equiv 2^j \pmod{N} \text{ for some integer } j \right\}.$$  (6.61)

This is a smaller group than $\Gamma_0(N)$, but still larger than (i.e., containing) $\Gamma_1(N)$ (recall the definitions from page 15). This seems to be the strongest result possible under these assumptions.

6.5.2 The proof

The article [15] considers not only ordinary modular forms but the more general forms with multiplicative (or Dirichlet) characters. Recall that a multiplicative character $\chi$ for a group $G$ is a homomorphism $\chi : G \to (\mathbb{C} \setminus \{0\}, \cdot)$ from $G$ to the multiplicative group of nonzero complex numbers; a Dirichlet character modulo $N$ is a Dirichlet character from the cyclic group $\mathbb{Z}/N\mathbb{Z}$, extended periodically to $\mathbb{Z}$. A modular form $f$ for $\Gamma_0(N)$ of weight $k$ and with character $\chi$ satisfies

$$f\left( \frac{a\tau + b}{c\tau + d} \right) = \chi(d) (c\tau + d)^k f(\tau)$$  (6.62)

for every Möbius transformation $\tau \mapsto \frac{a\tau + b}{c\tau + d}$ in $\Gamma_0(N)$. Ordinary modular forms correspond to the so-called trivial character $\chi \equiv 1$. For these more general forms, the Hecke operators become (cf. (6.32))

$$T(p) = \chi(p) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} + \sum_{i=0}^{p-1} \begin{pmatrix} 1 & i \\ 0 & p \end{pmatrix},$$  (6.63)

and the Atkin-Lehner operators are, as before,

$$U(p) = \sum_{i=0}^{p-1} \begin{pmatrix} p & i \\ 0 & p \end{pmatrix}.$$  (6.64)

The auxiliary function becomes

$$G(s) = \left( \frac{\sqrt{N}}{2\pi} \right)^s \Gamma(s + \frac{k-1}{2}) F(s),$$  (6.65)
a full Euler product would be
\[ F(s) = \prod_p \left(1 - a_p p^{-s} + \chi(p) p^{k-1-2s}\right)^{-1}, \] (6.66)
and (6.26) is modified to
\[ (f|k\gamma)(z) = \chi(a) (\det \gamma)^{k/2} (cz + d)^{-k} f\left(\frac{az + b}{cz + d}\right), \] (6.67)
after noting that \( N|c \) so that \( ad - bc = 1 \) becomes \( ad \equiv 1 \pmod{N} \) and so \( \chi(d^{-1}) = \chi(d) \). As before (cf. (6.23),(6.24)), the functional equation
\[ G(s) = \pm (-1)^{k/2} G(k-s) \] (6.68)
becomes, after an inverse Mellin transform, \( f|k \mathcal{F}_N = \pm f \). The partial product (6.60) becomes \( T(2)f = a_2 f \) with \( \chi(2) = 1 \). This is what necessitates the introduction of the group (6.61), as will be described shortly. Recall equation (6.44) (from the article [5]), which says that
\[ \left( \begin{array}{cc} 2 & 1 \\ N & N+1 \end{array} \right) \equiv \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \] (6.69)
in the group ring. Noting that
\[ \left( \begin{array}{cc} 2 & -1 \\ -N & N+1 \end{array} \right) = W(N)^{-1} \left( \begin{array}{cc} 2 & 1 \\ N & N+1 \end{array} \right) T^{-1}, \] (6.70)
\[ \left( \begin{array}{cc} 2 & 1 \\ -N & -N+1 \end{array} \right) = W(N)^{-1} \left( \begin{array}{cc} 2 & 1 \\ N & N+1 \end{array} \right), \] (6.71)
and
\[ \left( \begin{array}{cc} 2 & -1 \\ N & -N+1 \end{array} \right) = \left( \begin{array}{cc} 2 & 1 \\ N & N+1 \end{array} \right) T^{-1}, \] (6.72)
it follows that
\[ \left( \begin{array}{cc} 2 \alpha \\ \beta N \end{array} \right) \equiv \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \] (6.73)
whenever \( |\alpha| = |\beta| = 1 \). Next, one wishes to prove by induction that
\[ \left( \begin{array}{cc} 2^n \alpha \\ \beta N \end{array} \right) \equiv \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \] (6.74)
for every integer \( n \), where \( |\alpha|, |\beta| \leq 2^{n-1} \) and \( \ast \) is such that the determinant equals 1. To achieve this, the authors are forced to make a pairing assumption, namely that if \( A + B = C + D \), then either \( A \equiv C \) and \( B \equiv D \), or \( A \equiv D \) and \( B \equiv C \) for a certain set of matrices \( A, B, C \) and \( D \). Under this assumption, the induction works, showing that all matrices of the form (6.74) are congruent to the identity (recall that this is another way of expressing the appropriate transformation behaviour of \( f \)). Next, one wants to show that the matrices (6.74), together with the usual matrices \( T \) and \( W(N) \), generate the group (6.61).

To this end, the authors also have to assume that if \( (d, bN) = 1 \), there are integers \( n \) and \( k \) such that \( 2^n b \equiv 1 \pmod{d + kbN} \). Assuming this, take an element
\[ \gamma = \left( \begin{array}{cc} a & b \\ c & d \end{array} \right) \in \Gamma_{0,2}(N) \] (6.75)
where \( b \neq 1 \), and take a matrix \( \delta \) of the form (6.74), so that
\[
\delta W(N)^k \gamma = \left( \begin{array}{ccc}
* & 2^nb + \alpha(d + kbN) & \\
* & * & * \\
\end{array} \right).
\]
(6.76)

Now, \( \det \gamma = 1 \) and \( N|c \) forces \( d \) and \( bN \) to be relatively prime, so by assumption \( n \) and \( k \) can be chosen so that \( 2^nb + \alpha(d + kbN) = 1 \). Thus, in (6.75), one may take \( b = 1 \). Then \( \gamma \) takes the form
\[
\gamma = \left( \begin{array}{ccc}
a & & 1 \\
ad - 1 & d & \\
\end{array} \right),
\]
(6.77)
where \( a \) is congruent to \( 2^j \) modulo \( N \) for some integer \( j \), say \( a + kN = 2^j \). Then
\[
\gamma W(N)^k = \left( \begin{array}{ccc}
2^j & & 1 \\
ad - 1 + kdN & d & \\
\end{array} \right),
\]
(6.78)
so the matrices (6.74) and \( W(N) \) do indeed generate (6.61). The authors end the article by commenting on the pairing assumption made. The assumption is clearly not true for general matrices \( A, B, C \) and \( D \). The authors note, however, that similar pairing results hold in the context of Weil’s theorem. The other part of the proof, the existence of the integers \( n \) and \( k \) in (6.76), is not mentioned in the concluding remarks. It appears to be incorrect, as we describe next.

### 6.5.3 An error

The authors make an attempt to derive the existence of the integers \( n \) and \( k \) from Artin’s conjecture on primitive roots, which is unproved; they remark that the result needed for this theorem is much weaker, which is clearly the case. However, to the present author the deduction from Artin’s conjecture appears incorrect altogether. Firstly, they assume that one can get \( b \equiv 2^a \) (mod \( d + kbN \)) rather than \( 2^nb + \alpha(d + kbN) = 1 \), which seems to be what is required as one wants to have \( 2^nb + \alpha(d + kbN) = 1 \). We now need the definition of a primitive root:

**Definition** An integer \( a \) is called a *primitive root* modulo a prime \( p \) if the residue class of \( a \) modulo \( p \) is a generator of the multiplicative group \( ((\mathbb{Z}/p\mathbb{Z})^*, \cdot) \).

If \( a \) is a primitive root modulo \( p \), then, for every \( b \) not divisible by \( p \), there is an integer \( n \) such that \( a^n \equiv b \) (mod \( p \)). In particular, taking \( a = 2 \), there would be an integer \( n \) such that \( 2^n \equiv b \) (mod \( p \)). Taking instead \( b^{-1} \) (which works because \( p \) does not divide \( b \)), and assuming that \( p \) can be written \( p = d + kbN \), one can obtain \( 2^n b \equiv 1 \) (mod \( d + kbN \)), as needed. Apart from Artin’s conjecture, the authors seem to use Dirichlet’s theorem. We state both for convenience.

**Dirichlet’s theorem** If \( \alpha \) and \( \beta \) are relatively prime, positive integers, the arithmetic progression \( \alpha + n\beta, n \in \mathbb{N} \), contains infinitely many primes. In fact, for a fixed \( \beta \), the primes are equidistributed among the possible residue classes modulo \( \beta \), so that the primes of the form \( \alpha + n\beta \) have density \( 1/\varphi(\beta) \) among the primes, where \( \varphi \) is Euler’s totient function.
By assumption, \((d, bN) = 1\), so Dirichlet’s theorem applies, and there are infinitely many primes of the form \(d + kbN\).

**Artin’s conjecture, weak version** Any integer which is not a square and not equal to \(-1\) is a primitive root modulo infinitely many primes.

Artin did in fact conjecture more, namely that the set of primes for which a fixed \(a\) is a primitive root has positive density among the primes, and that when \(a \not\equiv 1 \pmod{4}\), this density equals

\[
\prod_p \left(1 - \frac{1}{p(p-1)}\right) \approx 0.37395581 \ldots,
\]

where \(p\) runs over the primes. In particular, 2 is a primitive root for approximately 37% of all primes. Now, the argument seems to be that one can find a \(k\) such that \(d + kbN\) is a prime for which 2 is a primitive root. Then the existence of the number \(n\) follows immediately. Of course, there seems to be no reason why the sequence \(d + kbN\) should not contain such a prime, or indeed a positive fraction of all primes. This is an unnecessarily strong result, because one does not need \(d + kbN\) to be prime or 2 to be a primitive root. But it still does not follow from Artin’s conjecture that it must happen: note that the density results allow the two sets of primes to avoid each other completely (one has \(\varphi(d + kbN) \geq 2 \Rightarrow 0.3739 \ldots + 1/\varphi(d + kbN) < 1\)). We remark that Artin’s conjecture has still not been proved for any single number. However, there is theoretical support for it: see [21]. For an introductory survey, see [29].

### 6.6 A converse theorem for \(\Gamma_0(13)\)

**6.6.1 Introduction and results**

Recall from 6.4 that the article [5] proved converse theorems for \(5 \leq N \leq 12\), \(14 \leq N \leq 17\), and \(N = 23\). Thus \(N = 13\) remained. This case was settled in 2006, in the article “A converse theorem for \(\Gamma_0(13)\)” by Conrey, Farmer, Odgers, and Snaith [6]. This article proves a converse theorem for \(\Gamma_0(13)\) by assuming, in addition to holomorphicity of \(f\) in \(\mathcal{H}\), convergence of the corresponding function \(F\) in some right half-plane, and a functional equation

\[
G(s) = \pm G(k - s), \quad (6.80)
\]

that \(F\) is expressible as

\[
F(s) = \left(1 - \frac{a_2}{2^s} + \frac{2^{k-1}}{2^s}\right)^{-1} \left(1 - \frac{a_3}{3^s} + \frac{3^{k-1}}{3^s}\right)^{-1} \sum_{(n,6)=1} \frac{a_n}{n^s}, \quad (6.81)
\]

that is, that \(F\) has a partial Euler product (at 2 and 3). \(\Gamma_0(13)\) is generated by four known matrices, the first three of which are easily dealt with using the same methods as in the previous articles. The fourth generator requires an analytic argument. No unproved assumptions are made in this article.
6.6. A converse theorem for $\Gamma_0(13)$

6.6.2 The proof

The group $\Gamma_0(13)$ is generated by the matrices

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$  \hspace{1cm} (6.82)

$$W(13) = \begin{pmatrix} 1 & 0 \\ 13 & 1 \end{pmatrix},$$  \hspace{1cm} (6.83)

$$\gamma_2 = \begin{pmatrix} 2 & -1 \\ 13 & -6 \end{pmatrix},$$  \hspace{1cm} (6.84)

$$\gamma_3 = \begin{pmatrix} 3 & -1 \\ 13 & -4 \end{pmatrix},$$  \hspace{1cm} (6.85)

and as before, periodicity and the functional equation imply the correct transformation behaviour under the first two. The proof for $\gamma_2$ uses the local factors at $p = 2$ and $p = 3$. Namely, (6.81) gives

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix} \equiv 2^{1-k/2}a_2 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$  \hspace{1cm} (6.86)

and

$$\begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \equiv 3^{1-k/2}a_3 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$  \hspace{1cm} (6.87)

and pre- and postmultiplying by

$$F(13) = \begin{pmatrix} 0 & -1 \\ 13 & 0 \end{pmatrix},$$  \hspace{1cm} (6.88)

one finds

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 \\ -13 & 1 \end{pmatrix} \equiv 2^{1-k/2}a_2 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$  \hspace{1cm} (6.89)

and

$$\begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ -13 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ -26 & 1 \end{pmatrix} \equiv 3^{1-k/2}a_3 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$  \hspace{1cm} (6.90)

Subtract (6.86) from (6.89) to obtain

$$\begin{pmatrix} 2 & 0 \\ -13 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix},$$  \hspace{1cm} (6.91)

and note that

$$\gamma_2 = W(13) \begin{pmatrix} 2 & 0 \\ -13 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}^{-1}.$$  \hspace{1cm} (6.92)

This shows that $\gamma_2 \equiv I$. Finally, the generator $\gamma_3$ requires an analytic treatment. Assuming $k > 0$, one can show (after deriving appropriate identities satisfied by $\gamma_3$) that if $f$ were not invariant under $|\gamma_3|$, the function $f(\tau)|(I - \gamma_3)$ would be invariant under a non-discrete subgroup of $\text{PSL}(2, \mathbb{R})$. This does not prove it to be identically zero, but the authors succeed in describing all pathologies which can occur and are able to conclude that $f(\tau)|(I - \gamma_3)$ must vanish, i.e. that $\gamma_3 \equiv I$. This proves the converse theorem for $\Gamma_0(13)$. 
Conclusions

This thesis has made a survey of some converse theorems for Dirichlet series, after introducing the necessary machinery for this. A converse theorem asserts that a Dirichlet series having certain properties must be the $L$-series of a modular form. The original work by Hecke [20] used only the functional equation satisfied by the function defined by the Dirichlet series. Weil’s contribution [45] was to introduce Dirichlet characters, thereby generalizing Hecke’s work. An alternative approach was pioneered by Conrey and Farmer, who assumed that the Dirichlet series has an Euler product [5]. This led to a renewed interest in the problem and some new results in 2005 [15] and 2006 [6]. However, there seems to be no obvious way to generalize the proofs to congruence groups of higher level. Additionally, the article [15] is forced to make some unproved assumptions, which are not believed to be strictly necessary. The proofs use the assumptions that a certain Dirichlet series satisfies some growth conditions, a functional equation and factors, at least partially, in an Euler product, to show the proper transformation behaviour of the potential modular form under every transformation in a generating set for the group in question. This obviously leads one to use as few generators as possible, which however makes it difficult to construct proofs that work for many groups simultaneously. Intuitively, minimizing the generating set destroys the inherent symmetry, and one is forced to make a detailed study of generating sets with few discernible regularities. It seems worthwhile, therefore, to explore more symmetric ways to generate the groups, and develop methods that work for arbitrary places, rather than localizing at specific primes.

The motivation for these problems are the conjectured properties of the Selberg class of Dirichlet series [7]. These, in turn, are believed to elucidate the Riemann hypothesis and its various generalizations. In a yet wider perspective, this is a part of the extensive Langlands program.
Bibliography


[18] H. Hamburger: Über die Riemannsche Funktionalgleichung der ζ-Funktion, Mathematische Zeitschrift 10, number 3-4, 240-254, 1921


[46] Wikipedia: pictures of a torus and a tessellation induced by the modular group, available as
and
respectively.
Appendix A

Bernoulli numbers

The Bernoulli numbers, used in section 5.2.1, have many applications in number theory. Unfortunately, several notational conventions exist. The one used here is the following: the Bernoulli numbers $B_n$ are defined by

$$\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n. \quad (A.1)$$

Thus, noting that

$$\frac{z}{e^z - 1} + \frac{z}{2} = \frac{z}{2} \coth \frac{z}{2} \quad (A.2)$$

is an even function, it follows that $B_1 = -\frac{1}{2}$, and that all other Bernoulli numbers with odd index are zero. One sometimes omits those numbers, so that all numbers are nonzero. Furthermore, as defined here, the Bernoulli numbers alternate in sign, which is sometimes hidden by using another definition. Next, we derive the formula (5.10) used in section 5.2.1. Note that

$$\pi z \cot \pi z = \pi iz \frac{e^{\pi iz} + e^{-\pi iz}}{e^{\pi iz} - e^{-\pi iz}} = \pi iz \left(1 + \frac{2}{e^{2\pi iz} - 1}\right). \quad (A.3)$$

Using (A.1), one finds

$$\pi z \cot \pi z = \sum_{n=0}^{\infty} \frac{B_{2n}}{(2n)!} (2\pi iz)^{2n}. \quad (A.4)$$

On the other hand, the product expression for the sine function [39],

$$\sin \pi z = \pi z \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{n^2}\right) \quad (A.5)$$

gives, after a logarithmic differentiation,

$$\pi z \cot \pi z = 1 - 2 \sum_{n=1}^{\infty} \zeta(2n) z^{2n}, \quad (A.6)$$

whence

$$\zeta(2n) = -\frac{(2\pi i)^{2n}}{2(2n)!} B_{2n}, \quad (A.7)$$

which is equation (5.10). Evidently, all \( B_n \) are rational numbers. The first few nonzero Bernoulli numbers are:

\[
\begin{aligned}
B_1 &= -\frac{1}{2}, \\
B_2 &= \frac{1}{6}, \\
B_4 &= -\frac{1}{30}, \\
B_6 &= \frac{1}{42}, \\
B_8 &= -\frac{1}{30}, \\
B_{10} &= \frac{5}{66}, \\
B_{12} &= -\frac{691}{2730}.
\end{aligned}
\]
Copyright

The publishers will keep this document online on the Internet - or its possible replacement - for a period of 25 years from the date of publication barring exceptional circumstances. The online availability of the document implies a permanent permission for anyone to read, to download, to print out single copies for your own use and to use it unchanged for any non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional on the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility. According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its WWW home page: http://www.ep.liu.se/

Upphovsrätt

Detta dokument hålls tillgängligt på Internet - eller dess framtidiga ersättare - under 25 år från publiceringsdatum under förutsättning att inga extraordinaärer omständigheter uppstår. Tillgång till dokumentet innebär tillstånd för var och en att läsa, ladda ner, skriva ut enstaka kopior för enskilt bruk och att använda det oförändrat för ickekommersiell forskning och för undervisning. Överföring av upphovsrätten vid en senare tidpunkt kan inte upphäva detta tillstånd. All annan användning av dokumentet kräver upphovsmannens medgivande. För att garantera äktheten, säkerheten och tillgängligheten finns det lösningar av teknisk och administrativ art. Upphovsmannens ideella rätt innefattar rätt att bli nämnd som upphovsman i den omfattning som god sed kräver vid användning av dokumentet på ovan beskrivna sätt samt skydd mot att dokumentet ändras eller presenteras i sådan form eller i sådant sammanhang som är kränkande för upphovsmannens litterära eller konstnärliga anseende eller egenart. För ytterligare information om Linköping University Electronic Press se förlagets hemsida http://www.ep.liu.se/

© 2009, Jonas Karlsson