Explicit computations with modular Galois representations

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. P. F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op maandag 15 december 2008
klokke 13.45 uur

door

Johannes Gerardus Bosman

geboren te Wageningen
in 1979

Samenstelling van de promotiecommissie:

Promotor:	prof. dr. S. J. Edixhoven (Universiteit Leiden)
Referent:	prof. dr. W. A. Stein (University of Washington)
Overige leden:	prof. dr. J.-M. Couveignes (Université de Toulouse 2)
	prof. dr. J. Klüners (Heinrich-Heine-Universität Düsseldorf)
	prof. dr. H. W. Lenstra, Jr. (Universiteit Leiden)
	prof. dr. P. Stevenhagen (Universiteit Leiden)
	prof. dr. J. Top (Rijksuniversiteit Groningen)
	prof. dr. S. M. Verduyn Lunel (Universiteit Leiden)
	prof. dr. G. Wiese (Universität Duisburg-Essen)

Explicit computations with modular Galois representations

THOMAS STIELTJES INSTITUTE
FOR MATHEMATICS

Johan Bosman, Leiden 2008

The research leading to this thesis was partly supported by NWO.

Contents

Preface **vii**

1 Preliminaries **1**
1.1 Modular forms . . . 1
1.1.1 Definitions . . . 1
1.1.2 Example: modular forms of level one . . . 4
1.1.3 Eisenstein series of arbitrary levels . . . 7
1.1.4 Diamond and Hecke operators . . . 10
1.1.5 Eigenforms . . . 14
1.1.6 Anti-holomorphic cusp forms . . . 16
1.1.7 Atkin-Lehner operators . . . 16
1.2 Modular curves . . . 17
1.2.1 Modular curves over $\mathbb{C}$ . . . 18
1.2.2 Modular curves as fine moduli spaces . . . 19
1.2.3 Moduli interpretation at the cusps . . . 21
1.2.4 Katz modular forms . . . 24
1.2.5 Diamond and Hecke operators . . . 27
1.3 Galois representations associated to newforms . . . 27
1.3.1 Basic definitions . . . 28
1.3.2 Galois representations . . . 29
1.3.3 ℓ-Adic representations associated to newforms . . . 30
1.3.4 Mod ℓ representations associated to newforms . . . 32
1.3.5 Examples . . . 34
1.4 Serre's conjecture . . . 35
1.4.1 Some local Galois theory . . . 35
1.4.2 The level . . . 38
1.4.3 The weight . . . 39
1.4.4 The conjecture . . . 41

2 Computations with modular forms **43**
2.1 Modular symbols . . . 43
2.1.1 Definitions . . . 43
2.1.2 Properties . . . 45
2.1.3 Hecke operators . . . 46

2.1.4 Manin symbols . . . 47
2.2 Basic numerical evaluations . . . 49
2.2.1 Period integrals: the direct method . . . 50
2.2.2 Period integrals: the twisted method . . . 51
2.2.3 Computation of q-expansions at various cusps . . . 52
2.2.4 Numerical evaluation of cusp forms . . . 55
2.2.5 Numerical evaluation of integrals of cusp forms . . . 56
2.3 Computation of modular Galois representations . . . 58
2.3.1 Computing representations for $\tau(p) \bmod \ell$ . . . 58
2.3.2 Computing $\tau(p) \bmod \ell$ from P_ℓ . . . 62
2.3.3 Explicit numerical computations . . . 62

3 A polynomial with Galois group $\mathrm{SL}_2(\mathbb{F}_{16})$ 69
3.1 Introduction . . . 69
3.1.1 Further remarks . . . 70
3.2 Computation of the polynomial . . . 71
3.3 Verification of the Galois group . . . 72
3.4 Does P indeed define $\overline{\rho}_f$? . . . 74
3.4.1 Verification of the level . . . 75
3.4.2 Verification of the weight . . . 76
3.4.3 Verification of the form f . . . 77
3.5 MAGMA code used for computations . . . 78

4 Some polynomials for level one forms 79
4.1 Introduction . . . 79
4.1.1 Notational conventions . . . 79
4.1.2 Statement of results . . . 80
4.2 Galois representations . . . 81
4.2.1 Liftings of projective representations . . . 81
4.2.2 Serre invariants and Serre's conjecture . . . 82
4.2.3 Weights and discriminants . . . 82
4.3 Proof of the theorem . . . 84
4.4 Proof of the corollary . . . 86
4.5 The table of polynomials . . . 87

Bibliography 89

Samenvatting 95

Curriculum vitae 101

Index 103

Preface

The area of modular forms is one of the many junctions in mathematics where several disciplines come together. Among these disciplines are complex analysis, number theory, algebraic geometry and representation theory, but certainly this list is far from complete. In fact, the phrase 'modular form' has no precise meaning since modular forms come in many types and shapes. In this thesis, we shall be working with classical modular forms of integral weight, which are known to be deeply linked with two-dimensional representations of the absolute Galois group of the field of rational numbers.

In the past decades an astonishing amount of research has been performed on the deep *theoretical* aspects of these modular Galois representations. The most well-known result that came out of this is the proof of Fermat's Last Theorem by Andrew Wiles. This theorem states that for any integer $n > 2$, the equation $x^n + y^n = z^n$ has no solutions in positive integers x, y and z. The fact that at first sight this theorem seems to have nothing to do with modular forms at all witnesses the depth as well as the broad applicability of the theory of modular Galois representations. Another big result has been achieved, namely a proof of Serre's conjecture by Chandrashekhar Khare, Jean-Pierre Wintenberger and Mark Kisin. Serre's conjecture states that every continuous two-dimensional odd irreducible residual representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ comes from a modular form. This can be seen as a vast generalisation of Wiles's result and in fact the proof also uses Wiles's ideas.

On the other hand, research on the *computational* aspects of modular Galois representations is still in its early childhood. At the moment of writing this thesis there is very little literature on this subject, though more and more people are starting to perform active research in this field. This thesis is part of a project, led by Bas Edixhoven, that focuses on the computations of Galois representations associated to modular forms. The project has a theoretical side, proving computability and giving solid runtime analyses, and an explicit side, performing actual computations. The main contributors to the theoretical part of the project are, at this moment of writing, Bas Edixhoven, Jean-Marc Couveignes, Robin de Jong and Franz Merkl. A preprint version of their work, which will eventually be published as a volume of the Annals of Mathematics Studies, is available [28]. As the title of this thesis already suggests, we will be dealing with the explicit side of the project. In the explicit calculations we will make some guesses and base ourselves on unproven heuristics. However, we will use Serre's conjecture to prove the correctness of our results afterwards.

The thesis consists of four chapters. In Chapter 1 we will recall the relevant parts of the theory of modular forms and Galois representations. It is aimed at a reader who hasn't studied this subject before but who wants to be able to read the rest of the thesis as well. Chapter 2 will be discussing computational aspects of this theory, with a focus on performing explicit computations. Chapter 3 consists of a published article that displays polynomials with Galois group $\mathrm{SL}_2(\mathbb{F}_{16})$, computed using the methods of Chapter 2. Explicit examples of such polynomials could not be computed by previous methods. Chapter 4 will appear in the final version of the manuscript [28]. In that chapter, we present some explicit results on mod ℓ representations for level one cusp forms. As an application, we improve a known result on Lehmer's non-vanishing conjecture for Ramanujan's tau function.

Notations and conventions

Throughout the thesis we will be using the following notational conventions. For each field k we fix an algebraic closure $\overline{k}$, keeping in mind that we can embed algebraic extensions of k into $\overline{k}$. Furthermore, for each prime number p, we regard $\overline{\mathbb{Q}}$ as a subfield of $\overline{\mathbb{Q}}_p$ and $\overline{\mathbb{F}}_p$ will be regarded as a fixed quotient of the integral closure of $\mathbb{Z}_p$ in $\overline{\mathbb{Q}}_p$. Furthermore, if λ is a prime of a local or global field, then $\mathbb{F}_\lambda$ will denote its residue field.

Chapter 1

Preliminaries

In this chapter we will set up some preliminaries that we will need in later chapters. No new material will be presented in this chapter and a reader who is familiar with modular forms can probably skip most of it without loss of understanding of the rest of this thesis. The main purpose of this chapter is to make a reader who is not familiar with modular forms or related subjects sufficiently comfortable with them. The presented material is well-known and the exposition will be far from complete. Proofs will usually be omitted. The main references for all of this chapter are [24] and the references therein, as well as [25]. In each section we will also give specific further references.

1.1 Modular forms

In this section we will briefly discuss what modular forms are. Apart from the main references given in the beginning, references for further reading include [54].

1.1.1 Definitions

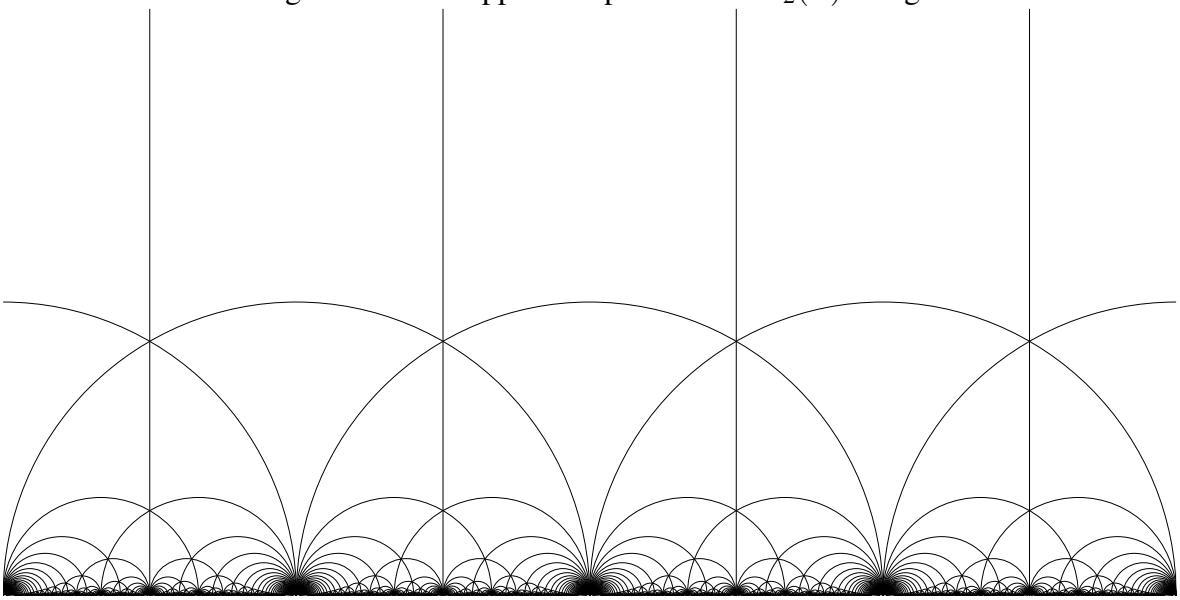
Consider the complex upper half plane $\mathfrak{H} := \{z \in \mathbb{C} : \Im z > 0\}$. On it we have an action of $\mathrm{SL}_2(\mathbb{Z})$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z := \frac{az+b}{cz+d}. \tag{1.1}$$

Note that this action is not faithful, but it does become faithful when factored through $\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z})/\pm I$. We can also add *cusps* to $\mathfrak{H}$. The cusps are the points in $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$. We will denote the completed upper half plane by $\mathfrak{H}^*$, so $\mathfrak{H}^* = \mathfrak{H} \cup \mathbb{P}^1(\mathbb{Q})$. We will extend the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathfrak{H}$ to an action on $\mathfrak{H}^*$: use the same fractional linear transformations.

It might be useful to note that $\mathrm{SL}_2(\mathbb{Z})$ acts transitively on the set of cusps: every cusp can be written as $\gamma\infty$ for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. The subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that fixes the cusp $\gamma\infty$ is the

Figure 1.1: The upper half plane with $\mathrm{SL}_2(\mathbb{Z})$-tiling



group

$$\gamma\left\{\pm\begin{pmatrix}1 & h\\ 0 & 1\end{pmatrix} : h\in\mathbb{Z}\right\}\gamma^{-1}.$$

Definition 1.1. Let $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ be a subgroup of finite index and consider a cusp $\gamma\infty$ with $\gamma\in\mathrm{SL}_2(\mathbb{Z})$. Then the *width* of $\gamma\infty$ with respect to Γ, or the width of $\gamma\infty$ in $\Gamma\setminus\mathfrak{H}^*$, is defined as the smallest positive integer h for which at least one of $\gamma\left(\begin{smallmatrix}1 & h\\ 0 & 1\end{smallmatrix}\right)\gamma^{-1}$ and $-\gamma\left(\begin{smallmatrix}1 & h\\ 0 & 1\end{smallmatrix}\right)\gamma^{-1}$ is in Γ.

Figure 1.1 is a useful picture to keep in mind when thinking about these things. It shows a tiling of the upper half plane along the $\mathrm{SL}_2(\mathbb{Z})$-action. Each tile here is an $\mathrm{SL}_2(\mathbb{Z})$-translate of the *fundamental domain*

$$\mathscr{F} := \left\{z\in\mathfrak{H} : -\frac{1}{2}\le\Re z\le\frac{1}{2} \text{ and } |z|\ge 1\right\}.$$

Sometimes in the literature parts of the boundary are left out in order that $\mathscr{F}$ contain exactly one point of each orbit of the $\mathrm{SL}_2(\mathbb{Z})$-action on $\mathfrak{H}$. We will not worry about sets of measure zero here; our definition enables us to view the topological space $\mathrm{SL}_2(\mathbb{Z})\setminus\mathfrak{H}$ as a quotient space of $\mathscr{F}$.

We can also use formula (1.1) to define an action of $\mathrm{GL}_2^+(\mathbb{R})$ on $\mathfrak{H}$ or of $\mathrm{GL}_2^+(\mathbb{Q})$ on $\mathfrak{H}^*$. Here the superscript $+$ means that we take the subgroup consisting of matrices with positive determinant.

We topologise $\mathfrak{H}^*$ in the following way: we take the usual topology on $\mathfrak{H}$ but a basis of open neighbourhoods for each cusp $\gamma\infty$ with $\gamma\in\mathrm{SL}_2(\mathbb{Z})$ consists of the sets

$$\{\gamma\infty\}\cup\gamma(\{z\in\mathfrak{H} : \Im z > M\}),$$

where M runs through $\mathbb{R}_{>0}$. With this topology, the set of cusps is discrete in $\mathfrak{H}^*$.

Definition 1.2. Let Γ be a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ of finite index and let k be an integer. A *modular form of weight k for* Γ is a holomorphic function $f : \mathfrak{H} \to \mathbb{C}$ satisfying the following conditions:

- $f(\frac{az+b}{cz+d}) = (cz+d)^k f(z)$ for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and all $z \in \mathfrak{H}$.
- f is holomorphic at the cusps. This means that for any matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, the function $(cz+d)^{-k} f(\frac{az+b}{cz+d})$ should be bounded in the region $\{z \in \mathbb{C} : \Im z \geq M\}$ for some (equivalently, any) $M > 0$.

The former condition is called the *modular transformation property* of f.

If $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ is of finite index, then the set of modular forms of weight k for the group Γ is denoted by $M_k(\Gamma)$. Under the usual addition and scalar multiplication of functions, $M_k(\Gamma)$ is a $\mathbb{C}$-vector space; it can in fact be shown to be of finite dimension.

We will often focus on the *cuspidal subspace* $S_k(\Gamma)$ of $M_k(\Gamma)$ that is defined as the set of $f \in M_k$ that vanish at the cusps. By "vanishing at the cusps" we mean that

$$\lim_{\Im z \to \infty} (cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right) = 0$$

should hold for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Elements of $S_k(\Gamma)$ are called *cusp forms*.

Now, let $N \in \mathbb{Z}_{>0}$ be given. Define the subgroup $\Gamma(N)$ of $\mathrm{SL}_2(\mathbb{Z})$ by

$$\Gamma(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod N \right\}.$$

Clearly, $\Gamma(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$ because it is the kernel of the reduction map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. A subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ that contains $\Gamma(N)$ for some N will be called a *congruence subgroup* of $\mathrm{SL}_2(\mathbb{Z})$. If Γ is a congruence subgroup then the smallest positive integer N for which $\Gamma \supset \Gamma(N)$ holds is called the *level* of Γ. Likewise, if f is a modular form for some congruence subgroup, we define its level to be the smallest positive integer N such that f is modular for the group $\Gamma(N)$.

Many special types of congruence subgroups of some level N turn out to be very interesting. Arguably, the two most interesting ones are

$$\Gamma_0(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \bmod N \right\}$$

and

$$\Gamma_1(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \bmod N \right\}.$$

One of the reasons to focus on these groups is that any modular form f of level N can be transformed into a modular form for $\Gamma_1(N^2)$ (and the same weight) by replacing it with $f(Nz)$. In fact we have an isomorphism

$$M_k(\Gamma(N)) \cong M_k\left(\Gamma_0(N^2) \cap \Gamma_1(N)\right) \subset M_k(\Gamma_1(N^2)) \tag{1.2}$$

defined by $f(z) \mapsto f(Nz)$.

Note that we have $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_1(N)$ for all N. If we plug this matrix into the transformation property of a modular form $f \in M_k(\Gamma_1(N))$, then $f(z+1) = f(z)$ follows. In other words, f is periodic with period 1. Hence f is a holomorphic function of

$$q = q(z) := e^{2\pi i z}.$$

We therefore have a power series expansion

$$f(z) = \sum_{n \geq 0} a_n(f) q^n,$$

the so-called *q-expansion* of f. The absence of terms with negative exponent is equivalent with f being holomorphic at ∞. If f is a cusp form, then it vanishes at ∞ and hence $a_0(f) = 0$. Be aware of the fact that $a_0 = 0$ does not in general imply that f is a cusp form because there are other cusps than ∞. The function from $\mathbb{Z}_{>0}$ to $\mathbb{C}$ defined by $n \mapsto a_n(f)$ has very interesting arithmetic properties for many modular forms f, as we shall see later.

1.1.2 Example: modular forms of level one

Let us give some examples of modular forms of level one now, that is modular forms for the full group $\mathrm{SL}_2(\mathbb{Z})$. Note that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. So to check the modular transformation properties in this case it suffices to check $f(z+1) = f(z)$ and $f(-1/z) = z^k f(z)$.

Another interesting thing to observe here is that $z \in \mathfrak{H}$ defines a lattice

$$\Lambda_z := \mathbb{Z}z + \mathbb{Z} \subset \mathbb{C}.$$

For $z, w \in \mathfrak{H}$ there is a $\lambda \in \mathbb{C}^\times$ with $\Lambda_z = \lambda \Lambda_w$ if and only if there is a $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $z = \gamma(w)$. On the other hand, given a lattice $\Lambda \subset \mathbb{C}$ we can choose a basis ω_1, ω_2 with $\Im(\omega_2/\omega_1) > 0$. Then we have $\Lambda = \omega_1 \Lambda_{\omega_2/\omega_1}$. This gives us a bijective correspondence between the $\mathrm{SL}_2(\mathbb{Z})$-equivalence classes of $\mathfrak{H}$ and the $\mathbb{C}^\times$-equivalence classes of the set of rank 2 lattices in $\mathbb{C}$.

We can use this to formulate the modular transformation property of a function $f : \mathfrak{H} \to \mathbb{C}$ in terms of lattices. Let $f : \mathfrak{H} \to \mathbb{C}$ be a function satisfying $f(\frac{az+b}{cz+d}) = (cz+d)^k f(z)$ for all

$\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and all $z \in \mathfrak{H}$. Then we define the function $F = F_f$ from the set of rank 2 lattices in $\mathbb{C}$ to $\mathbb{C}$ by

$$F(\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2) := \omega_1^{-k} f(\omega_2/\omega_1) \quad \text{where } \Im(\omega_2/\omega_1) > 0.$$

This function F then satisfies $F(\lambda\Lambda) = \lambda^{-k}F(\Lambda)$ for all $\lambda \in \mathbb{C}$ and all Λ. Conversely, given a function F from the set of rank 2 lattices in $\mathbb{C}$ to $\mathbb{C}$ that satisfies $F(\lambda\Lambda) = \lambda^{-k}F(\Lambda)$ for all $\lambda \in \mathbb{C}$ and all Λ, we define $f = f_F$ by

$$f(z) = F(\mathbb{Z}z + \mathbb{Z}).$$

The function f will then satisfy the weight k modular transformation property for $\mathrm{SL}_2(\mathbb{Z})$ and in fact the assignments $f \mapsto F_f$ and $F \mapsto f_F$ are inverse to each other.

Eisenstein series

Now that we have given definitions of modular forms, it becomes time that we write down some explicit examples. Let us first note that there are no non-zero modular forms of odd weight and level one; this can be seen by plugging in the matrix $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, which yields the identity $f(z) = (-1)^k f(z)$. So if we want to write down a modular form we should at least do this in even weight. For reasons that we will make clear later, there cannot exist nonzero modular forms of negative weight and no non-constant modular forms of weight 0. Also, in level one there are no non-zero modular forms of weight 2.

If $k \geq 4$ is even, then

$$G_k(z) := \frac{(k-1)!}{2(2\pi i)^k} \sum_{m,n \in \mathbb{Z}}{}' \frac{1}{(mz+n)^k} \tag{1.3}$$

is a modular form of weight k, the so-called *normalised Eisenstein series* of weight k and level one (priming the summation sign here means that we ignore the terms whose denominator is equal to zero). One can in fact write down $G_k(z)$ in terms of lattices. The formula becomes then

$$G_k(\Lambda) = \frac{(k-1)!}{2(2\pi i)^k} \sum_{z \in \Lambda}{}' z^{-k}$$

and we readily see that it does satisfy the weight k modular transformation property for $\mathrm{SL}_2(\mathbb{Z})$. The reason for using the normalisation factor $(k-1)!/(2(2\pi i)^k)$ becomes clear if one writes down the q-expansion for G_k:

$$G_k = -\frac{B_k}{2k} + \sum_{n \geq 1} \sigma_{k-1}(n) q^n. \tag{1.4}$$

Here B_k is the k-th Bernoulli number, defined by

$$\frac{x}{e^x - 1} = \sum_{k \geq 0} \frac{B_k}{k!} x^k.$$

and $\sigma_{k-1}(n)$ is defined as $\sum_{d|n} d^{k-1}$.

We see that the arithmetic function $n \mapsto \sigma_{k-1}(n)$ arises as the coefficients of a modular form, something that not everyone would expect right after reading the definition of a modular form.

Why can't we take $k = 2$ here? This is because the series (1.3) does not converge absolutely in that case and verifying the modular transformation property boils down to changing the order of summation. If we define G_2 by the q-expansion (1.4), then we get a well-defined holomorphic function on $\mathfrak{H}$ that 'almost' satisfies a modular transformation property for $\mathrm{SL}_2(\mathbb{Z})$: we have

$$G_2\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 G_2(z) - \frac{c(cz+d)}{4\pi i}$$

for all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. The 'almost' modularity of G_2 is still very useful within the theory of modular forms.

Discriminant modular form

The spaces $M_k(\mathrm{SL}_2(\mathbb{Z}))$ for $k \in \{4,6,8,10\}$ can be shown to be one-dimensional, so they are generated by G_k. In particular there are no non-zero cusp forms there. The lowest weight where we do have a cusp form of level one is $k = 12$ (for higher levels, however, there are non-zero cusp forms of lower weight):

$$\Delta(z) := 8000G_4^3 - 147G_6^2 = q\prod_{n\geq 1}(1-q^n)^{24}.$$

This form is called the *discriminant modular form* or *modular discriminant* and it is a generator for the space $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$. If we write it out as a series

$$\Delta(z) = \sum_{n\geq 1}\tau(n)q^n = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 + \cdots$$

then $\tau(n)$ is called the *Ramanujan tau function*. The tau function will play an important role in this thesis. Ramanujan observed some very remarkable properties of it. Among these properties, the following ones occur, which he was unable to prove.

- For coprime integers m and n we have $\tau(mn) = \tau(m)\tau(n)$.
- For prime powers we have a recurrence $\tau(p^{r+1}) = \tau(p)\tau(p^r) - p^{11}\tau(p^{r-1})$.
- For all prime numbers p we have the estimation $|\tau(p)| \leq 2p^{11/2}$.

The first two of these properties were proved by Mordell in 1917; they determine $\tau(n)$ in terms of $\tau(p)$ for p prime. The third property was proved by Deligne in 1974; its proof uses very deep results from algebraic geometry. These properties witness once more the interesting arithmetic behaviour of q-coefficients of modular forms.

Other properties found by Ramanujan and improved by others (cf. [83, Section 1] and [64, Section 4.5]) are congruence properties. For $\ell \in \{2,3,5,7,23,691\}$ there exist simple formulas for $\tau(n)$ modulo ℓ or a power of ℓ. The following summarises what is known about this for $\ell \neq 23$:

$$\begin{array}{ll}
\tau(n) \equiv \sigma_{11}(n) \mod 2^{11} & \text{for } n \equiv 1 \mod 8, \\
\tau(n) \equiv 1217\sigma_{11}(n) \mod 2^{13} & \text{for } n \equiv 3 \mod 8, \\
\tau(n) \equiv 1537\sigma_{11}(n) \mod 2^{12} & \text{for } n \equiv 5 \mod 8, \\
\tau(n) \equiv 705\sigma_{11}(n) \mod 2^{14} & \text{for } n \equiv 7 \mod 8, \\
\tau(n) \equiv n^{-610}\sigma_{1231}(n) \mod 3^6 & \text{for } n \equiv 1 \mod 3, \\
\tau(n) \equiv n^{-610}\sigma_{1231}(n) \mod 3^7 & \text{for } n \equiv 2 \mod 3, \\
\tau(n) \equiv n^{-30}\sigma_{71}(n) \mod 5^3 & \text{for } n \not\equiv 0 \mod 5, \\
\tau(n) \equiv n\sigma_9(n) \mod 7 & \text{for } n \equiv 0,1,2,4 \mod 7, \\
\tau(n) \equiv n\sigma_9(n) \mod 7^2 & \text{for } n \equiv 3,5,6 \mod 7, \\
\tau(n) \equiv \sigma_{11}(n) \mod 691 & \text{for all } n.
\end{array}$$

Modulo 23 we have the following congruences for $p \neq 23$ prime:

$$\begin{array}{ll}
\tau(p) \equiv 0 \mod 23 & \text{if } \left(\frac{p}{23}\right) = -1, \\
\tau(p) \equiv \sigma_{11}(p) \mod 23^2 & \text{if } p \text{ is of the form } a^2 + 23b^2, \\
\tau(p) \equiv -1 \mod 23 & \text{otherwise.}
\end{array}$$

Later in this thesis we will study $\tau(p) \bmod \ell$ for other values of ℓ.

1.1.3 Eisenstein series of arbitrary levels

Having seen some examples in level one, we now turn back to the subgroups $\Gamma_0(N)$ and $\Gamma_1(N)$ of $\mathrm{SL}_2(\mathbb{Z})$. In this subsection we will define what Eisenstein series are for these subgroups. The situation is analogous to the level one case, though slightly more complicated. We will make use of Dirichlet characters, which will in this subsection be assumed to be primitive and take values in $\mathbb{C}^\times$. If a Dirichlet character is evaluated at an integer not coprime with its conductor, then the value is defined to be 0. Details for this subsection can be found in [25, Chapter 4].

The case $k \geq 3$

For $N \in \mathbb{Z}_{>0}$, $k \in \mathbb{Z}_{\geq 3}$ and $\overline{c}, \overline{d} \in \mathbb{Z}/N\mathbb{Z}$ we define

$$G_k^{(\overline{c},\overline{d})}(z) := \sum_{\substack{m \equiv c \bmod N \\ n \equiv d \bmod N}}{}' \frac{1}{(mz+n)^k}. \tag{1.5}$$

This defines a modular form of weight k for $\Gamma(N)$.

To get forms with nice q-expansions, we have to take suitable linear combinations of the forms $G_k^{(\overline{c},\overline{d})}$. Choose two Dirichlet characters ψ and ϕ, of conductors $N(\psi)$ and $N(\phi)$ say,

that satisfy the conditions

$$N(\psi)N(\phi) \mid N \quad \text{and} \quad \psi(-1)\phi(-1) = (-1)^k. \tag{1.6}$$

We then define

$$G_k^{\psi,\phi} := \frac{(-N(\phi))^k (k-1)!}{2(2\pi i)^k g(\phi^{-1})} \sum_{c=1}^{N(\psi)} \sum_{d=1}^{N(\phi)} \sum_{e=1}^{N(\psi)} G_k^{(\overline{cN(\psi)}, \overline{d+eN(\psi)})},$$

where the pair $(\overline{cN(\psi)}, \overline{d+eN(\psi)})$ is an element of $(\mathbb{Z}/(N(\psi)N(\phi)\mathbb{Z}))^2$ and for any $\mathbb{C}$-valued Dirichlet character χ, the number $g(\chi)$ denotes its Gauss sum:

$$g(\chi) := \sum_{v \in (\mathbb{Z}/N(\chi)\mathbb{Z})^\times} \chi(v) \exp\left(\frac{2\pi i v}{N(\chi)}\right). \tag{1.7}$$

The q-expansion of $G_k^{\psi,\phi}$ is as follows:

$$G_k^{\psi,\phi} = -\frac{\delta(\psi) B_{k,\phi}}{2k} + \sum_{n \geq 1} \sigma_{k-1}^{\psi,\phi}(n) q^n, \tag{1.8}$$

where $\delta(\psi)$ equals 1 if ψ is trivial and 0 otherwise, $B_{k,\phi}$ is a so-called generalised Bernoulli number defined by

$$\sum_{v \in (\mathbb{Z}/N(\phi)\mathbb{Z})^\times} \phi(n) \frac{x e^{vx}}{e^{N(\phi)x} - 1} = \sum_{k \geq 0} \frac{B_{k,\phi}}{k!} x^k$$

and $\sigma_{k-1}^{\psi,\phi}(n)$ is a character-twisted sum of $(k-1)$-st powers of divisors, defined as

$$\sigma_{k-1}^{\psi,\phi}(n) = \sum_{d \mid n} \psi(n/d)\phi(d) d^{k-1}.$$

The function $G_k^{\psi,\phi}$ is called a *normalised Eisenstein series with characters ψ and ϕ*. It is an element of $M_k(\Gamma_1(N(\psi)N(\phi)))$. In particular, it is an element of $M_k(\Gamma_1(N))$ and the same holds for $G_k^{\psi,\phi}(dz)$ for every $d \mid \frac{N}{N(\psi)N(\phi)}$. Furthermore, $G_k^{\psi,\phi}$ is in $M_k(\Gamma_0(N))$ if and only if the character $\psi\phi$ is trivial.

The cases $k = 1$ and $k = 2$

Recall from the level one situation that G_2, defined by a q-series, is not a modular form, though it is not really far from being one. A similar picture occurs in arbitrary level: the series (1.5) do not converge absolutely for $k \in \{1,2\}$, but the q-series (1.8) do define holomorphic functions on $\mathfrak{H}$ that are 'almost' modular. In fact it will turn out to be much nicer than it seems to be at first sight. Assume $k \in \{1,2\}$, take $N \in \mathbb{Z}_{>0}$ and let ψ and ϕ be $\mathbb{C}^\times$-valued Dirichlet characters that satisfy (1.6).

Let us first treat the case $k = 2$. Define $G_2^{\psi,\phi}$ by the q-series (1.8). Then $G_2^{\psi,\phi}$ is in $M_2(\Gamma_1(N))$ unless both ψ and ϕ are trivial, in which case $G_2^{\psi,\phi}(z) - dG_2^{\psi,\phi}(dz) = G_2(z) - dG_2(dz)$ is in $M_2(\Gamma_1(N))$ for all $d \mid N$. Again, the series is modular for $\Gamma_0(N)$ if and only if $\psi\phi$ is trivial.

In weight 1 the convergence problems of (1.5) are even worse but still we can do almost the same thing. We alter the definition of the q-series slightly: put

$$G_1^{\psi,\phi} := -\frac{\delta(\phi)B_{1,\psi} + \delta(\psi)B_{1,\phi}}{2} + \sum_{n\geq 1} \sigma_0^{\psi,\phi}(n)q^n.$$

This turns out to be a modular form in $M_1(\Gamma_1(N))$ in all cases.

Eisenstein subspace

Now that we have defined for each space $M_k(\Gamma_1(N))$ what its Eisenstein series are, we will define its *Eisenstein subspace* as the subspace generated by these series:

Definition 1.3. Let k and N be positive integers with $k \neq 2$. The Eisenstein subspace $E_k(\Gamma_1(N))$ of $M_k(\Gamma_1(N))$ is defined as the subspace generated by the modular forms $G_k^{\psi,\phi}(dz)$ defined above where (ψ,ϕ) runs through the set of pairs of Dirichlet characters satisfying (1.6) and for given (ψ,ϕ), the number d runs through all divisors of $N/(N(\psi)N(\phi))$.

Definition 1.4. Let N be a positive integer. The Eisenstein subspace $E_2(\Gamma_1(N))$ of $M_2(\Gamma_1(N))$ is defined as the subspace generated by the following modular forms:

- The forms $G_k^{\psi,\phi}(dz)$ defined above where (ψ,ϕ) runs through the set of pairs of Dirichlet characters that are not both trivial and that satisfy (1.6) and for given (ψ,ϕ), the number d runs through all divisors of $N/(N(\psi)N(\phi))$.
- The forms $G_2(z) - dG_2(dz)$ where d runs through divisors of N, except $d = 1$.

The given generators for the spaces actually do give a basis for each space, provided that in the case $k = 1$ we take each form $G_1^{\psi,\phi} = G_1^{\phi,\psi}$ only once. Furthermore, we define $E_k(\Gamma_0(N))$ to be $M_k(\Gamma_0(N)) \cap E_k(\Gamma_1(N))$ and this is actually generated by the Eisenstein series that lie in $M_k(\Gamma_0(N))$.

The Eisenstein subspace satisfies a very nice property:

Theorem 1.1. *Let k and N be positive integers and let Γ be either $\Gamma_0(N)$ or $\Gamma_1(N)$. Then every $f \in M_k(\Gamma)$ can be written in a unique way as $g + h$ with $g \in E_k(\Gamma)$ and $h \in S_k(\Gamma)$.*

In particular, Eisenstein series are not cusp forms and knowing a complete description of Eisenstein series reduces the study of modular forms to that of cusp forms. The q-expansions of cusp forms are in general far less explicit but far more interesting than those of Eisenstein series.

1.1.4 Diamond and Hecke operators

The arithmetic structure of modular forms turns out to be related to interesting operators on the spaces $S_k(\Gamma_1(N))$, called *diamond operators* and *Hecke operators*. The operators are in fact defined on all of $\mathbb{M}_k(\Gamma_1(N))$, preserving $E_k(\Gamma_1(N))$ as well. However, the treatments for S_k and E_k differ at a few points and since we more or less 'know' E_k already, we will stick to $S_k(\Gamma_1(N))$ from now. Details for this subsection can be found in [25, Chapter 5].

Most operators on modular forms can be formulated in terms of a notation called the *slash operator*. For $k \in \mathbb{Z}$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{R})$ we define the following operation on the space of functions $f : \mathfrak{H} \to \mathbb{C}$:

$$(f|_k\gamma)(z) := \det(\gamma)^{k-1}(cz+d)^{-k} f(\gamma z).$$

It must be noted that in the literature there appears to be no consensus about the normalisation factor $\det(\gamma)^{k-1}$; some textbooks use $\det(\gamma)^{k/2}$ instead. For a function f the modular transformation property of weight k for $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ can be formulated in terms of the slash operator as $f|_k\gamma = f$ for all $\gamma \in \Gamma$. Be aware of the fact that slash operators in general don't leave the spaces $S_k(\Gamma)$ invariant.

Diamond operators

Note that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ and that for the quotient we have

$$\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times \quad \text{by } \overline{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \mapsto \overline{d}. \tag{1.9}$$

It follows from this normality that $\gamma \in \Gamma_0(N)$ leaves the spaces $S_k(\Gamma_1(N))$ invariant under the weight k slash action. Since the action of the subgroup $\Gamma_1(N)$ is trivial so this defines an action of $(\mathbb{Z}/N\mathbb{Z})^\times$ on $S_k(\Gamma_1(N))$:

$$\langle d\rangle f := f|_k \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where we can choose any matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$ mapping to $\overline{d}$ under (1.9). The operator $\langle d\rangle$ is called a *diamond operator*.

Let $\varepsilon : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ be a character. Then we define the subspace $S_k(N,\varepsilon)$ of $S_k(\Gamma_1(N))$ as

$$S_k(N,\varepsilon) := \left\{ f \in S_k(\Gamma_1(N)) : \langle d\rangle f = \varepsilon(d) f \quad \text{for all } d \in (\mathbb{Z}/N\mathbb{Z})^\times \right\}$$

and call it the ε-eigenspace of $S_k(\Gamma_1(N))$. Note that if ε is the trivial character, then we have $S_k(N,\varepsilon) = S_k(\Gamma_0(N))$. If $f \in S_k(\Gamma_1(N))$ lies inside $S_k(N,\varepsilon)$ then we say that f is a *modular form with character* ε. Now, the diamond action of $(\mathbb{Z}/N\mathbb{Z})^\times$ on $S_k(\Gamma_1(N))$ is a

representation of $(\mathbb{Z}/N\mathbb{Z})^\times$ on a finite-dimensional $\mathbb{C}$-vector space and thus is a direct sum of irreducible representations, hence we have

$$S_k(\Gamma_1(N)) = \bigoplus_{\varepsilon:(\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times} S_k(N,\varepsilon).$$

Note that we always have $\langle -1\rangle = (-1)^k$ so that $S_k(N,\varepsilon)$ can only be non-zero for ε with $\varepsilon(-1) = (-1)^k$.

Hecke operators

Congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$ have the property that any two of them are commensurable, which means that their intersection has finite index in both of them. Also, for any congruence subgroup $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ and any $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ we have that $\gamma^{-1}\Gamma\gamma \cap \mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ and that $\gamma^{-1}\Gamma\gamma$ is commensurable with Γ. It follows that for any two congruence subgroups Γ_1 and Γ_2 and any $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$ the left action of Γ_1 on $\Gamma_1\gamma\Gamma_2$ has only a finite number of orbits. If we choose representatives $\gamma_1, \ldots, \gamma_r \in \mathrm{GL}_2^+(\mathbb{Q})$ for these orbits then the operator

$$T_\gamma = T_{\Gamma_1,\Gamma_2,k,\gamma} : S_k(\Gamma_1) \to S_k(\Gamma_2)$$

given by

$$T_\gamma f = \sum_{i=1}^{r} f|_k \gamma_i \tag{1.10}$$

is well-defined and depends only on the double coset $\Gamma_1\gamma\Gamma_2$. Note that the diamond operator $\langle d\rangle$ is equal to T_γ if we choose $\gamma \in \Gamma_0(N)$ with lower right entry congruent to d mod N.

Now, let p be a prime number and consider the operator T_p on $S_k(\Gamma_1(N))$ defined as

$$T_p := T_\gamma \quad \text{for } \gamma = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}.$$

It is this operator that we call a *Hecke operator*. If we write it out according to the definition of T_γ then we have

$$T_p f = (\langle p\rangle f)\big|_k \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} + \sum_{j=0}^{p-1} f\big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix}, \tag{1.11}$$

where we take the convention $\langle p\rangle f = 0$ for $p \mid N$. It can be shown that the Hecke operators on $S_k(\Gamma_1(N))$ commute with the diamond operators and with each other. In particular the subspaces $S_k(N,\varepsilon)$ are preserved; hence we can speak of T_p as operators on $S_k(N,\varepsilon)$, with $S_k(\Gamma_0(N))$ being a special case of this. The formula (1.11) then becomes

$$T_p f = \varepsilon(p) f\big|_k \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} + \sum_{j=0}^{p-1} f\big|_k \begin{pmatrix} 1 & j \\ 0 & p \end{pmatrix},$$

for $f \in S_k(N,\varepsilon)$.

If we use the lattice interpretation for the level one case, we can formulate T_p in terms of lattices. Take $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ and let F be the corresponding function on the set of full rank lattices in $\mathbb{C}$. Then the function corresponding to $T_p f$ is equal to

$$T_p F(\Lambda) = p^{k-1} \sum_{\substack{\Lambda' \subset \Lambda \\ [\Lambda:\Lambda']=p}} F(\Lambda'), \tag{1.12}$$

i.e. we sum over all sublattices of index p. A similar interpretation exists in arbitrary levels; we shall address this later, in Subsection 1.2.5.

We can also define operators T_n for arbitrary positive integers n. We do this by means of a recursion formula:

$$\begin{array}{ll} T_1 = 1, & \\ T_{mn} = T_m T_n & \text{for } m,n \text{ coprime}, \\ T_{p^r} = T_p^r & \text{for } p \mid N \text{ prime and } r \in \mathbb{Z}_{>1}, \\ T_{p^{r+1}} = T_p T_{p^r} - \langle p \rangle p^{k-1} T_{p^{r-1}} & \text{for } p \nmid N \text{ prime and } r \in \mathbb{Z}_{>0}. \end{array} \tag{1.13}$$

One motivation for this definition is that in the lattice interpretation formula (1.12) we can simply replace p with n.

We can in fact describe the Hecke operators in terms of q-expansions. Take $N \in \mathbb{Z}_{>0}$ and $f \in S_k(\Gamma_1(N))$. For all $n \in \mathbb{Z}_{>0}$ we have

$$a_m(T_n f) = \sum_{\substack{d \mid \gcd(m,n) \\ \gcd(d,N)=1}} d^{k-1} a_{mn/d^2}(\langle d \rangle f).$$

This formula has some interesting special cases. First of all, for $m = 1$ we get

$$a_1(T_n f) = a_n(f). \tag{1.14}$$

Also, for p prime and $f \in S_k(N, \varepsilon)$ we have

$$a_n(T_p f) = \begin{cases} a_{pn}(f) & \text{for } p \nmid n, \\ a_{pn}(f) + \varepsilon(p) p^{k-1} a_{n/p}(f) & \text{for } p \mid n. \end{cases}$$

Petersson inner product

Let $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ be of finite index. We can define an inner product (i.e. a positive definite hermitian form) on $S_k(\Gamma)$ that is very natural in some sense. If we write $z = x + iy$ then the measure $\mu := dxdy/y^2$ is $\mathrm{GL}_2^+(\mathbb{R})$-invariant on $\mathfrak{H}$ and the integral $\int_{\Gamma \backslash \mathfrak{H}} \mu$ converges to $[\mathrm{PSL}_2(\mathbb{Z}) : \mathrm{P}\Gamma]\pi/3$. The measure μ is called the *hyperbolic measure* on $\mathfrak{H}$. Also, for $f \in S_k(\Gamma)$ the function $|f(z)|^2 y^k$ is Γ-invariant and bounded on $\mathfrak{H}$, hence the measure

$$\mu_f := |f(z)|^2 y^{k-2} dxdy \quad \text{where } z = x + iy$$

is a Γ-invariant measure on $\mathfrak{H}$ such that the integral $\int_{\Gamma\backslash\mathfrak{H}} \mu_f$ converges to a positive real number. Now we define the Petersson inner product on $S_k(\Gamma)$ as follows:

$$(f,g) := \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \mathrm{P}\Gamma]} \int_{\Gamma\backslash\mathfrak{H}} f(z)\overline{g(z)}y^{k-2}dxdy \tag{1.15}$$

for $f,g \in S_k(\Gamma)$, i.e. it is a scaled inner product associated to the Hermitian form $f \mapsto \int_{\Gamma\backslash\mathfrak{H}} \mu_f$. The normalisation factor $[\mathrm{PSL}_2(\mathbb{Z}) : \mathrm{P}\Gamma]^{-1}$ is used so that the value of the integral does not depend on the chosen group Γ for which f and g are modular.

We can in fact use the formula (1.15) for the Petersson inner product to define a sesquilinear pairing on $M_k(\Gamma) \times S_k(\Gamma)$ (note that this would not work on $M_k(\Gamma) \times M_k(\Gamma)$ as the integral diverges there). For $\Gamma \in \{\Gamma_0(N), \Gamma_1(N)\}$ the set of $f \in M_k(\Gamma)$ with $(f,g) = 0$ for all $g \in S_k(\Gamma)$ is exactly the Eisenstein subspace $E_k(\Gamma)$ defined in Subsection 1.1.3.

From now on, we return to the case $\Gamma = \Gamma_1(N)$. The Petersson inner product behaves particularly nicely with respect to the Hecke operators. Take $\gamma \in \mathrm{GL}_2^+(\mathbb{Q})$. Then the adjoint of T_γ with respect to the Petersson inner product is equal to T_{γ^*} where

$$\gamma^* = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

i.e.

$$(T_\gamma f, g) = (f, T_\gamma^* g) \quad \text{where } T_\gamma^* = T_{\gamma^*}.$$

For the diamond operators this boils down to

$$\langle d \rangle^* = \langle d \rangle^{-1}$$

If we now let W_N be the operator $f \mapsto N^{1-k/2} f|_k \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$ on $S_k(\Gamma_1(N))$ then we have

$$T_n^* = W_N T_n W_N^{-1}. \tag{1.16}$$

We will study the operator W_N in more detail in Subsection 1.1.7. In the special case $\gcd(n,N) = 1$ formula (1.16) simplifies to

$$T_n^* = \langle n \rangle^{-1} T_n \quad \text{if } \gcd(n,N) = 1.$$

In particular for n coprime to N the operators T_n and T_n^* commute.

Hecke algebra

The diamond and Hecke operators on $S_k(\Gamma_1(N))$ generate a subring of $\mathrm{End}_{\mathbb{C}} S_k(\Gamma_1(N))$ which we call the *Hecke algebra* of $S_k(\Gamma_1(N))$ and which is commutative. We will usually denote the Hecke algebra by $\mathbb{T}$, where it is understood which modular forms space is involved. We will also be considering its subalgebra $\mathbb{T}'$ that is generated by all the $\langle d \rangle$ and

T_n with $\gcd(n,N) = 1$. If confusion could arise we will write $\mathbb{T}_k(N)$ and $\mathbb{T}'_k(N)$ respectively.

The structure of $\mathbb{T}$ is important in the study of $S_k(\Gamma_1(N))$. It can be shown that $\mathbb{T}$ is a free $\mathbb{Z}$-module of rank $\dim S_k(\Gamma_1(N))$. Consider the pairing

$$\mathbb{T} \times S_k(\Gamma_1(N)) \to \mathbb{C}, \quad (T, f) \mapsto a_1(Tf).$$

For any ring A we put $\mathbb{T}_A := \mathbb{T} \otimes A$. From formula (1.14) it follows immediately that the induced pairing $\mathbb{T}_{\mathbb{C}} \times S_k(\Gamma_1(N)) \to \mathbb{C}$ is perfect. In particular we have

$$S_k(\Gamma_1(N)) \cong \mathrm{Hom}_{\mathbb{Z}-\mathrm{Mod}}(\mathbb{T}, \mathbb{C}) \tag{1.17}$$

Under this isomorphism, the action of $\mathbb{T}$ on $S_k(\Gamma_1(N))$ comes from the following action of $\mathbb{T}$ on $\mathrm{Hom}_{\mathbb{Z}-\mathrm{Mod}}(\mathbb{T}, \mathbb{Z})$: let $T \in \mathbb{T}$ send $\phi \in \mathrm{Hom}_{\mathbb{Z}-\mathrm{Mod}}(\mathbb{T}, \mathbb{Z})$ to $T' \mapsto \phi(TT')$. It can be shown that $\mathrm{Hom}(\mathbb{T}_{\mathbb{Q}}, \mathbb{Q})$ is in this way a free $\mathbb{T}_{\mathbb{Q}}$-module of rank one so that in fact $S_k(\Gamma_1(N))$ is free of rank one as a $\mathbb{T}_{\mathbb{C}}$-module. For each subring A of $\mathbb{C}$, we can identify $\mathrm{Hom}_{\mathbb{Z}-\mathrm{Mod}}(\mathbb{T}, A)$ with the A-module of modular forms whose q-expansion has coefficients in A.

1.1.5 Eigenforms

The commutativity of all the T_n, T_n^*, $\langle d \rangle$ and $\langle d \rangle^*$ for n and d coprime to N has an interesting consequence:

Theorem 1.2. *For $k, N \in \mathbb{Z}_{>0}$ the space $S_k(\Gamma_1(N))$ has a basis that is orthogonal with respect to the Petersson inner product and whose elements are eigenvectors for all the operators in $\mathbb{T}'$.*

Theorem 1.2 would fail if we took all the Hecke operators in $\mathbb{T}$, i.e. also the T_n with $\gcd(n,N) > 1$. This is because those operators are in general not semi-simple, so we do not get a decomposition of our vector space into eigenspaces. Forms that are eigenvectors for all the operators in $\mathbb{T}$ are called *eigenforms*. If a form is an eigenvector for all the operators in $\mathbb{T}'$, we will call it a $\mathbb{T}'$-eigenform. Each $\mathbb{T}'$-eigenform is an eigenvector for the diamond operators, so must lie inside some space $S_k(N, \varepsilon)$. An eigenform f is called *normalised* if $a_1(f) = 1$. From (1.14) and the commutativity of $\mathbb{T}$ it follows easily that $f \in S_k(\Gamma_1(N))$ is a normalised eigenform if and only if the map $\mathbb{T} \to \mathbb{C}$ corresponding to f as in (1.17) is a ring homomorphism.

Consider M and N with $M \mid N$. For each divisor d of N/M we have a map

$$\alpha_d : S_k(\Gamma_1(M)) \to S_k(\Gamma_1(N)) \quad \text{defined by } f(z) \mapsto f(dz).$$

The map α_d is called a *degeneracy map*. Note that for $d = 1$ it is just the inclusion of $S_k(\Gamma_1(M))$ into $S_k(\Gamma_1(N))$. The subspace of $S_k(\Gamma_1(N))$ generated by all the $\alpha_d(f)$ for $M \mid N$, $M < N$, $d \mid N/M$ is called the *old subspace* of $S_k(\Gamma_1(N))$ and is denoted by $S_k(\Gamma_1(N))^{\mathrm{old}}$.

The orthogonal complement of $S_k(\Gamma_1(N))^{\mathrm{old}}$ with respect to the Petersson inner product is called the *new subspace* and denoted by $S_k(\Gamma_1(N))^{\mathrm{new}}$. Its eigenforms have interesting properties:

Theorem 1.3. *Let $f \in S_k(\Gamma_1(N))^{\text{new}}$ be an eigenform. Then $\mathbb{C} \cdot f$ is an eigenspace of $S_k(\Gamma_1(N))$ and $a_1(f) \neq 0$. Furthermore, $S_k(\Gamma_1(N))^{\text{new}}$ is generated by its eigenforms.*

This is called the *multiplicity one theorem*. In fact, in the new subspace there is no distinction between eigenforms for $\mathbb{T}$ and eigenforms for $\mathbb{T}'$. The theorem allows us to put the normalisation $a_1 = 1$ on eigenforms in the new subspace. New eigenforms f that satisfy $a_1(f) = 1$ are called *newforms*. If we combine this with (1.14) then we see

Theorem 1.4. *Let N and k be positive integers and let $f \in S_k(\Gamma_1(N))$ be a newform. Then the eigenvalue of the Hecke operator T_n on f is equal to the q-coefficient $a_n(f)$.*

If $f \in S_k(\Gamma_1(M))$ a $\mathbb{T}'_k(M)$-eigenform, then for all d the form $\alpha_d(f) \in S_k(\Gamma_1(dM))$ is a $\mathbb{T}'_k(dM)$-eigenform. We furthermore have a decomposition:

$$S_k(\Gamma_1(N)) = \bigoplus_{M|N} \bigoplus_{d|\frac{N}{M}} \alpha_d \left(S_k(\Gamma_1(M))^{\text{new}}\right)$$

that allows us to write down an interesting basis for $S_k(\Gamma_1(N))$:

Theorem 1.5. *Let N and k be given positive integers. Then the following set is a basis for $S_k(\Gamma_1(N))$ consisting of $\mathbb{T}'$-eigenforms.*

$$\bigcup_{M|N} \bigcup_{d|\frac{N}{M}} \{\alpha_d(f) : f \text{ is a newform in } S_k(\Gamma_1(M))\}.$$

The field K_f

If $f \in S_k(\Gamma_1(N))$ is a newform with character ε, then the values of ε together with the coefficients $a_n(f)$ generate a field

$$K_f := \mathbb{Q}(\varepsilon, a_1(f), a_2(f), \ldots)$$

which is known to be a number field. It can be shown that for any embedding $\sigma : K_f \hookrightarrow \mathbb{C}$ the function $\sigma f := \sum \sigma(a_n) q^n$ is a newform in $S_k(\Gamma_1(N))$ with character $\sigma\varepsilon$. To a newform $f \in S_k(N, \varepsilon)$ we can attach a ring homomorphism

$$\theta_f : \mathbb{T} \to K_f$$

defined by

$$\theta_f(\langle d \rangle) = \varepsilon(d) \quad \text{and} \quad \theta_f(T_p) = a_p,$$

as in (1.17). We define

$$I_f := \ker(\theta_f),$$

which is a prime ideal of $\mathbb{T}$ called the *Hecke ideal* of f. It is known that $\operatorname{im} \theta_f$ is an order in K_f but it need not be the maximal order.

1.1.6 Anti-holomorphic cusp forms

From time to time we will also be considering *anti-holomorphic cusp forms*. A function $f : \mathfrak{H} \to \mathbb{C}$ is called an anti-holomorphic cusp form of some level N and weight k if $z \mapsto \overline{f(z)}$ is in $S_k(\Gamma_1(N))$. The space of anti-holomorphic cusp forms of level N and weight k is denoted by $\overline{S}_k(\Gamma_1(N))$. We let the diamond and Hecke operators act on $\overline{S}_k(\Gamma_1(N))$ by the formulas

$$\langle d\rangle \overline{f} = \overline{\langle d\rangle f} \quad \text{and} \quad T_p \overline{f} = \overline{T_p f},$$

where we denote by $\overline{f}$ the function $z \mapsto \overline{f(z)}$. The spaces $\overline{S}_k(N,\varepsilon)$ are now defined as

$$\begin{aligned}\overline{S}_k(N,\varepsilon) &= \left\{\overline{f} : f \in S_k(N,\overline{\varepsilon})\right\} \\ &= \left\{f \in \overline{S}_k(\Gamma_1(N)) : \langle d\rangle f = \varepsilon(d) f \text{ for all } d \in (\mathbb{Z}/N\mathbb{Z})^\times\right\}.\end{aligned}$$

If we have a simultaneous eigenspace inside $S_k(\Gamma_1(N))$ for the diamond and Hecke operators then we also have an eigenspace with conjugate eigenvalues and of the same dimension (which could be the same space if all these eigenvalues are real). It follows that we have a decomposition of $S_k(\Gamma_1(N)) \oplus \overline{S}_k(\Gamma_1(N))$ into eigenspaces with the same eigenvalues as in the decomposition of $S_k(\Gamma_1(N))$, but the dimension of each such eigenspace is twice the dimension of its restriction to $S_k(\Gamma_1(N))$.

1.1.7 Atkin-Lehner operators

The main reference for this subsection is [3].

Besides diamond and Hecke operators, there is another interesting type of operators on $S_k(\Gamma_1(N))$, namely the *Atkin-Lehner operators*. Let Q be a positive divisor of N such that $\gcd(Q, N/Q) = 1$. Let $w_Q \in \mathrm{GL}_2^+(\mathbb{Q})$ be any matrix of the form

$$w_Q = \begin{pmatrix} Qa & b \\ Nc & Qd \end{pmatrix} \tag{1.18}$$

with $a,b,c,d \in \mathbb{Z}$ and $\det(w_Q) = Q$. The assumption $\gcd(Q,N/Q) = 1$ ensures that such a w_Q exists. A straightforward verification shows $f|_k w_Q \in S_k(\Gamma_1(N)))$. Now, given Q, this $f|_k w_Q$ still depends on the choice of a,b,c,d. However, we can use a normalisation in our choice of a,b,c,d which will ensure that $f|_k w_Q$ only depends on Q. Be aware of the fact that different authors use different normalisations here. The one we will be using is

$$a \equiv 1 \bmod N/Q, \quad b \equiv 1 \bmod Q, \tag{1.19}$$

which is the normalisation used in [3]. We define

$$W_Q(f) := Q^{1-k/2} f|_k w_Q = \frac{Q^{k/2}}{(Ncz+Qd)^k} f\left(\frac{Qaz+b}{Ncz+Qd}\right), \tag{1.20}$$

which is now independent of the choice of w_Q and call W_Q an *Atkin-Lehner operator*.

An unfortunate thing about these Atkin-Lehner operators is that they do not preserve the spaces $S_k(N,\varepsilon)$. But we can say something about it. Let $\varepsilon : (\mathbb{Z}/NZ)^\times \to \mathbb{C}^\times$ be a character and suppose that $f \in S_k(N,\varepsilon)$. By the Chinese Remainder Theorem, one can write ε in a unique way as $\varepsilon = \varepsilon_Q \varepsilon_{N/Q}$ such that ε_Q is a character on $(\mathbb{Z}/Q\mathbb{Z})^\times$ and $\varepsilon_{N/Q}$ is a character on $(\mathbb{Z}/(N/Q)\mathbb{Z})^\times$. It is a fact that

$$W_Q(f) \in S_k(N, \overline{\varepsilon}_Q \varepsilon_{N/Q}).$$

Also, there is a relation between the q-expansions of f and $W_Q(f)$:

Theorem 1.6. *Let $f \in S_k(N,\varepsilon)$ be a newform. Take $Q \mid N$ with $\gcd(Q, N/Q) = 1$. Then*

$$W_Q(f) = \lambda_Q(f) g$$

with $\lambda_Q(f) \in \mathbb{C}$ an algebraic number of absolute value 1 and $g \in S_k(N, \overline{\varepsilon}_Q \varepsilon_{N/Q})$ a newform. Suppose now that n is a positive integer and write $n = n_1 n_2$ where n_1 consists only of prime factors dividing Q and n_2 consists only of prime factors not dividing Q. Then we have

$$a_n(g) = \varepsilon_{N/Q}(n_1)\overline{\varepsilon}_Q(n_2)\overline{a_{n_1}(f)} a_{n_2}(f).$$

The number $\lambda_Q(f)$ in the above theorem is called a *pseudo-eigenvalue* for the Atkin-Lehner operator. In some cases there exists a closed expression for it.

Theorem 1.7. *Let $f \in S_k(N,\varepsilon)$ be a newform and suppose q is a prime that divides N exactly once. Then we have*

$$\lambda_q(f) = \begin{cases} g(\varepsilon_q) q^{-k/2} \overline{a_q(f)} & \text{if } \varepsilon_q \text{ is non-trivial,} \\ -q^{1-k/2} \overline{a_q(f)} & \text{if } \varepsilon_q \text{ is trivial.} \end{cases}$$

Here, $g(\varepsilon_q)$ is the Gauss sum of ε_q.

Theorem 1.8 ([2, Theorem 2])**.** *Let $f \in S_k(N,\varepsilon)$ be a newform with N square-free. For $Q \mid N$ we have*

$$\lambda_Q(f) = \varepsilon(Qd - \frac{N}{Q}a) \prod_{q \mid Q} \varepsilon(Q/q)\lambda_q(f).$$

Here, a and d are defined by (1.18). Moreover, this identity holds without any normalisation assumptions on the entries of w_Q, as long as we define $\lambda_q(f)$ by the formula given in Theorem 1.7.

1.2 Modular curves

In this section we will very briefly discuss modular curves. Apart from the main references given in the beginning, we use [22] and [36] as further references on this subject. We will use a little bit of algebro-geometric language. but we'll keep it as simple as possible, trying to explain properties that we need to understand why the calculations in later chapters work.

1.2.1 Modular curves over $\mathbb{C}$

Let $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ be a subgroup of finite index. If one divides out the group action of Γ on $\mathfrak{H}$ one obtains a Riemann surface

$$Y_\Gamma := \Gamma \setminus \mathfrak{H}.$$

If we add the cusps to Y_Γ and use $(q|_0\gamma^{-1})^{1/w(\gamma\infty)}$ as a local parameter at the cusp $\gamma\infty$ we obtain another Riemann surface

$$X_\Gamma := \Gamma \setminus \mathfrak{H}^*,$$

which happens to be compact. This compactness implies that X_Γ is in fact (the analytification of) a projective algebraic curve over $\mathbb{C}$, the open subset $Y_\Gamma \subset X_\Gamma$ being an affine curve.

For Γ equal to $\Gamma_0(N)$, $\Gamma_1(N)$ or $\Gamma(N)$ we write Y_Γ as $Y_0(N)$, $Y_1(N)$ or $Y(N)$ and X_Γ as $X_0(N)$, $X_1(N)$ or $X(N)$ respectively. These are the curves in which we are primarily interested.

The curves $Y_0(N)$, $Y_1(N)$ and $Y(N)$ have moduli interpretations. Take $z \in \mathfrak{H}$ and consider the lattice $\Lambda_z = \mathbb{Z}z + \mathbb{Z}$, as we did in Subsection 1.1.2. Then $\mathbb{C}/\Lambda_z$ is a complex elliptic curve and in this way $\mathrm{SL}_2(\mathbb{Z}) \setminus \mathfrak{H}$ is in bijection with the set of all isomorphism classes of elliptic curves over $\mathbb{C}$. This gives in all three cases the moduli interpretation for $N = 1$. In general, $Y_0(N)(\mathbb{C}) = \Gamma_0(N) \setminus \mathfrak{H}$ is in bijection with the set of isomorphism classes of pairs (E, C) where E is an elliptic curve over $\mathbb{C}$ and $C \subset E(\mathbb{C})$ is a cyclic subgroup of order N. The bijection is obtained by

$$z \mapsto (\mathbb{C}/\Lambda_z, \frac{1}{N}\mathbb{Z} \bmod \Lambda_z).$$

The additional information C that we attach to E is called a *level structure*.

Likewise, for $Y_1(N)(\mathbb{C}) = \Gamma_1(N) \setminus \mathfrak{H}$ the map

$$z \mapsto (\mathbb{C}/\Lambda_z, \frac{1}{N} \bmod \Lambda_z).$$

defines a bijection with the set of isomorphism classes of pairs (E, P) with E an elliptic curve over $\mathbb{C}$ and $P \in E(\mathbb{C})$ a point of order N.

To describe the moduli interpretation of $Y(N)$, we use the Weil pairing on elliptic curves over $\mathbb{C}$. The sign convention we use is such that the Weil e_N-pairing on the N-torsion of $\mathbb{C}/\Lambda$ is defined as

$$e_N(z, w) = \exp\left(\pi i N \frac{\bar{z}w - z\overline{w}}{\mathrm{covol}(\Lambda)}\right).$$

Then the map

$$z \mapsto (\mathbb{C}/\Lambda_z, \frac{1}{N} \bmod \Lambda_z, \frac{z}{N} \bmod \Lambda_z)$$

defines a bijection between $Y(N)(\mathbb{C}) = \Gamma(N) \setminus \mathfrak{H}$ and the set of isomorphism classes of triples (E, P, Q) where E is an elliptic curve over $\mathbb{C}$ and $P, Q \in E(\mathbb{C})[N]$ are points that satisfy $e_n(P, Q) = \exp(2\pi i/N)$.

In view of (1.2), the curve $Y(N)$ is isomorphic to Y_Γ with $\Gamma = \Gamma_0(N^2) \cap \Gamma_1(N)$. The map $z \mapsto Nz$ defines an isomorphism $Y_\Gamma \to Y(N)$. In terms of moduli, Y_Γ parametrises triples (E, C, P) with $E/\mathbb{C}$ an elliptic curve, $C \subset E(\mathbb{C})$ cyclic of order N^2 and $P \in C$ a point of order N. Let us describe what the given isomorphism $Y_\Gamma \to Y(N)$ sends (E, C, P) to. Choose a generator P' for C with $P = NP'$ and a $Q \in E(\mathbb{C})[N^2]$ with $e_{N^2}(P', Q) = \exp(2\pi i/N^2)$. Then the image of (E, C, P) is the triple $(E/\langle NP \rangle, P \bmod NP, NQ \bmod NP)$.

1.2.2 Modular curves as fine moduli spaces

In the previous subsection we spoke about bijections between points of $Y_\Gamma(\mathbb{C})$ and isomorphism classes of elliptic curves with certain level structures. It turns out that this can be put in a more general setting, which is what we will do in the present subsection.

For an arbitrary scheme S, an elliptic curve over S is defined to be a proper smooth group scheme E over S of which all the geometric fibres are elliptic curves. For a fixed positive integer N that we use for our level structures, we will usually work with schemes in which N is invertible, i.e. schemes over $\mathbb{Z}[1/N]$, which is the treatment of [22]. Getting rid of this condition is done in the standard work [36] and makes things much more technical.

So let N be a positive integer, let $S/\mathbb{Z}[1/N]$ a scheme and let E/S be an elliptic curve. Then a point of order N of E/S is meant to be a section $P \in E(S)[N]$ whose pull-back to all geometric fibres of E/S defines a point of order N. Define a contravariant functor

$$F_1(N) : \underline{\mathrm{Sch}}_{\mathbb{Z}[1/N]} \to \underline{\mathrm{Set}}$$

from the category of schemes over $\mathbb{Z}[1/N]$ to the category of sets as follows. We send a scheme S to the set of isomorphism classes of pairs (E, P) where E is an elliptic curve over S and P a point of order N of E/S. And we send a morphism $T \to S$ to the map $F_1(N)(S) \to F_1(N)(T)$ that sends every pair $(E, P)/S$ to its pull-back along $T \to S$.

Theorem 1.9 (Igusa)**.** *Let $N > 3$ be an integer. Then there exists a smooth affine scheme $Y_1(N)$ over $\mathbb{Z}[1/N]$, an elliptic curve $\mathbb{E}$ over $Y_1(N)$ and a point $\mathbb{P}$ of $\mathbb{E}/Y_1(N)$ of order N that satisfies the following universal property: for all schemes $S/\mathbb{Z}[1/N]$ and pairs (E, P) with E/S an elliptic curve and P a point of order N of E/S there are unique morphisms $S \to Y_1(N)$ and $E \to \mathbb{E}$ such that the following diagram is commutative with Cartesian inner square:*

$$\begin{array}{ccc} E & \longrightarrow & \mathbb{E} \\ {\scriptstyle P}\uparrow\downarrow & \square & \downarrow\uparrow{\scriptstyle \mathbb{P}} \\ S & \longrightarrow & Y_1(N) \end{array}$$

Moreover, the geometric fibres of $Y_1(N)/\mathbb{Z}[1/N]$ are irreducible curves.

Note that we abusively use the same notation $Y_1(N)$ as in the previous subsection; we will write subscripts in cases where this abuse might lead to confusion. The scheme $Y_1(N)$ of the theorem represents the functor $F_1(N)$: pulling back $(\mathbb{E}, \mathbb{P})/Y_1(N)$ along morphisms

$S \to Y_1(N)$ defines a functorial bijection between $Y_1(N)(S)$ and $F_1(N)(S)$. Because we can give such an isomorphism of functors, or equivalently, a universal $(\mathbb{E},\mathbb{P})$, we say that $Y_1(N)$ is a *fine moduli space* for the functor $F_1(N)$.

The complex curve $Y_1(N)$ from the previous subsection, together with its moduli description, is canonically isomorphic to the base change $Y_1(N)_{\mathbb{C}}$ of $Y_1(N)_{\mathbb{Z}[1/N]}$ to $\mathbb{C}$. In fact, over $\mathbb{C}$, the universal elliptic curve $\mathbb{E}_{\mathbb{C}}/Y_1(N)_{\mathbb{C}}$ can be described analytically as follows: Consider $\mathbb{C}\times\mathfrak{H}$ as line bundle over $\mathfrak{H}$ and embed $\mathbb{Z}^2\times\mathfrak{H}$ into it by

$$\mathbb{Z}^2\times\mathfrak{H} \hookrightarrow \mathbb{C}\times\mathfrak{H}, \quad ((m,n),z) \mapsto ((mz+n),z).$$

Call the image of this embedding Λ. The quotient $(\mathbb{C}\times\mathfrak{H})/\Lambda$ is an elliptic curve E over $\mathfrak{H}$ whose fibre over $z\in\mathfrak{H}$ is $\mathbb{C}/\Lambda_z$. The section $P:\mathfrak{H}\to E$ defined by $z\mapsto 1/N$ has order N. We have an action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{C}\times\mathfrak{H}$ as follows:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(w,z) := \left(\frac{w}{cz+d}, \frac{az+b}{cz+d}\right).$$

This action respects Λ and therefore induces an action on E. The subgroup of $\mathrm{SL}_2(\mathbb{Z})$ respecting the section P is exactly $\Gamma_1(N)$ and we can in fact describe $\mathbb{E}_{\mathbb{C}}/Y_1(N)_{\mathbb{C}}$ as the quotient of $E/\mathfrak{H}$ by the action of $\Gamma_1(N)$:

$$\mathbb{E}_{\mathbb{C}} \cong \Gamma_1(N)\setminus((\mathbb{C}\times\mathfrak{H})/\Lambda). \tag{1.21}$$

Let us note that from Theorem 1.9 it follows that $Y_1(N)$ has a model over $\mathbb{Q}$ and that for each field extension $K/\mathbb{Q}$ the set $Y_1(N)(K)$ of K-rational points of $Y_1(N)_{\mathbb{Q}}$ is in bijection with the set of isomorphism classes of pairs (E,P) where E is an elliptic curve over K and $P\in E(K)$ is a K-rational point of order N. We furthermore see that for $p\nmid N$ the curve $Y_1(N)_{\mathbb{Q}}$ has a non-singular reduction $Y_1(N)_{\mathbb{F}_p}$ that parametrises all pairs (E,P) with E an elliptic curve over a field K of characteristic p and $P\in E(K)$ a point of order N.

There is another functor that people sometimes use; this is the functor

$$F_\mu(N) : \underline{\mathrm{Sch}} \to \underline{\mathrm{Set}}.$$

It takes a scheme S to the set of pairs (E,ι) where E/S is an elliptic curve and $\iota:\mu_{N,S}\to E$ is a closed immersion of group schemes over S. There exists a fine moduli space $Y_\mu(N)/\mathbb{Z}[1/N]$ for $F_\mu(N)$ as well. Also here we have an isomorphism of $Y_\mu(N)_{\mathbb{C}}$ with the complex curve $Y_1(N)_{\mathbb{C}}$; it is defined by sending z to $(\mathbb{C}/\Lambda_z, \exp(2\pi ik/N)\mapsto k/N \bmod \Lambda)$. In fact, we have an isomorphism of schemes

$$Y_1(N) \cong Y_\mu(N) \tag{1.22}$$

defined as follows. Let $S/\mathbb{Z}[1/N]$ be a scheme and take $(E,P)\in Y_1(S)$. We have to make a point $(E',\iota')\in Y_\mu(S)$. Put $E'=E/\langle P\rangle$ with quotient map $\phi:E\to E'$. For each closed

immersion of group schemes $\iota : \mu_{N,S} \to E'$ we have an endomorphism of $\mu_{N,S}$ that is defined by sending $Q \in \mu_{N,S}(T)$ to $e_N(P,(\iota Q)')$ for any S-scheme T, where $(\iota Q)'$ denotes any point of $E(T)$ that maps to ιQ along ϕ. We take for ι' the ι that makes this endomorphism the identity. Over $\mathbb{C}$ the isomorphism (1.22) can be defined by sending $z \in \mathfrak{H}$ to $w_N(z) = -1/Nz$.

For $Y(N)$ with $N > 2$ there is a similar description as the fine moduli space over $\mathbb{Z}[1/N]$ parametrising all pairs $(E/S, \phi)$ where $\phi : (\mathbb{Z}/N\mathbb{Z})_S \to E(S)[N]$ is an isomorphism of group schemes. In this case, the $Y(N)$ from the previous subsection is a disjoint union of $\phi(N)$ copies of the base change $Y_1(N)_{\mathbb{C}}$ of $Y_1(N)_{\mathbb{Z}[1/N]}$ to $\mathbb{C}$: one for each possible value of the Weil pairing.

One cannot construct $Y_0(N)$ as the fine moduli space parametrising pairs (E,C) of elliptic curve and cyclic subgroups of order N in any sensible meaning. The obstruction lies in the fact that such pairs always have the non-trivial automorphism -1. However, we can do the following. Let the group $G = (\mathbb{Z}/N\mathbb{Z})^\times$ act on $Y_1(N)$ by letting $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ act as $(E,P) \mapsto (E,dP)$ on moduli and define $Y_0(N)$ as the quotient $G \setminus Y_1(N)$. Although $Y_0(N)$ is not a fine moduli space, it is true that for all fields K with $\mathrm{char}(K) \nmid N$ the set $Y_0(N)(K)$ is naturally in bijection with the set of $\overline{K}$-isomorphism classes of pairs (E,C) where E is an elliptic curve over K and $C \subset E$ is a cyclic subgroup of order N defined over K. Here as well $Y_0(N)$ from the previous subsection is canonically isomorphic to the base change of $Y_0(N)_{\mathbb{Z}[1/N]}$ to $\mathbb{C}$.

1.2.3 Moduli interpretation at the cusps

In Subsection 1.2.1 we defined the compact Riemann surfaces $X_0(N)$, $X_1(N)$ and $X(N)$ but so far we only gave moduli descriptions for $Y_0(N)$, $Y_1(N)$ and $Y(N)$. In this subsection we will explain the approach of [22] to extend the moduli interpretation to the cusps.

Néron polygons and generalised elliptic curves

Let n be a positive integer and let k be a field. A *Néron n-gon over k* is defined to be a singular connected curve over k that can be constructed as follows: take n copies of $\mathbb{P}^1_k$, indexed by $\mathbb{Z}/n\mathbb{Z}$ and identify for each $i \in \mathbb{Z}/n\mathbb{Z}$ the point ∞ of the i-th $\mathbb{P}^1$ with the point 0 of the $(i+1)$-st $\mathbb{P}^1$ such that this intersection point is an ordinary double point.

For $a \in \mathbb{P}^1_k(\overline{k})$ and $i \in \mathbb{Z}/n\mathbb{Z}$ we denote the point a of the i-th $\mathbb{P}^1$ of a Néron n-gon by (a,i). The choice of projective coordinates on $\mathbb{P}^1$ allows us to identify $\mathbb{P}^1_k - \{0,\infty\}$ with $\mathbb{G}_{m,k}$, which acts on $\mathbb{P}^1_k$ by $(a,b) \mapsto ab$. This way we give the smooth locus C^{sm} of a Néron n-gon C the structure of a commutative group scheme, where addition is defined as

$$(a,i) + (b,j) := (ab, i+j). \tag{1.23}$$

We use this same formula to equip a Néron n-gon C with an action of C^{sm}.

Note that a Néron n-gon C together with its addition (1.23), admits an action of the group $\mu_n(k)$ by letting $\zeta \in \mu_n(k)$ act as $(a,i) \mapsto (\zeta^i a, i)$. Furthermore, we have an automorphism ι defined on it that sends (a,i) to $(a^{-1},-i)$. In fact

$$\mathrm{Aut}(C,+) \cong \mu_n(k) \times \langle \iota \rangle \tag{1.24}$$

is the group of automorphisms of C that respect the addition.

We are now ready to define the notion of a generalised elliptic curve.

Definition 1.5. Let S be a scheme. Then a *generalised elliptic curve* over S is a scheme E over S that is proper, flat, of finite presentation that comes equipped with a morphism $E^{\mathrm{sm}} \times_S E \xrightarrow{+} E$ that makes E^{sm} into a commutative group scheme acting on E and such that each geometric fibre of E/S is either an elliptic curve or a Néron polygon equipped with an action as in (1.23).

Definition 1.6. If E is a generalised elliptic curve over a scheme S, then *a point of order N* of E/S is meant to be section in $E^{\mathrm{sm}}(S)[N]$ whose pull-back to all geometric fibres defines a point of order N such that the subgroup generated by it meets all irreducible components.

The notion of generalised elliptic curves enables us to generalise Igusa's theorem to $X_1(N)$:

Theorem 1.10 (see [22, Chapter IV])**.** *Let $N > 4$ be an integer. Then there exists a proper smooth scheme $X_1(N)$ over $\mathbb{Z}[1/N]$, a generalised elliptic curve $\mathbb{E}$ over $X_1(N)$ and a point $\mathbb{P}$ of $\mathbb{E}/X_1(N)$ of order N that satisfies the following universal property: for all schemes $S/\mathbb{Z}[1/N]$ and pairs (E,P) with E/S a generalised elliptic curve and $P \in E(S)$ a point of order N there are unique morphisms $S \to X_1(N)$ and $E \to \mathbb{E}$ such that the following diagram is commutative with Cartesian inner square:*

$$\begin{array}{ccc} E & \longrightarrow & \mathbb{E} \\ {\scriptstyle P}\uparrow\downarrow & \square & \downarrow\uparrow{\scriptstyle \mathbb{P}} \\ S & \longrightarrow & X_1(N) \end{array}$$

Moreover, the geometric fibres of $X_1(N)/\mathbb{Z}[1/N]$ are irreducible curves.

The scheme $Y_1(N)$ is naturally an open subscheme of $X_1(N)$ and the complement is called the *cuspidal locus* of $X_1(N)$. We can also extend $Y_\mu(N)$ to cusps and get a scheme $X_\mu(N)$ parametrising pairs (E,ι) of generalised elliptic curves over S together with closed immersions $\iota : \mu_{N,S} \to E$. We require that the image of ι meets the geometric fibres of E in all components. The isomorphism (1.22) extends to an isomorphism $X_1 \cong X_\mu$.

As with $Y_0(N)$, we define $X_0(N)$ by dividing out the group action of $(\mathbb{Z}/N\mathbb{Z})^\times$ defined by $d : (E,P) \mapsto (E,dP)$. Furthermore, there also exists for $N > 2$ a scheme $X(N)$ that is a fine moduli space for pairs $(E,\phi)/S/\mathbb{Z}[1/N]$ with E/S a generalised elliptic curve and $\phi : (\mathbb{Z}/N\mathbb{Z})^2_S \to E^{\mathrm{sm}}$ a closed immersion of S-group schemes meeting all irreducible components of all geometric fibres of E.

Tate curves

We will give an informal discussion on the Tate curve now. Precise results can be found in [22, Chapter VII]. See also [74, Chapter V] for a more elementary and explicit approach. The idea is that for an elliptic curve $E = \mathbb{C}/\Lambda$ over $\mathbb{C}$ we have $E \cong \mathbb{C}^\times/q^{\mathbb{Z}}$ with $q = \exp(2\pi i z)$. An explicit Weierstrass equation for E is

$$E : y^2 + xy = x^3 + a_4(q)x + a_6(q) \tag{1.25}$$

with

$$a_4(q) = -5\sum_{n\geq 1}\sigma_3(n)q^n \quad \text{and} \quad a_6(q) = -\frac{1}{12}\sum_{n\geq 1}(5\sigma_3(n) + 7\sigma_5(n))q^n.$$

An isomorphism $\mathbb{C}^\times/q^{\mathbb{Z}} \to E$ can be given by

$$t \mapsto \left(\sum_{n\in\mathbb{Z}}\frac{q^n t}{(1-q^n t)^2} - 2\sum_{n\geq 1}\sigma_1(n)q^n, \quad \sum_{n\in\mathbb{Z}}\frac{(q^n t)^2}{(1-q^n t)^3} + \sum_{n\geq 1}\sigma_1(n)q^n\right),$$

where of course we send $t \in q^{\mathbb{Z}}$ to $0 \in E$. This isomorphism leads to the following identification of differentials on $\mathbb{C}^\times/q^{\mathbb{Z}}$ and E:

$$\frac{dt}{t} = \frac{dx}{2y+x}.$$

We will use this t-coordinate notation whenever it makes sense.

The Weierstrass equation (1.25) defines a generalised elliptic curve over $\mathbb{Z}[[q]]$. Also, for any $w \in \mathbb{Z}_{>0}$ we can regard (1.25) as a Weierstrass equation for an elliptic curve over the ring $\mathbb{Z}((q^{1/w}))$. We call this the *Tate curve* E_q over $\mathbb{Z}[[q]]$ and $\mathbb{Z}((q^{1/w}))$ respectively. The idea is now that if we move our favourite cusp of width w to ∞ and see $q^{1/w}$ as a local parameter there, then E_q can be seen as a (formal completion of a) universal elliptic curve over a punctured neighbourhood of our cusp. This can in fact be used to describe cusps of $X_1(N)$ over arbitrary fields, not just $\mathbb{C}$.

Let now $N > 4$ and w be integers with $w \mid N$. Let k be a field of characteristic not dividing N that contains all N-th roots of unity and put $R = k[[q^{1/w}]]$ and $K = k((q^{1/w}))$. The Néron model $\mathscr{E}_q$ of E_q over K is the smooth locus of a generalised elliptic curve over R whose special fibre $\overline{\mathscr{E}}_q$ is a Néron w-gon over k. We have canonical isomorphisms $E_q(K) \cong K^\times/q^{\mathbb{Z}}$ and $E_{q,0}(K) \cong R^\times/q^{\mathbb{Z}}$, where the latter is the subset of $E_q(K)$ consisting of points whose specialisation lies in the 0-component of the smooth locus. The component group of $\overline{\mathscr{E}}_q$ is canonically isomorphic to

$$\overline{\mathscr{E}}_q(k)/\overline{\mathscr{E}}_q^0(k) \cong E_q(K)/E_{q,0}(K) \cong (q^{1/w})^{\mathbb{Z}}/q^{\mathbb{Z}} \cong \mathbb{Z}/w\mathbb{Z}.$$

Using the identification $E_q(K) \cong K^\times/q^{\mathbb{Z}}$ we get an isomorphism from $\mu_N(k) \times \mathbb{Z}/w\mathbb{Z}$ to $E_q(K)[N]$, hence a homomorphism to $\overline{\mathscr{E}}_q(k)$, defined by $(\zeta, i) \mapsto \zeta q^{i/w}$. This gives us a description for all the cusps: to write down a cusp of $X_1(N)(k)$ is suffices to write down a $w \mid N$

and a point (ζ, i) of order N of $E_q(K)$ satisfying $\gcd(i,w) = 1$; this last condition is necessary so as to meet the requirement that the subgroup generated by it meets all the components of the special fibre. Be aware of the fact that this does not lead to a unique notation for cusps because of (1.24).

Let us work out what this means for a cusp $\gamma\infty$ with $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ in the upper half plane model for $X_1(N)_{\mathbb{C}}$. Write $z \in \mathfrak{H}$ as $\gamma\omega$ with $\omega \in \mathfrak{H}$ and let $w = w(\gamma) = N/\gcd(c,N)$ be the width of $\gamma\infty$ in $X_1(N)$. If we put $q_\gamma = \exp(2\pi i\omega)$ then $q_\gamma^{1/w}$ is a local parameter for $X_1(N)_{\mathbb{C}}$ at $\gamma\infty$. The fibre of $(\mathbb{E},\mathbb{P})$ above $z = \gamma\omega$ is then uniquely isomorphic to

$$(\mathbb{E},\mathbb{P})_z \cong \left(\mathbb{C}/\Lambda_\omega, \frac{c\omega + d}{N}\right).$$

In terms of the parameter $q_\gamma^{1/w}$ this can be written as

$$(\mathbb{E},\mathbb{P})_z \cong \left(\mathbb{C}^\times/q^{\mathbb{Z}}, \zeta_N^d q_\gamma^{c/N}\right) = \left(\mathbb{C}^\times/q^{\mathbb{Z}}, \zeta_N^d (q_\gamma^{1/w})^{c/\gcd(c,N)}\right),$$

where we have put $\zeta_N = \exp(2\pi i/N)$. Our conclusion is that $\mathbb{E}_{\gamma\infty}$ is the Néron w-gon with $w = N/\gcd(N,c)$ and for the point of order N on it we have

$$\mathbb{P}_{\gamma\infty} = \left(\exp(2\pi i d/N), \frac{c}{\gcd(c,N)}\right).$$

Note that the cusp does not uniquely determine the number d, but the different choices lead to isomorphic objects.

Let us note that in this way we can see that the cusp $0 \in \mathfrak{H}^*$ is defined over $\mathbb{Q}$: it corresponds to $(c,d) = (1,0)$ and thus to an N-gon with the point $(1,1)$ on it, which is invariant under the action of $\mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$. The cusp $\infty \in \mathfrak{H}^*$ is not defined over $\mathbb{Q}$: it corresponds to $(c,d) = (0,1)$ and thus to a 1-gon with the point $(\zeta_N, 0)$ on it, whose isomorphism class in only invariant under the stabiliser subgroup of $\mathbb{Q}(\zeta_N + \zeta_N^{-1})$.

1.2.4 Katz modular forms

The algebraic description of modular curves allows us to give an algebraic description of modular forms as global sections of certain line bundles over modular curves. These sections are sometimes called *Katz modular forms* and in particular they allow us to speak about modular forms for $\Gamma_1(N)$ over any $\mathbb{Z}[1/N]$-algebra.

Let S be a scheme and let E/S be a generalised elliptic curve. The curve E has a sheaf of relative differentials $\Omega^1_{E/S}$ as well as a zero section $0 : S \to E$. We put

$$\omega_{E/S} := 0^*\Omega^1_{E/S},$$

which is a line bundle on S. In particular, for $N > 4$ and $k \in \mathbb{Z}$ we can consider the line bundle $\omega^{\otimes k}_{\mathbb{E}_{\mathbb{C}}/Y_1(N)_{\mathbb{C}}}$ on $Y_1(N)_{\mathbb{C}}$, using the notation of (1.21). Using the same construction of $\omega_{E/S}$

in an analytic context, the sheaf $\omega^{\otimes k}_{((\mathbb{C}\times\mathfrak{H})/\Lambda)/\mathfrak{H}}$ is a free $\mathscr{O}_{\mathfrak{H}}$-module of rank 1, generated by $(dw)^{\otimes k}$, where w denotes the coordinate on the factor $\mathbb{C}$. In particular, any holomorphic function $f:\mathfrak{H}\to\mathbb{C}$ can be seen as the section $f(z)(dw)^{\otimes k}$ of $\omega^{\otimes k}_{((\mathbb{C}\times\mathfrak{H})/\Lambda)/\mathfrak{H}}$ and vice versa. The action of $\gamma=\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)\in \mathrm{SL}_2(Z)$ on $(\mathbb{C}\times\mathfrak{H})/\Lambda$ sends $f(z)(dw)^{\otimes k}$ to $(cz+d)^{-k}f(\gamma z)(dw)^{\otimes k}$. Using that $\Gamma_1(N)$ acts freely on $\mathfrak{H}$, we see now that $H^0(Y_1(N)(\mathbb{C}),\omega^{\otimes k})$ is isomorphic to the space of holomorphic functions on $\mathfrak{H}$ that satisfy the weight k modular transformation property for $\Gamma_1(N)$.

Now, we extend this to $\mathbb{E}_{\mathbb{C}}/X_1(N)_{\mathbb{C}}$. Global sections of $H^0(X_1(N)(\mathbb{C}),\omega^{\otimes k})$ can still be seen as holomorphic functions $f:\mathfrak{H}\to\mathbb{C}$ satisfying the weight k modular transformation property for $\Gamma_1(N)$. Using the description of neighbourhoods of cusps as Tate curves, one can see that the extra condition at the cusps is simply that f has to be holomorphic at the cusps. So we have an isomorphism

$$M_k(\Gamma_1(N))\cong H^0\left(X_1(N)_{\mathbb{C}},\omega^{\otimes k}_{\mathbb{E}_{\mathbb{C}}/X_1(N)_{\mathbb{C}}}\right).$$

Cusps forms are modular forms that vanish at the cusps, so we have

$$S_k(\Gamma_1(N))\cong H^0\left(X_1(N)_{\mathbb{C}},\omega^{\otimes k}_{\mathbb{E}_{\mathbb{C}}/X_1(N)_{\mathbb{C}}}(-\mathrm{cusps})\right).$$

Here, cusps denotes the divisor of all cusps, all counted with multiplicity 1. The above isomorphisms inspire us to write down the definition of Katz modular forms

Definition 1.7. Let $N>4$ and k be integers. Let A be a $\mathbb{Z}[1/N]$-algebra. Then the space of Katz modular forms for $\Gamma_1(N)$ over A is defined to be the A-module

$$M_k(\Gamma_1(N),A):=H^0\left(X_1(N)_A,\omega^{\otimes k}_{\mathbb{E}_A/X_1(N)_A}\right)$$

and the space of Katz cusp forms over A is defined as the A-module

$$S_k(\Gamma_1(N),A):=H^0\left(X_1(N)_A,\omega^{\otimes k}_{\mathbb{E}_A/X_1(N)_A}(-\mathrm{cusps})\right).$$

Let us remark that there is an isomorphism of line bundles

$$\omega^{\otimes 2}_{\mathbb{E}/X_1(N)}\xrightarrow{\sim}\Omega^1_{X_1(N)/\mathbb{Z}[1/N]}(\mathrm{cusps}),$$

called the *Kodaira-Spencer isomorphism*, see [35, Subsection A1.3.17]. Over $\mathbb{C}$ it is defined by $f(z)(dw)^{\otimes 2}\mapsto(2\pi i)^{-1}f(z)dz$. It is compatible with base-change. A consequence of this isomorphism is

$$S_2(\Gamma_1(N),A)\cong H^0\left(X_1(N)_A,\Omega^1_{X_1(N)_A/A}\right),$$

which is something that we shall use later in our calculations.

q-expansions

We can define the q-expansion of a Katz modular form of level N and weight k algebraically. Let A be an algebra over $\mathbb{Z}[1/N,\zeta_N]$ and consider the Tate curve E_q over $A[[q]]$ together with the point $t=\zeta_N \bmod q^{\mathbb{Z}}$ on it. By Theorem 1.10, the pair $(E_q,\zeta_N \bmod q^{\mathbb{Z}})$ is the base-change of $\mathbb{E}/X_1(N)$ along an $A[[q]]$-valued point of $X_1(N)$. This base-change gives a pull-back homomorphism

$$M_k(\Gamma_1(N),A)=H^0\left(X_1(N)_A,\omega^{\otimes k}_{\mathbb{E}_A/X_1(N)_A}\right)\to H^0\left(\operatorname{Spec}A[[q]],\omega^{\otimes k}_{E_q/A[[q]]}\right).$$

The latter object is a free module over $A[[q]]$ generated by $(dt/t)^{\otimes k}$, where dt/t is the standard differential on E_q. So we obtain a homomorphism of A-modules

$$M_k(\Gamma_1(N),A)\to A[[q]]\left(\frac{dt}{t}\right)^{\otimes k}.$$

Applying this homomorphism and dropping the factor $(dt/t)^{\otimes k}$ defines for $f\in M_k(\Gamma_1(N),A)$ its q-expansion in $A[[q]]$. Formation of this q-expansion commutes with base-change. Over $\mathbb{C}$ this q-expansion coincides with the usual q-expansion of $f\in M_k(\Gamma_1(N))$ since the pair $(E_q,\zeta_N \bmod q^{\mathbb{Z}})$ corresponds to a neighbourhood of the cusp ∞.

A thorn in the eye here is that the ring A has to contain a primitive N-th root of unity, while we wish to work, for instance, over $\mathbb{Q}$. Luckily, we can resolve this problem. So let A be a $\mathbb{Z}[1/N]$-algebra. Remember that we have an isomorphism

$$X_1(N)\cong X_\mu(N).$$

This induces an isomorphism

$$M_k(\Gamma_1(N),A)\cong H^0\left(X_\mu(N),\omega^{\otimes k}_{(\mathbb{E}/\langle\mathbb{P}\rangle)_A/X_\mu(N)_A}\right).$$

Now, consider the pair (E_q,ι) over $A[[q]]$ with ι the canonical injection $\mu_{N,A}\hookrightarrow E_q$ via the t-coordinate. We repeat the above argument and obtain a map

$$M_k(\Gamma_1(N),A)\to A[[q]].$$

Over $\mathbb{C}$, the q-series of $f\in M_k(\Gamma_1(N),A)$ obtained in this way coincides with the usual q-expansion of $W_N(f)$. So we have the following proposition:

Proposition 1.1. *Let N and k be positive integers with $N>4$. Let A be a subring of $\mathbb{C}$ in which N is invertible. Then the image of the canonical map*

$$M_k(\Gamma_1(N),A)\to M_k(\Gamma_1(N))$$

consist exactly of those forms f for which the q-expansion of $W_N(f)$ has coefficients in A.

1.2.5 Diamond and Hecke operators

On the modular curve $X_1(N)$ we have a diamond operator $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ that we have in fact already mentioned before. It acts on a pair (E,P) by

$$\langle d\rangle(E,P) \mapsto (E,dP).$$

By pull-back it defines an operator on the space $S_k(\Gamma_1(N),\mathbb{Q})$ for any $\mathbb{Z}[1/N]$-algebra A. Over $\mathbb{C}$ this coincides with the usual diamond operator on $S_k(\Gamma_1(N))$.

Hecke operators are defined on the Jacobian $J_1(N)_\mathbb{Q}$ of $X_1(N)_\mathbb{Q}$ as follows. For a positive integer n, we let T_n be the endomorphism of $J_1(N)_\mathbb{Q}$ induced by the following map on divisors:

$$T_n : (E,P) \mapsto \sum_{\substack{C\subset E \text{ subgroup of order } n,\\ C\cap\langle P\rangle=0}} (E/C,\, P \bmod C).$$

Here, E is a true elliptic curve, not a generalised one. Now choose a rational point Q in $X_1(N)_\mathbb{Q}$, for instance (N-gon, $(1,1)$). Then we can embed $X_1(N)_\mathbb{Q}$ into $J_1(N)_\mathbb{Q}$ by sending P to $P-Q$. This embedding induces an isomorphism

$$H^0\left(J_1(N)_\mathbb{Q}, \Omega^1_{J_1(N)_\mathbb{Q}/\mathbb{Q}}\right) \xrightarrow{\sim} H^0\left(X_1(N)_\mathbb{Q}, \Omega^1_{X_1(N)_\mathbb{Q}/\mathbb{Q}}\right) \cong S_2(\Gamma_1(N),\mathbb{Q})$$

which is independent of the choice of Q. The Hecke operators on $J_1(N)_\mathbb{Q}$ induce operators on the space $S_2(\Gamma_1(N),\mathbb{Q})$ via this isomorphism. Over $\mathbb{C}$, they coincide with the usual Hecke operators on $S_2(\Gamma_1(N))$. For a general definition of Hecke operators on the space $S_k(\Gamma_1(N),\mathbb{Q})$, see [35, 1.11].

Eichler-Shimura relation

Consider the modular curve $X_1(N)$ and let p be a prime not dividing N. On the Jacobian $J_1(N)_{\mathbb{F}_p}$ of $X_1(N)_{\mathbb{F}_p}$ we have several operators. First of all, we have the Frobenius operator Frob_p, defined on coordinates by $x \mapsto x^p$. This operator has a dual Ver_p, called the *Verschiebung*. It satisfies $\mathrm{Frob}_p \circ \mathrm{Ver}_p = \mathrm{Ver}_p \circ \mathrm{Frob}_p = p$ as endomorphisms of $J_1(N)_{\mathbb{F}_p}$. Viewing the Jacobian as a covariant (Albanese) functor of curves, the diamond operator $\langle p\rangle$ on $X_1(N)_{\mathbb{F}_p}$ defines an operator on $J_1(N)_{\mathbb{F}_p}$ that we shall also denote by $\langle p\rangle$. Furthermore, the Hecke operator T_p on $J_1(N)_\mathbb{Q}$ defines an operator on the Néron model of $J_1(N)_\mathbb{Q}$ over $\mathbb{Z}$. The fibre of this Néron model over p is $J_1(N)_{\mathbb{F}_p}$ so we have an operator T_p on $J_1(N)_{\mathbb{F}_p}$ as well. The following relation between all these operators holds in $\mathrm{End}(J_1(N)_{\mathbb{F}_p})$:

$$T_p = \mathrm{Frob}_p + \langle p\rangle \,\mathrm{Ver}_p\,. \tag{1.26}$$

This relation is called the *Eichler-Shimura relation* in $\mathrm{End}(J_1(N)_{\mathbb{F}_p})$.

1.3 Galois representations associated to newforms

Modular forms turn out to be strongly related to the representation theory of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, in particular to the 2-dimensional representations over finite fields and ℓ-adic fields. As

in the previous sections, we will not present the material in its most general and complete form. Interested readers could consult for example [19] or [65] for a general treatment of representation theory and [62] or [85] for Galois representations.

1.3.1 Basic definitions

Let G be a group and let K be a field. Assume that both G and K are equipped with a topology; when for groups or fields considered in this text no standard topology exists or no topology has been specified, the topology will be assumed to be discrete. For $n \in \mathbb{Z}_{\geq 0}$, an *$n$-dimensional linear representation of G over K* is a continuous homomorphism

$$\rho : G \to \mathrm{GL}_n(K)$$

or, equivalently, a continuous linear action of G on an n-dimensional vector space over K. A topology on $\mathrm{GL}_n(K)$ is defined in the following way: embed $\mathrm{GL}_n(K)$ into $\mathrm{M}_n(K) \times \mathrm{M}_n(K)$ by $g \mapsto (g, g^{-1})$ and give $\mathrm{M}_n(K) \times \mathrm{M}_n(K) \cong K^{n^2}$ the product topology. This is to ensure that the map $g \mapsto g^{-1}$ will be continuous.

The conventions here are not completely standard. In the literature, infinite-dimensional and non-continuous representations are considered as well. Representations of G on two K-vector spaces V and V' are called isomorphic if there is a linear isomorphism between V and V' that respects the G-action.

A representation $\rho : G \to \mathrm{GL}(V)$ is said to be *irreducible* if V is nonzero and the only subspaces of V fixed by G are 0 and V. It is said to be *absolutely irreducible* if the representation $G \to \mathrm{GL}(V \otimes_K \overline{K})$ obtained from ρ is irreducible. A representation $\rho : G \to \mathrm{GL}(V)$ is said to be *semi-simple* if it can be written as a direct sum of irreducible representations. If G is a finite group then any finite-dimensional representation of G over a field of characteristic not dividing $\#G$ is semi-simple (Maschke's theorem). An example of a representation that is not semi-simple can be obtained as follows: Let p be any prime number and take $\rho : \mathbb{Z}/p\mathbb{Z} \to \mathrm{GL}_2(\mathbb{F}_p)$ defined by

$$\rho(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \tag{1.27}$$

The following two theorems on semi-simple representations are important to us.

Theorem 1.11 (cf. [10, Proposition 3.12])**.** *Let G be a group, let K be a field of characteristic 0 and let ρ and ρ' be n-dimensional semi-simple representations of G over K. If $\mathrm{tr}(\rho(g)) = \mathrm{tr}(\rho'(g))$ holds for all $g \in G$ then ρ and ρ' are isomorphic.*

Theorem 1.12 (Brauer-Nesbitt, [19, Theorem 30.16])**.** *Let G be a finite group and let ρ and ρ' be finite-dimensional semi-simple representations of G over a field. If for all $g \in G$ the characteristic polynomials of $\rho(g)$ and $\rho'(g)$ coincide, then ρ and ρ' are isomorphic.*

To any finite-dimensional representation $\rho : G \to \mathrm{GL}(V)$ we can attach a semi-simple representation $\rho^{\mathrm{ss}} : G \to \mathrm{GL}(V)$ as follows. There is a maximal chain $0 = V_0 \subsetneq V_1 \subsetneq \cdots \subsetneq V_r = V$

of G-stable subspaces. The action of G on V induces an action on each successive quotient V_{i+1}/V_i and we define ρ^{ss} to be the action of G on the direct sum of these successive quotients. The representation ρ^{ss} is called the *semi-simplification* of ρ; by the Jordan-Hölder theorem it is well-defined, i.e. independent of the chosen chain. In any case, the process of semi-simplification does not affect the function $g \mapsto \mathrm{charpol}(\rho(g))$.

1.3.2 Galois representations

Let now G be the group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ with its Krull topology. For each prime p, we fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$. This defines an embedding $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \hookrightarrow \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ whose image is a decomposition group D_p at p; we will identify $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ with D_p. Every representation ρ of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ defines a representation ρ_p of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ by restriction. A representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathrm{GL}_n(K)$ is called *unramified* if it is trivial on its inertia subgroup. In that case it factors through the quotient $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \twoheadrightarrow \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$ and we have a well-defined element $\rho(\mathrm{Frob}_p) \in \mathrm{GL}_2(K)$. A representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(K)$ is called unramified at p if the restriction of ρ at p is unramified; this notion is independent of the choice of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \hookrightarrow \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. If ρ is unramified at p then $\rho(\mathrm{Frob}_p)$ is well-defined up to conjugacy; in particular $\mathrm{charpol}(\rho(\mathrm{Frob}_p))$ will be well-defined in that case.

One-dimensional Galois representations

The Kronecker-Weber theorem allows us to classify the 1-dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. The maximal abelian extension of $\mathbb{Q}$ is the field $\mathbb{Q}(\mu_\infty)$ obtained by adjoining all roots of unity in $\overline{\mathbb{Q}}$ to $\mathbb{Q}$. Its Galois group $\mathrm{Gal}(\mathbb{Q}(\mu_\infty)/\mathbb{Q})$ is canonically isomorphic to $\hat{\mathbb{Z}}^\times$; the isomorphism $\hat{\mathbb{Z}}^\times \xrightarrow{\sim} \mathrm{Gal}(\mathbb{Q}(\mu_\infty)/\mathbb{Q})$ is given by letting $\alpha \in \hat{\mathbb{Z}}^\times$ send a root unity ζ to ζ^α (which is well-defined). This implies that for any topological field K, giving a 1-dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is equivalent to giving a continuous homomorphism $\hat{\mathbb{Z}}^\times \to K^\times$.

A particular example that is interesting to us is the case $K = \overline{\mathbb{Q}}_\ell$. We canonically have a surjection $\hat{\mathbb{Z}}^\times \twoheadrightarrow \mathbb{Z}_\ell^\times$ and an embedding $\mathbb{Z}_\ell^\times \hookrightarrow \mathbb{Q}_\ell^\times \subset \overline{\mathbb{Q}}_\ell^\times$. Composing these two homomorphisms gives a $\mathbb{Q}_\ell^\times$-valued character of $\hat{\mathbb{Z}}^\times$ that corresponds to a 1-dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ that is known as the *ℓ-adic cyclotomic character* and that is denoted by χ_ℓ. The representation χ_ℓ is unramified outside ℓ and for all primes $p \neq \ell$ we have

$$\chi_\ell(\mathrm{Frob}_p) = p \in \mathbb{Q}_\ell^\times.$$

This representation factors through $\mathrm{Gal}(\mathbb{Q}(\mu_{\ell^\infty})/\mathbb{Q})$, where $\mathbb{Q}(\mu_{\ell^\infty})$ is the extension of $\mathbb{Q}$ obtained by adjoining all roots of unity of ℓ-primary order.

For each $N \in \mathbb{Z}_{>0}$ we have a canonical surjection $\hat{\mathbb{Z}}^\times \twoheadrightarrow (\mathbb{Z}/N\mathbb{Z})^\times$ and we can write down a character of $\mathbb{Z}^\times$ by writing down a character $\varepsilon : (\mathbb{Z}/N\mathbb{Z})^\times \to \mu(\overline{\mathbb{Q}}_\ell)$. By abuse of notation, we will also write the corresponding character of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ as ε. It is unramified outside N, factors through $\mathrm{Gal}(\mathbb{Q}(\mu_N)/\mathbb{Q})$ and using our abusive notation it satisfies $\varepsilon(\mathrm{Frob}_p) = \varepsilon(p)$ for all $p \nmid N$. In particular we can make 1-dimensional representations of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ over $\overline{\mathbb{Q}}_\ell$

of the form $\varepsilon\chi_\ell^n$ where ε is associated to a character of $(\mathbb{Z}/N\mathbb{Z})^\times$ for some N and n is an integer.

We can also take $K = \overline{\mathbb{F}}_\ell$. Any continuous homomorphism $\varepsilon : \hat{\mathbb{Z}}^\times \to \overline{\mathbb{F}}_\ell^\times$ factors as

$$\varepsilon : \hat{\mathbb{Z}}^\times \twoheadrightarrow (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{F}_\lambda^\times \subset \overline{\mathbb{F}}_\ell^\times$$

for some $N \in \mathbb{Z}_{>0}$ and some finite extension $\mathbb{F}_\lambda$ of $\mathbb{F}_\ell$. Again if we denote the corresponding character of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ by ε as well then we have the abusively written identity $\varepsilon(\mathrm{Frob}_p) = \varepsilon(p) \in \overline{\mathbb{F}}_\ell^\times$ for $p \nmid N$. A special example is the mod ℓ cyclotomic character $\overline{\chi}_\ell$. Here we take $N = \ell$ and use the canonical map $(\mathbb{Z}/\ell\mathbb{Z})^\times \to \mathbb{F}_\ell^\times \subset \overline{\mathbb{F}}_\ell^\times$. It satisfies $\overline{\chi}_\ell(\mathrm{Frob}_p) = p \in \mathbb{F}_\ell^\times$ for $p \neq \ell$. This corresponds to the well-known canonical isomorphism $\mathrm{Gal}(\mathbb{Q}(\mu_\ell)/\mathbb{Q}) \xrightarrow{\sim} (\mathbb{Z}/\ell\mathbb{Z})^\times$.

1.3.3 ℓ-Adic representations associated to newforms

It was a conjecture of Ramanujan and Petersson that for a newform f of level N and weight k, the inequality

$$|a_p| \leq 2p^{(k-1)/2}$$

holds for all primes $p \nmid N$. The inequality $|\tau(p)| \leq 2p^{11/2}$ mentioned in Subsection 1.1.2 is a special case of this, conjectured by Ramanujan; Petersson formulated the conjecture for more general newforms. Later, Serre refined this conjecture to a more delicate conjecture about Galois representations, which was already known to hold by Eichler and Shimura for weight $k = 2$, and later proved by Deligne for weights $k > 2$ [21] and by Deligne and Serre for $k = 1$ [23]. The proven form of the conjecture is as follows:

Theorem 1.13. *Let k and N be positive integers. Let $f \in S_k(\Gamma_1(N))$ be a newform and let K_f be the coefficient field of f. Choose a rational prime ℓ and a prime λ of K_f lying over ℓ. Then there is an irreducible representation*

$$\rho = \rho_{f,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(K_{f,\lambda})$$

that is unramified outside $N\ell$ and such that for each prime $p \nmid N\ell$ the characteristic polynomial of $\rho(\mathrm{Frob}_p)$ satisfies

$$\mathrm{charpol}(\rho(\mathrm{Frob}_p)) = x^2 - a_p(f)x + \varepsilon_f(p)p^{k-1}.$$

Furthermore, the representation ρ is unique up to isomorphism and for each $p \nmid N\ell$ the complex roots of $\mathrm{charpol}(\mathrm{Frob}_p)$ *both have their absolute value equal to $p^{(k-1)/2}$.*

The representation ρ in the theorem is called the *λ-adic representation associated to f*. It is clear that this theorem implies the conjecture of Ramanujan and Petersson, as the trace is the sum of the roots of the characteristic polynomial. Also, it follows from this theorem that $\rho = \rho_{f,\lambda}$ is *odd*, which means that for a complex conjugation $c \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ we have $\det\rho(c) = -1$. This holds because of $\det\rho = \varepsilon_f\chi_\ell^{k-1}$ and the fact that the character and the weight of a newform have the same parity. Let us for completeness say what happens with $|a_p(f)|$ for $p \mid N$.

Theorem 1.14. *Let $f \in S_k(N,\varepsilon)$ be a newform and let p be a prime dividing N. Then we have*

$$|a_p(f)| = \begin{cases} p^{(k-1)/2} & \text{if } N(\varepsilon) \nmid \frac{N}{p}, \\ p^{(k-2)/2} & \text{if } N(\varepsilon) \mid \frac{N}{p} \text{ and } p^2 \nmid N, \\ 0 & \text{if } N(\varepsilon) \mid \frac{N}{p} \text{ and } p^2 \mid N. \end{cases}$$

For a proof of this, see [58, Theorems 2 & 3 and Corollary 1] or [50, Theorem 3].

We will now indicate where the representations $\rho_{f,\lambda}$ can be found. Let $f \in S_k(\Gamma_1(N))$ be a newform, let $\mathbb{T}$ be the Hecke algebra associated to $S_k(\Gamma_1(N))$ and consider the map $\theta_f : \mathbb{T} \to \mathbb{C}$ defined by $T_n \mapsto a_n(f)$ and $\langle d \rangle \mapsto \varepsilon_f(d)$. Also, choose a rational prime ℓ and a prime $\lambda \mid \ell$ of K_f.

For $k = 2$ we can find the representation as follows. First of all, we have the ℓ-adic Tate module of $J_1(N)$:

$$T_\ell(J_1(N)) := \varprojlim_n J_1(N)(\overline{\mathbb{Q}})[\ell^n],$$

where the maps in the projective system are multiplication by ℓ. This is a free $\mathbb{Z}_\ell$-module of rank $2g(X_1(N))$, equipped with an linear action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Let $\mathbb{T}$ be the Hecke algebra associated to $S_2(\Gamma_1(N))$. Integration defines a perfect pairing between $H_1(X_1(N)(\mathbb{C}),\mathbb{C})$ and $S_2(\Gamma_1(N)) \oplus \overline{S}_2(\Gamma_1(N))$. Also, $\mathbb{T}$ acts on $H_1(X_1(N)(\mathbb{C}),\mathbb{Z}) \cong H_1(J_1(N)(\mathbb{C}),\mathbb{Z})$ and this action is self-adjoint with respect to the integration pairing. If follows that $T_\ell(J_1(N)) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$ is a free $\mathbb{T}_{\mathbb{Q}_\ell}$-module of rank 2. We can describe the space $V_{f,\lambda}$ as the tensor product of $T_\ell(J_1(N))$ and $K_{f,\ell}$ over $\mathbb{T}_{\mathbb{Z}_\ell}$. Here K_f obtains its $\mathbb{T}$-module structure via θ_f and it gets the action $\rho'_{f,\lambda}$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ from the one on $T_\ell(J_1(N))$.

Now, let $p \nmid N\ell$ be a prime. By proper smooth base-change, the action of $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $T_\ell(J_1(N))$ coincides with the action of Frob_p on $T_\ell(J_1(N)_{\mathbb{F}_p})$. From the Eichler-Shimura relation (1.26) it already follows that $\rho'_{f,\lambda}(\mathrm{Frob}_p)$ is a root of $x^2 - a_p(f)x + \varepsilon_f(p)p$. Now, if $\rho'_{f,\lambda}(\mathrm{Frob}_p)$ is not a scalar matrix, this already shows that $x^2 - a_p(f)x + \varepsilon_f(p)p$ is indeed its characteristic polynomial. Using the Weil pairing on $T_\ell(J_1(N)_{\mathbb{Q}})$ one can show that $\det \rho'_{f,\lambda} = \varepsilon_f \chi_\ell$ so that in general we have $\mathrm{charpol}(\rho'_{f,\lambda}) = x^2 - a_p(f)x + \varepsilon_f(p)p$ and thus $\rho_{f,\lambda} \cong \rho'_{f,\lambda}$.

For $k > 2$ the construction is more technical and uses étale cohomology. Replace N by a multiple rN with $\gcd(r,N) = 1$ if this is necessary to obtain $N > 4$. Consider the universal elliptic curve $\pi : \mathbb{E}_{\overline{\mathbb{Q}}} \to Y_1(N)_{\overline{\mathbb{Q}}}$ and the ℓ-adic étale sheaf

$$\mathscr{F}_{k,\ell} := \mathrm{Sym}^{k-2} R^1\pi_* \mathbb{Q}_\ell.$$

This is a locally free sheaf of $\mathbb{Q}_\ell$-vector spaces of dimension $k-1$. Now put

$$W_\ell := \mathrm{Hom}_{\mathbb{Q}_\ell}\left(H^1_{\text{ét}}(X_1(N)_{\overline{\mathbb{Q}}}, j_*\mathscr{F}_{k,\ell}), \mathbb{Q}_\ell\right)$$

with $j : Y_1(N) \hookrightarrow X_1(N)$ the natural embedding. It can be shown that there are natural actions of $\mathbb{T}_k(N)$ and $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on W_ℓ that allow us to obtain $\rho_{f,\lambda}$ as the tensor product of W_ℓ with $K_{f,\lambda}$ over $\mathbb{T}_\mathbb{Q}$. We won't be using this construction in our calculations and we refer to [21] for the details.

In the case $k = 1$ no direct geometric construction is known, but a proof of existence was given by Deligne and Serre [23]. The essential idea of their proof is as follows. For any prime ℓ, a form of weight one is congruent to a form of weight ℓ modulo ℓ, a case in which the existence of a representation is already known. Reducing mod a prime above ℓ we get a representation $\overline{\rho}_{f,\ell}$ over $\overline{\mathbb{F}}_\ell$. Combining asymptotic properties of $a_p(f)$ mod ℓ for large ℓ and $|a_p(f)|$ they concluded that the set $\{a_p(f)\}$ should be finite and that in fact a representation over K_f should exist for f. So not only over all $K_{f,\lambda}$ there exists a representation in this case but also over $\mathbb{C}$.

1.3.4 Mod ℓ representations associated to newforms

The representations $\rho_{f,\lambda}$ are uncountable objects. This implies that we will not be able to compute them precisely, except in some special cases. So if we want to compute them then we have to approximate them, like one approximates real numbers by floating point numbers. The approximations that we will study are representations $\overline{\rho} = \overline{\rho}_{f,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_\lambda)$ that have $\mathrm{charpol}(\overline{\rho}(\mathrm{Frob}_p))$ congruent to $X^2 - a_p(f)X + \varepsilon(p)p^{k-1} \bmod \lambda$.

Let G be a compact group, let K be an ℓ-adic field with residue field k and let $\rho : G \to \mathrm{GL}_n(K)$ be a semi-simple representation. From the compactness of G it follows that K^n has a G-stable $\mathcal{O}_K$-sublattice: if $\Lambda \subset K^n$ is any $\mathcal{O}_K$-lattice, then the $\mathcal{O}_K$-module generated by $G\Lambda$ is a G-stable $\mathcal{O}_K$-lattice. Reducing this lattice modulo the prime λ of K we obtain a 2-dimensional representation of G over k. This representation depends in general on the choice of the lattice. However, the Brauer-Nesbitt theorem shows that its semi-simplification $\overline{\rho}$ is unique up to isomorphism (note that since k is finite, the representation factors through a finite quotient of G). This semi-simple representation $\overline{\rho}$ is called the *reduction* of ρ modulo λ.

This shows that the above mentioned representations $\overline{\rho}_{f,\lambda}$ at least do exist. We can also find them concretely. Assume for this that $\overline{\rho}_{f,\lambda}$ is absolutely irreducible, which is the most interesting case anyway.

The case $k = 2$

The above mentioned construction of $\rho_{f,\lambda}$ suggests that we should look inside Jacobians of modular curves.

Theorem 1.15 (Boston-Lenstra-Ribet [9, Theorem 2])**.** *Let $f \in S_2(\Gamma_1(N))$ be a newform and let λ be a prime of K_f such that $\overline{\rho}_{f,\lambda}$ is absolutely irreducible. Let $\mathbb{T}$ be the Hecke algebra associated to $S_2(\Gamma_1(N))$ and consider the map $\overline{\theta}_{f,\lambda} : \mathbb{T} \to \mathbb{F}_\lambda$ defined by $T_n \mapsto a_n \bmod \lambda$ and*

$\langle d \rangle \mapsto \varepsilon_f(d) \bmod \lambda$. *Let* $\mathfrak{m} = \mathfrak{m}_f \subset \mathbb{T}$ *be the kernel of* $\overline{\theta}_{f,\lambda}$. *Then the* $(\mathbb{T}/\mathfrak{m})[\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})]$-*module* $J_1(N)(\overline{\mathbb{Q}})[\mathfrak{m}]$ *is a direct sum of copies of* $\overline{\rho}_{f,\lambda}$.

If we take $\mathfrak{m}$ as in the theorem then from the construction of $\rho_{f,\lambda}$ it follows a priori that $\overline{\rho}_{f,\lambda}$ is an irreducible constituent of $J_1(N)(\overline{\mathbb{Q}})[\mathfrak{m}^r]$ for some $r > 0$. An argument of Mazur [52, Section 14] shows that we can in fact take $r = 1$ here, showing that the number of copies in Theorem 1.15 is positive. The map $\overline{\theta}_{f,\lambda}$ mentioned in Theorem 1.15 need not be surjective. So it may happen that the representation $\overline{\rho}_{f,\lambda}$ is actually defined over a field that is smaller than $\mathbb{F}_\lambda$.

The case $k \neq 2$

If we write $N = N'\ell^n$ with $\ell \nmid N'$ then it can be shown that there is a newform f' of weight k and level dividing N', a prime λ' of $K_{f'}$ and embeddings of $\mathbb{F}_\lambda$ and $\mathbb{F}'_\lambda$ into $\overline{\mathbb{F}}_\ell$ such that for all n coprime to N we have in $\overline{\mathbb{F}}_\ell$ an equality of $a_n(f) \bmod \lambda$ with $a_n(f') \bmod \lambda'$. For a proof of this see [61, Theorem 2.1] and [12, Proposition 1.1]. In other words, without loss of generality we can and do assume $\ell \nmid N$.

If we let the weight vary, we can find more congruences. In fact, [61, Theorem 2.2] states that for $k \leq \ell + 1$ there is a newform f' of level dividing $N\ell$ and weight 2 such that in the notation as above, $a_n(f) \bmod \lambda$ is equal to $a_n(f') \bmod \lambda'$ for n coprime to $N\ell$. So also in this case, we can find the representation inside the Jacobian of a modular curve. If we have $k > \ell + 1$ then the representation $\overline{\rho} = \overline{\rho}_{f,\lambda}$ might not always be present inside the ℓ-torsion of some $J_1(M)$ but there is a twist

$$\overline{\rho} \otimes \overline{\chi}_\ell^n : \sigma \mapsto \overline{\rho}(\sigma)\overline{\chi}_\ell^n(\sigma)$$

which does belong to a form of weight at most $\ell + 1$, hence can be reduced to weight 2 again; see [27, Section 7].

In conclusion, if $\overline{\rho}_{f,\lambda}$ is absolutely irreducible, we can always reduce to weight 2 and work inside the ℓ-torsion of the Jacobian of some modular curve $X_1(M)$.

Multiplicity one

The number of copies of $\overline{\rho}_{f,\lambda}$ in Theorem 1.15 is called the *multiplicity* of $\overline{\rho}_{f,\lambda}$. In general, let $f \in S_k(\Gamma_1(N))$ be a newform and λ is a prime of K_f such that $\overline{\rho} = \overline{\rho}_{f,\lambda}$ is absolutely irreducible. Then we define the multiplicity of $\overline{\rho}_{f,\lambda}$ as the multiplicity of its twist that is associated to a weight 2 form of minimal level. This multiplicity is equal to 1 in most cases, exceptions are only possible if a list of very strong conditions are satisfied.

Theorem 1.16 (Multiplicity one theorem, cf. [13, Theorem 6.1])**.** *Let* N *and* k *be positive integers and let* $f \in S_k(\Gamma_1(N))$ *be a newform. Furthermore, let* $\ell \nmid N$ *be a prime and suppose* $2 \leq k \leq \ell + 1$. *Take a prime* λ *of* K_f *above* ℓ *such that* $\overline{\rho} = \overline{\rho}_{f,\ell}$ *is an absolutely irreducible representation of multiplicity not equal to one. Then* k *is equal to* ℓ, *the representation* $\overline{\rho}$ *is unramified at* ℓ *and* $\overline{\rho}(\mathrm{Frob}_\ell)$ *is a scalar matrix.*

With some possible exceptions for $\ell = 2$, the converse of the theorem also holds; for a proof of this, see [87, Corollary 4.5]. For computational examples on representations of multiplicity not equal to one, see [41].

1.3.5 Examples

Let us give some examples of Galois representations associated to modular forms now. If one relaxes Theorem 1.13 a bit and does neither demand the representation to be irreducible nor the roots of $\rho(\mathrm{Frob}_p)$ to have absolute value $p^{(k-1)/2}$ then the Eisenstein series $G_k^{\psi,\phi}$ have Galois representations as well. From the q-expansion (1.8) of $G = G_k$ one can immediately read off that

$$\rho_{G,\ell} = \begin{pmatrix} \psi & 0 \\ 0 & \phi\chi_\ell^{k-1} \end{pmatrix}$$

is an ℓ-adic representation for G, where we denote a Dirichlet character and its associated $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-character by the same symbol.

The Ramanujan tau function

For the Ramanujan tau function, we also have representations. We are unable to write down the ℓ-adic ones so we'll display some of the mod ℓ representations for the tau function. The congruences for $\tau(n)$ described in Subsection 1.1.2 enable us to write down explicit representations $\overline{\rho}_{\Delta,\ell}$ for $\ell \in \{2,3,5,7,23,691\}$. For $\ell \in \{2,3,5,7,691\}$ they are reducible, for instance

$$\overline{\rho}_{\Delta,5} \sim \begin{pmatrix} \overline{\chi}_5 & 0 \\ 0 & \overline{\chi}_5^2 \end{pmatrix} \quad \text{and} \quad \overline{\rho}_{\Delta,691} \sim \begin{pmatrix} 1 & 0 \\ 0 & \overline{\chi}_{691}^{11} \end{pmatrix}.$$

For $\ell = 23$ we have to do a little more work to write it down. Consider the field $\mathbb{Q}(\sqrt{-23})$ and let H be its Hilbert class field. The field H is a splitting field of $x^3 - x - 1$ over $\mathbb{Q}$ and has Galois group $\mathrm{Gal}(H/\mathbb{Q}) \cong S_3$; we fix an isomorphism of these two groups. Consider the space $V \subset \mathbb{F}_{23}^3$ consisting of the vectors whose coordinates sum up to zero. The group S_3 acts on $\mathbb{F}_{23}^3$ by permuting the basis vectors and V is stable under this action. We claim that $\overline{\rho}_{\Delta,23}$ is the composition

$$\overline{\rho}_{\Delta,23} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Gal}(H/\mathbb{Q}) \cong S_3 \to \mathrm{GL}(V) \cong \mathrm{GL}_2(\mathbb{F}_{23}).$$

Indeed: primes p with $\left(\frac{p}{23}\right) = \left(\frac{-23}{p}\right) = -1$ are inert in $\mathbb{Q}(\sqrt{-23})$ so are sent to a transposition in S_3; transpositions have trace 0 in $\mathrm{GL}(V)$. Primes of the form $a^2 + 23b^2$ are known to split completely in H so are sent to the identity matrix which has trace 2. The other primes $p \neq 23$ have $\left(\frac{-23}{p}\right) = 1$ but do not split completely in H so must be sent to a 3-cycle which has trace -1.

Another interesting case is $\ell = 11$. From the above we know that we can obtain $\rho_{\Delta,11}$ as the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on a 2-dimensional subspace of $J_1(11)(\overline{\mathbb{Q}})[11]$. Because of

$g(X_1(11)) = 1$, the space $J_1(11)(\overline{\mathbb{Q}})[11]$ is 2-dimensional itself and $E := J_1(11)$ is an elliptic curve; a minimal Weierstrass equation for it is

$$E : y^2 - y = x^3 - x^2.$$

So $\overline{\rho}_{\Delta,11}$ is isomorphic to the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $E[11]$. In particular we have the congruence $\tau(p) \equiv p + 1 - \#E(\mathbb{F}_p) \bmod 11$ for $p \neq 11$. Schoof's algorithm [63] can be used to compute $\#E(\mathbb{F}_p) \bmod \ell$ efficiently for $p \neq 11$ and small $\ell \neq p$.

Remarks

Serre [64] has explained that the existence of simple congruences for $\tau(p)$ depends on what type of representation $\overline{\rho}_{\Delta,\ell}$ is. As already remarked, it is reducible for $\ell \in \{2,3,5,7,691\}$. Furthermore, it is *dihedral* for $\ell = 23$: a representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_q)$ is called dihedral if it is irreducible and over an algebraic closure of $\mathbb{F}_q$ its image is contained in a subgroup conjugate to $\left\{ \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \right\} \cup \left\{ \begin{pmatrix} 0 & * \\ * & 0 \end{pmatrix} \right\}$.

For all other primes ℓ, the representation $\overline{\rho}_{\Delta,\ell}$ is *non-exceptional*: a Galois representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_q)$ with $q = p^r$ is called exceptional if its image does not contain a subgroup conjugate to $\mathrm{SL}_2(\mathbb{F}_p)$. So $\overline{\rho}_{\Delta,\ell}$ is exceptional for $\ell \in \{2,3,5,7,23,691\}$ and for all other ℓ its image contains $\mathrm{SL}_2(\mathbb{F}_\ell)$. We have seen that reducible and dihedral representations are exceptional. These are not the only types of exceptional representations; there are also representations whose projective image is contained in a group isomorphic to the symmetry group of a regular polyhedron, but these do not occur very often. For more details on the exceptional representations for $\tau(p)$ and related functions, the reader is referred to [83] and [84].

1.4 Serre's conjecture

Let ℓ be a prime and let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ be an odd irreducible representation. Serre made the striking conjecture that such a ρ can always be obtained from a modular form, of a prescribed level and weight. In this section we will give the definitions for the level and the weight of the representation, which are called its *Serre invariants*; they depend on local properties of ρ. After this, we will formulate the conjecture, which is nowadays a theorem. The main reference for this material is [70]; other references include [27], [20], [42], [62] and [37].

1.4.1 Some local Galois theory

In this subsection we shall give some basic definitions from local Galois theory that we shall be using later on. However, to understand this material well, it is recommended to study [67], especially [67, Chapter IV].

Let K be a field that is complete with respect to a discrete valuation $v = v_K$, having perfect residue field κ. A field satisfying these conditions will be called a *local field* here. We also take as convention that discrete valuations map $K^\times$ surjectively to $\mathbb{Z}$. The ring of elements of a local field K with nonnegative valuation will be denoted by $\mathscr{O}_K$ and $\pi = \pi_K$ will denote a uniformiser of K.

Lower numbering

Let L/K be a finite Galois extension of local fields, with residue fields λ/κ. For $s \in [-1, +\infty[$, define the subgroups G_s and G_s^+ of $\mathrm{Gal}(L/K)$ as

$$G_s = \{\sigma \in \mathrm{Gal}(L/K) : v_L(\sigma\pi_L - \pi_L) \geq s+1\},$$
$$G_s^+ = \{\sigma \in \mathrm{Gal}(L/K) : v_L(\sigma\pi_L - \pi_L) > s+1\};$$

this does not depend on the choice of π_L. In particular, G_{-1} is equal to $\mathrm{Gal}(L/K)$ and G_{-1}/G_{-1}^+ is canonically isomorphic to $\mathrm{Gal}(\lambda/\kappa)$. If s is not an integer, then we have $G_s^+ = G_s$ and if s is an integer, we have $G_s^+ = G_{s+1}$.

The group G_0 is called the *inertia* subgroup of $\mathrm{Gal}(L/K)$ and is usually denoted by I. The group G_0^+ is called the *wild ramification* subgroup of $\mathrm{Gal}(L/K)$ and we denote usually by I_w. The wild ramification group can only be non-trivial if $p = \mathrm{char}(\kappa)$ is positive; in that case it is the unique Sylow p-subgroup of I. Also, G_0/G_0^+ is called the *tame ramification* or *tame inertia* subquotient of $\mathrm{Gal}(L/K)$ and is denoted by I_t.

We have an injective homomorphism

$$\theta_0 = \theta_0^{L/K} : I_t \hookrightarrow \mathscr{O}_L^\times/(1+\pi_L\mathscr{O}_L) \cong \lambda^\times, \quad \overline{\sigma} \mapsto \frac{\sigma\pi}{\pi} \bmod (1+\pi_L\mathscr{O}_L),$$

which is independent of the choice of a uniformiser of L. The group G_{-1}/G_{-1}^+ acts by conjugation on G_0/G_0^+; via θ_0, this action is compatible with the natural action of $\mathrm{Gal}(\lambda/\kappa)$ on $\lambda^\times$. To be more precise, for $\sigma \in G_{-1}$ and $\tau \in G_0$ the following formula holds:

$$\theta_0(\overline{\sigma\tau\sigma^{-1}}) = \sigma(\theta_0(\overline{\tau})), \tag{1.28}$$

where the action of G_{-1} on $\lambda^\times$ is the one that is obtained from the canonical isomorphism $G_{-1}/G_{-1}^+ \cong \mathrm{Gal}(\lambda/\kappa)$.

Upper numbering

Let again a finite Galois extension L/K of local fields be given and consider its lower numbering filtration. Define a function $\phi : [-1, +\infty[\to [-1, +\infty[$ by

$$\phi(s) = \int_0^s \frac{\#G_t}{\#G_0} dt.$$

This is a concave piecewise linear strictly increasing function. In particular it has an inverse, which we will call ψ. Now the upper numbering is defined by

$$G^s = G_{\psi(s)} \quad \text{and} \quad G^{s+} = G^+_{\psi(s)}.$$

The jumps in this filtration have rational index, not necessarily at integers. The real-valued indices allow us to use integrals in order to compactify a lot of notation. Note that for $s \in [-1,0]$ we have $G_s = G^s$ and $G^+_s = G^{s+}$.

If L is an infinite Galois extension of K, then we can still define an upper numbering on $\mathrm{Gal}(L/K)$: the upper numbering is compatible with taking Galois subfields, thus with taking quotients of Galois groups. Therefore, we can simply take projective limits to obtain an upper numbering $\mathrm{Gal}(L/K)^s$ and $\mathrm{Gal}(L/K)^{s+}$ that is compatible with taking finite Galois subextensions of L/K. In particular, we can speak of $I(L/K)$, $I_w(L/K)$ and $I_t(L/K)$

Tame characters

We will now restrict to the case $K = \mathbb{Q}_\ell$ and study the structure of the tame ramification group of $\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell$. For every finite Galois extension $L/\mathbb{Q}_\ell$ with residue field λ there is a canonical embedding $\theta_0^{L/\mathbb{Q}_\ell} : I_t(L/\mathbb{Q}_\ell) \hookrightarrow \lambda^\times$ as we saw above. If $M/L/\mathbb{Q}_\ell$ is a tower of Galois extensions with $\mu/\lambda/\mathbb{F}_\ell^\times$ the corresponding extensions of residue fields, then the diagram

$$\begin{array}{ccc} I_t(M/\mathbb{Q}_\ell) & \twoheadrightarrow & I_t(L/\mathbb{Q}_\ell) \\ \downarrow{\scriptstyle \theta_0} & & \downarrow{\scriptstyle \theta_0} \\ \mu^\times & \xrightarrow{\text{Norm}} & \lambda^\times \end{array}$$

commutes. If we put $L = \mathbb{Q}_\ell(\zeta_m, \sqrt[m]{\ell})$ with $m = \ell^n - 1$ then $\theta_0^{L/\mathbb{Q}_\ell}$ maps $I_t(L/\mathbb{Q}_\ell)$ isomorphically to $\mathbb{F}_{\ell^n}^\times$. This gives us an isomorphism

$$I_t(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \cong \varprojlim_n \mathbb{F}_{\ell^n}^\times,$$

where the maps in the projective system are the norm maps $\mathbb{F}_{\ell^n} \to \mathbb{F}_{\ell^m}$ for $m \mid n$.

Giving a character $\phi : I_t(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \overline{\mathbb{F}}_\ell^\times$ boils thus down to giving an n and a homomorphism of groups $\mathbb{F}_{\ell^n}^\times \to \overline{\mathbb{F}}_\ell^\times$. The smallest n that can be used here is called the *level* of ϕ. For a given n, exactly n of the homomorphisms $\mathbb{F}_{\ell^n}^\times \to \overline{\mathbb{F}}_\ell^\times$ come from field embeddings $\mathbb{F}_{\ell^n} \hookrightarrow \overline{\mathbb{F}}_\ell$; if a character $\psi : I_t(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \overline{\mathbb{F}}_\ell^\times$ can be given in this way, then we call ψ a *fundamental character* of level n.

Every character $\phi : I_t(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \overline{\mathbb{F}}_\ell^\times$ is a power of any fundamental character of the same level. The fundamental character of level 1 is the restriction of the mod ℓ cyclotomic $\overline{\chi}_\ell$ character to I, which we will abusively write as $\overline{\chi}_\ell$ as well.

Peu/très ramifiée

Let $L/\mathbb{Q}_\ell$ be a Galois extension whose wild ramification group I_w is killed by ℓ. Let $K \subset L$ be the maximal tamely ramified subextension, i.e. the fixed field of I_w and consider the extension $L(\zeta_\ell)/K(\zeta_\ell)$. By Kummer theory, there is a unique subgroup $A < K(\zeta_\ell)^\times/K(\zeta_\ell)^{\times\ell}$ with $L(\zeta_\ell) = K(\zeta_\ell)(\sqrt[\ell]{A})$. If A is a subgroup of $\mathscr{O}^\times_{K(\zeta_\ell)} \bmod K(\zeta_\ell)^{\times\ell}$ then we say that the extension $L/\mathbb{Q}_\ell$ is *peu ramifiée* and otherwise that it is *très ramifiée*. A representation ρ of $\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell)$ is called peu/très ramifiée if the field extension $\overline{\mathbb{Q}}_\ell^{\ker(\rho)}/\mathbb{Q}_\ell$ is.

1.4.2 The level

Let V be a finite dimensional vector space over $\overline{\mathbb{F}}_\ell$ and let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(V)$ be a representation. For a prime $p \neq \ell$ we consider the representation $\rho|_{D_p}$ of $G = \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$ and set

$$n(p,\rho) = \int_{-1}^{+\infty} \dim(V/V^{G^s})ds. \tag{1.29}$$

It is a non-trivial fact that $n(p,\rho)$ is a non-negative integer (cf. [67, Ch. VI]), equal to 0 for all but finitely many p. We define the *level* $N(\rho)$ of ρ as

$$N(\rho) := \prod_{p\neq\ell \,\text{prime}} p^{n(p,\rho)}. \tag{1.30}$$

The integer defined in this way is known as the prime-to-ℓ part of the Artin conductor of ρ.

We can also use the lower numbering to define the level. The field $K := \overline{\mathbb{Q}}^{\ker(\rho)}$ is a finite Galois extension of $\mathbb{Q}$ and the representation ρ factors through $\mathrm{Gal}(K/\mathbb{Q})$. Again, let $p \neq \ell$ be a prime and choose a prime $\mathfrak{p}$ of K above p. Then $G = \mathrm{Gal}(K_\mathfrak{p}/\mathbb{Q}_p)$ can be seen as a subgroup of $\mathrm{Gal}(K/\mathbb{Q})$. The formula (1.29) is equivalent to

$$n(p,\rho) = \sum_{i=0}^{\infty} \frac{\dim(V/V^{G_i})}{[G_0 : G_i]}.$$

In any case, we can read off from these formulas that $n(p,\rho) = 0$ if and only if ρ is unramified at p and $n(p,\rho) = \dim(V/V^I)$ if and only if ρ is (at most) tamely ramified at p.

This definition of level comes from Artin L-series. Let V be a finite-dimensional vector space over $\mathbb{C}$ and let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(V)$ be a representation. For each prime p, consider the subspace V^{I_p} of V. The action of Frob_p on V^{I_p} is well-defined up to conjugacy. We define the L-series of ρ to be

$$L(\rho,s) := \prod_{p\,\text{prime}} \det\left(1 - p^{-s}\rho(\mathrm{Frob}_p); V^{I_p}\right).$$

This series converges absolutely and uniformly in any right half plane $\{s \in \mathbb{C} : \Re(s) > 1+\delta\}$ with $\delta > 0$ and it has a meromorphic continuation to all of $\mathbb{C}$. Any complex conjugation in

$\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ will be sent to a matrix with eigenvalues equal to 1 and -1; let $n+$ and $n-$ their respective multiplicities. Define the completed L-series to be

$$\Lambda(\rho,s) := N(\rho)^{s/2}\Big(\pi^{s/2}\Gamma(\frac{s}{2})\Big)^{n+}\Big(\pi^{(s+1)/2}\Gamma(\frac{s+1}{2})\Big)^{n-}L(\rho,s),$$

where $N(\rho)$ is defined by the same formulas as above except that we don't exclude a prime called ℓ in the product (1.30). If we let ρ' be the representation obtained by composing ρ with complex conjugation in $\mathrm{GL}(V)$ then we have a functional equation

$$\Lambda(\rho,s) = W(\rho)\Lambda(\rho',1-s)$$

where $W(\rho) \in \mathbb{C}$ has absolute value 1. For details on these matters, the reader is referred to [56, Chapter VII].

1.4.3 The weight

The weight of ρ is defined in terms of $\rho|_{D_\ell}$. Serre's original definition [70, Section 2] differs slightly from Edixhoven's one in [27, Section 4]. The difference is due to the fact that Serre considers only classical modular forms, whereas Edixhoven considers the more geometric Katz modular forms. Spaces of Katz modular forms in positive characteristic can sometimes be bigger than their classical counterparts. Because of this, Serre avoids the cases $k = 1$ and odd k for $\ell = 2$. It is however true that those Katz modular forms can always be lifted to classical modular forms, but the weight may have to be adjusted.

A representation $\mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ can have several shapes and to define the weight it seems inevitable to do an investigation on the possible shapes that can occur. Using the fact that $\mathrm{im}(\rho|_{I_\ell})$ is an extension of a cyclic group by an ℓ-group one can show that $\rho|_{I_\ell}$ has to be reducible. It follows that $(\rho|_{I_\ell})^{\mathrm{ss}}$ the direct sum of two characters, which have to be tame as the order of the image is coprime to ℓ:

$$(\rho|_{I_\ell})^{\mathrm{ss}} \sim \begin{pmatrix} \phi & 0 \\ 0 & \phi' \end{pmatrix},$$

say. Using (1.28) one can show that either ϕ and ϕ' are both of level 1 or ϕ and ϕ' are both of level 2. To define the weight we will distinguish on these two cases, starting with the level 2 case as it has less subcases than the level 1 case.

The case that ϕ and ϕ' have level 2

From (1.28) it follows that ϕ and ϕ' are each others ℓ-th power and in fact that $\rho|_{D_\ell}$ is dihedral. If we choose a fundamental character ψ of level 2 then we can find $a,b \in \{0,\ldots,\ell-1\}$ with

$$\phi = \psi^{a+\ell b} \quad \text{and} \quad \phi' = \psi^{\ell a+b}.$$

We define the weight of ρ now as

$$k(\rho) := 1 + \ell\cdot\min(a,b) + \max(a,b).$$

Let us remark that choosing another ψ just exchanges a and b and furthermore that a and b are distinct as otherwise the level of ϕ and ϕ' would be 1.

The case that ϕ and ϕ' have level 1 and $\rho|_{I_\ell}$ is tamely ramified

In this case ϕ and ϕ' are powers of the cyclotomic character $\overline{\chi}_\ell$ and $\rho|_{I_\ell}$ is semi-simple, so we can write

$$\rho|_{I_\ell} \sim \begin{pmatrix} \overline{\chi}_\ell^a & 0 \\ 0 & \overline{\chi}_\ell^b \end{pmatrix}$$

with $a,b \in \{0,\dots,\ell-2\}$. There is a difference between the definitions of Serre and Edixhoven. Edixhoven puts

$$k(\rho) := 1 + \ell \cdot \min(a,b) + \max(a,b)$$

and Serre's definition is the same except for $a = b = 0$ where he puts $k(\rho) := \ell$.

The case that ϕ and ϕ' have level 1 and $\rho|_{I_\ell}$ is wildly ramified

Here, ϕ and ϕ' are again powers of $\overline{\chi}_\ell$, but $\rho|_{I_\ell}$ is not semi-simple. We write

$$\rho|_{I_\ell} \sim \begin{pmatrix} \overline{\chi}_\ell^a & * \\ 0 & \overline{\chi}_\ell^b \end{pmatrix},$$

with $a \in \{1,\dots,\ell-1\}$ and $b \in \{0,\dots,\ell-2\}$.

Suppose first that we have $a = b+1$ and $\rho|_{D_\ell}$ is très ramifiée over $\mathbb{Q}_\ell$. Then we have again a difference between Serre and Edixhoven. Edixhoven puts

$$k(\rho) := \ell + \ell \cdot \min(a,b) + \max(a,b)$$

and Serre's definition has one exception to Edixhoven's one: Serre puts $k(\rho) := 4$ in the case $\ell = 2$. In all other cases (i.e. if either $a \neq b+1$ holds or ρ is peu ramifiée at ℓ) the weight is defined by

$$k(\rho) := 1 + \ell \cdot \min(a,b) + \max(a,b).$$

Remarks

Sticking to Edixhoven's definitions, we have $1 \leq k(\rho) \leq \ell^2 - 1$ in all cases. We have $k(\rho) = 1$ if and only if ρ is unramified at ℓ. There is a twist $\rho \otimes \overline{\chi}_\ell^n$ of minimal weight. This minimal weight is at most $\ell+1$ and is called the *reduced weight* of ρ; it is denoted by $\tilde{k}(\rho)$. For a representation ρ that is wildly ramified at ℓ, an interesting theorem of Moon and Taguchi relates the reduced weight of ρ to the ℓ-part of the discriminant of the number field $\overline{\mathbb{Q}}^{\ker(\rho)}$:

Theorem 1.17 (Moon & Taguchi, [55, Theorem 3])**.** *Consider a wildly ramified representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$. Let $n \in \mathbb{Z}$ satisfy $\tilde{k} := \tilde{k}(\rho) = k(\rho \otimes \overline{\chi}_\ell^n)$. Define a number d by $d = \gcd(b, \tilde{k}-1, \ell-1)$ and define $m \in \mathbb{Z}$ by letting ℓ^m be the wild ramification degree of $K := \overline{\mathbb{Q}}_\ell^{\ker(\rho)}$ over $\mathbb{Q}_\ell$. Then we have*

$$v_\ell(\mathscr{D}_{K/\mathbb{Q}_\ell}) = \begin{cases} 1 + \frac{\tilde{k}-1}{\ell-1} - \frac{\tilde{k}-1+d}{(\ell-1)\ell^m} & \text{if } 2 \leq \tilde{k} \leq \ell, \\ 2 + \frac{1}{(\ell-1)\ell} - \frac{2}{(\ell-1)\ell^m} & \text{if } \tilde{k} = \ell+1, \end{cases}$$

where $\mathscr{D}_{K/\mathbb{Q}_\ell}$ denotes the different of K over $\mathbb{Q}_\ell$ and v_ℓ is normalised by $v_\ell(\ell) = 1$.

1.4.4 The conjecture

Let us now state the conjecture. It has a weak form and a strong form.

Conjecture 1.1 (Serre's conjecture, weak form, [70, Conjecture 3.2.3])**.** *Consider an odd irreducible representation* $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$. *Then there exists a newform* f *of some level and some weight, a prime* λ *of* K_f *above* ℓ *and an embedding* $\mathbb{F}_\lambda \hookrightarrow \overline{\mathbb{F}}_\ell$ *such that* $\rho \cong \overline{\rho}_{f,\lambda}$ *holds, where we view* $\overline{\rho}_{f,\lambda}$ *as a representation over* $\overline{\mathbb{F}}_\ell$ *via the embedding* $\mathbb{F}_\lambda \hookrightarrow \overline{\mathbb{F}}_\ell$.

Conjecture 1.2 (Serre's conjecture, strong form, [70, Conjecture 3.2.6])**.** *In the notation and statement of Conjecture 1.1 there exists an* f *of level dividing* $N(\rho)$ *and weight* $k(\rho)$.

It is a result of many people that the weak version is equivalent to the strong version; instead of compiling a complete list of names here, we refer to the overview article [42]. Serre's conjecture has been proven subsequently for level one in [38], for representations of odd level over fields of odd characteristic in [39] and finally in general in [43]. In all cases, the main ideas originate from the proof of the modularity theorem for elliptic curves by Taylor and Wiles [86].

Chapter 2

Computations with modular forms

In this chapter we will discuss several aspects of computations with modular forms. Let us warn the reader on beforehand that we will focus on how to compute in practice, not on theoretical aspects of computability. What in theory can be proven to be computable, can often not be computed in practice and what in practice can be computed, can often not be proven to be computable in theory.

2.1 Modular symbols

Modular symbols provide a way of doing symbolic calculations with modular forms, as well as the homology of modular curves. In this section as well, our intention is to give the reader an idea of what is going on rather than a complete and detailed account of the material. For more details and further reading on the subject of modular symbols, the reader could take a look at [51], [72] and [53]. A computational approach to the material can be found in [78] and [79].

2.1.1 Definitions

Let A be the free abelian group on the symbols $\{\alpha,\beta\}$ with $\alpha,\beta \in \mathbb{P}^1(\mathbb{Q})$. Consider the subgroup $I \subset A$ generated by all elements of the forms

$$\{\alpha,\beta\}+\{\beta,\gamma\}+\{\gamma,\alpha\}, \quad \{\alpha,\beta\}+\{\beta,\alpha\}, \quad \text{and} \quad \{\alpha,\alpha\}.$$

We define the group

$$\mathbb{M}_2 := (A/I)/\text{torsion}$$

as the quotient of A/I by its torsion subgroup. By a slight abuse of notation, we will denote the class of $\{\alpha,\beta\}$ in this quotient also by $\{\alpha,\beta\}$. We have an action $\mathrm{GL}_2^+(\mathbb{Q})$ on $\mathbb{M}_2$ by

$$\gamma\{\alpha,\beta\} := \{\gamma\alpha,\gamma\beta\},$$

where γ acts on $\mathbb{P}^1(\mathbb{Q})$ by fractional linear transformations.

For $k \geq 2$, we consider also the abelian group $\mathbb{Z}[x,y]_{k-2} \subset \mathbb{Z}[x,y]$ of homogeneous polynomials of degree $k-2$ and we let matrices in $\mathrm{GL}_2^+(\mathbb{Q})$ with integer coefficients act on it on the left by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} P(x,y) := P(dx - by, -cx + ay).$$

We define

$$\mathbb{M}_k := \mathbb{Z}[x,y]_{k-2} \otimes \mathbb{M}_2,$$

and we equip $\mathbb{M}_k$ with the component-wise action of integral matrices in $\mathrm{GL}_2^+(\mathbb{Q})$ (that is $\gamma(P \otimes \alpha) = \gamma(P) \otimes \gamma(\alpha)$).

Definition 2.1. Let $k \geq 2$ be an integer. Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a subgroup of finite index and let $I \subset \mathbb{M}_k$ be the subgroup generated by all elements of the form $\gamma x - x$ with $\gamma \in \Gamma$ and $x \in \mathbb{M}_k$. Then we define the space of *modular symbols* of weight k for Γ to be the quotient of M_k/I by its torsion subgroup and we denote this space by $\mathbb{M}_k(\Gamma)$:

$$\mathbb{M}_k(\Gamma) := (\mathbb{M}_k/I)/\text{torsion}.$$

In the special case $\Gamma = \Gamma_1(N)$, which we will mostly be interested in, $\mathbb{M}_k(\Gamma)$ is called the space of modular symbols of weight k and level N. The class of $\{\alpha,\beta\}$ in $\mathbb{M}_k(\Gamma)$ will be denoted by $\{\alpha,\beta\}_\Gamma$ or, if no confusion exists, by $\{\alpha,\beta\}$.

The group $\Gamma_0(N)$ acts naturally on $\mathbb{M}_k(\Gamma_1(N))$ and hence induces an action of $(\mathbb{Z}/N\mathbb{Z})^\times$ on $\mathbb{M}_k(\Gamma_1(N))$. We denote this action by the diamond symbol $\langle d \rangle$. The operator $\langle d \rangle$ on $\mathbb{M}_k(\Gamma_1(N))$ is called a *diamond operator*. This leads to the notion of modular symbols with character.

Definition 2.2. Let $\varepsilon : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ be a Dirichlet character. Denote by $\mathbb{Z}[\varepsilon] \subset \mathbb{C}$ the subring generated by all values of ε. Let $I \subset \mathbb{M}_k(\Gamma_1(N)) \otimes \mathbb{Z}[\varepsilon]$ be the $\mathbb{Z}[\varepsilon]$-submodule generated by all elements of the form $\langle d \rangle x - \varepsilon(d)x$ with $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and $x \in \mathbb{M}_k(\Gamma_1(N))$. Then we define the space $\mathbb{M}_k(N,\varepsilon)$ of modular symbols of weight k, level N and character ε as the $\mathbb{Z}[\varepsilon]$-module

$$\mathbb{M}_k(N,\varepsilon) := \big(\mathbb{M}_k(\Gamma_1(N)) \otimes \mathbb{Z}[\varepsilon]/I\big)/\text{torsion}.$$

We denote the elements of $\mathbb{M}_k(N,\varepsilon)$ by $\{\alpha,\beta\}_{N,\varepsilon}$ or simply by $\{\alpha,\beta\}$. If ε is trivial, then we have $\mathbb{M}_k(N,\varepsilon) \cong \mathbb{M}_k(\Gamma_0(N))$.

Let $\mathbb{B}_2$ be the free abelian group on the symbols $\{\alpha\}$ with $\alpha \in \mathbb{P}^1(\mathbb{Q})$ with action of $\mathrm{SL}_2(\mathbb{Z})$ by $\gamma\{\alpha\} = \{\gamma\alpha\}$ and define $\mathbb{B}_k := \mathbb{Z}[x,y]_{k-2} \otimes \mathbb{B}_2$ with component-wise $\mathrm{SL}_2(\mathbb{Z})$-action. Elements of $\mathbb{B}_k$ are called *boundary modular symbols*. For a subgroup $\Gamma < \mathrm{SL}_2(\mathbb{Z})$ of finite index, we define $\mathbb{B}_k(\Gamma)$ as

$$\mathbb{B}_k(\Gamma) := (\mathbb{B}_k/I)/\text{torsion}$$

where I is the subgroup of $\mathbb{B}_k$ generated by all elements $\gamma x - x$ with $\gamma \in \Gamma$ and $x \in \mathbb{B}_k$. We define $\mathbb{B}_k(N,\varepsilon)$ to be the quotient of $(\mathbb{B}_k(\Gamma_1(N)) \otimes \mathbb{Z}[\varepsilon])/I$ by its torsion submodule,

where I is the $\mathbb{Z}[\varepsilon]$-submodule of $\mathbb{B}_k(\Gamma_1(N)) \otimes \mathbb{Z}[\varepsilon]$ generated by the elements $\gamma x - \varepsilon(\gamma)x$ with $\gamma \in \Gamma_0(N)$.

We have *boundary homomorphisms* $\delta : \mathbb{M}_k(\Gamma) \to \mathbb{B}_k(\Gamma)$ and $\delta : \mathbb{M}_k(N,\varepsilon) \to \mathbb{B}_k(N,\varepsilon)$, defined by

$$\delta\left(P \otimes \{\alpha,\beta\}\right) = P \otimes \{\beta\} - P \otimes \{\alpha\}.$$

The spaces of *cuspidal modular symbols*, denoted by $\mathbb{S}_k(\Gamma)$ and $\mathbb{S}_k(N,\varepsilon)$ respectively are defined as the kernel of δ.

2.1.2 Properties

One can interpret the symbol $\{\alpha,\beta\}$ as a smooth path in $\mathfrak{H}^*$ from the cusp α to the cusp β, lying in $\mathfrak{H}$ except for the endpoints α and β. It can be shown that this interpretation induces an isomorphism

$$\mathbb{M}_2(\Gamma) \cong H_1\left(X_\Gamma, \text{cusps}, \mathbb{Z}\right).$$

Here the homology is taken of the topological pair $(X_1(N), \text{cusps})$. We also get an isomorphism

$$\mathbb{S}_2(\Gamma) \cong H_1\left(X_\Gamma, \mathbb{Z}\right).$$

So we immediately see that there is a perfect pairing

$$\left(\mathbb{S}_2(\Gamma(N)) \otimes \mathbb{C}\right) \times \left(S_2(\Gamma(N)) \oplus \overline{S}_2(\Gamma(N))\right) \to \mathbb{C}$$

defined by

$$\left(\{\alpha,\beta\}, f \oplus g\right) \mapsto \int_\alpha^\beta \left(f\frac{dq}{q} + g\frac{d\overline{q}}{\overline{q}}\right).$$

More generally, there is a pairing

$$\mathbb{M}_k(\Gamma_1(N)) \times \left(S_k(\Gamma_1(N)) \oplus \overline{S}_k(\Gamma_1(N))\right) \to \mathbb{C} \tag{2.1}$$

defined by

$$\left(P \otimes \{\alpha,\beta\}, f \oplus g\right) \mapsto 2\pi i \int_\alpha^\beta \left(f(z)P(z,1)dz - g(z)P(\bar{z},1)d\bar{z}\right),$$

which becomes perfect if we restrict and tensor the left factor to $\mathbb{S}_k(\Gamma(N)) \otimes \mathbb{C}$. This pairing induces a pairing

$$\left(\mathbb{M}_k(N,\varepsilon)\right) \times \left(S_k(N,\varepsilon) \oplus \overline{S}_k(N,\varepsilon)\right) \to \mathbb{C}$$

which is perfect when the left factor is restricted and tensored to $\mathbb{S}_k(N,\varepsilon) \otimes_{\mathbb{Z}[\varepsilon]} \mathbb{C}$. From now on we will denote all these pairings with the notation

$$(x, f) \mapsto \langle x, f \rangle.$$

The star involution

On the spaces $\mathbb{M}_k(\Gamma_1(N))$ and $\mathbb{M}_k(N,\varepsilon)$ we have an involution ι^* defined by

$$\iota^*(P(x,y)\otimes\{\alpha,\beta\}) := -P(x,-y)\otimes\{-\alpha,-\beta\},$$

which is called the *star involution*. It preserves cuspidal subspaces. We define $\mathbb{S}_k(\Gamma_1(N))^+$ and $\mathbb{S}_k(\Gamma_1(N))^-$ subspaces of $\mathbb{S}_k(\Gamma_1(N))$ where ι^* acts as $+1$ and -1 respectively and we use similar definitions for $\mathbb{S}_k(N,\varepsilon)^\pm$. It can be shown that the pairing (2.1) induces perfect pairings

$$(\mathbb{S}_k(\Gamma_1(N))^+\otimes\mathbb{C})\times S_k(\Gamma_1(N))\to\mathbb{C}$$

and

$$(\mathbb{S}_k(\Gamma_1(N))^-\otimes\mathbb{C})\times \overline{S}_k(\Gamma_1(N))\to\mathbb{C}$$

and similarly for the spaces with character. This allows us to work sometimes in modular symbols spaces of half the dimension of the full cuspidal space.

2.1.3 Hecke operators

Hecke operators on modular symbols are defined in a similar way as on modular forms (see Subsection 1.1.4). Let $k\geq 2$ and $N\geq 1$ be given. Then for $\gamma\in \mathrm{GL}_2^+(\mathbb{Q})\cap \mathrm{M}_2(\mathbb{Z})$ we define an operator T_γ on $\mathbb{M}_k(\Gamma_1(N))$ by letting $\gamma_1,\ldots,\gamma_r$ be double coset representatives for $\Gamma_1(N)\setminus\Gamma_1(N)\gamma\Gamma_1(N)$ and putting

$$T_\gamma(x) := \sum_{i=1}^{r}\gamma_i x \quad \text{for } x\in\mathbb{M}_k(\Gamma_1(N)). \tag{2.2}$$

It follows from [72, Theorem 4.3] that this operator is well-defined. For a prime number p we put $T_p = T_\gamma$ for $\gamma = \left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)$ and for positive integers n we define T_n by means of the relations (1.13). The operators T_n are called Hecke operators.

The Hecke operators preserve the subspace $\mathbb{S}_k(\Gamma_1(N))$ and induce an action on the spaces $\mathbb{M}_k(N,\varepsilon)$ and $\mathbb{S}_k(N,\varepsilon)$. Furthermore, from [72, Theorem 4.3] one can conclude that the diamond and Hecke operators are self-adjoint with respect to the pairings defined in the previous subsection:

$$\langle Tx, f\rangle = \langle x, Tf\rangle. \tag{2.3}$$

for any modular symbol x, cusp form f and diamond or Hecke operator T for which this relation is well-defined. Furthermore, the Hecke operators commute with the star involution ι^*.

In conclusion, we see how we can write cusp forms spaces as the dual of modular symbols spaces. The computation of Hecke operators on these modular symbols spaces would enable us to compute q-expansions of cusp forms: q-coefficients of newforms can be computed once we can compute the eigenvalues of Hecke operators. But because of (2.3) this reduces to the computation of the eigenvalues of Hecke operators on modular symbols spaces. In computations one often works with the spaces $\mathbb{S}_k(N,\varepsilon)^+\otimes_{\mathbb{Z}[\varepsilon]}\mathbb{Q}(\varepsilon)$ because these have smaller

dimension than $\mathbb{S}_k(\Gamma_1(N)) \otimes \mathbb{Q}$. Since we also know how all cusp forms arise from newforms of possibly lower level (see Theorem 1.5), this allows us to compute the q-expansions of a basis for the spaces $S_k(\Gamma_1(N))$ and $S_k(N,\varepsilon)$. For precise details on how these computations work, please read [79, Chapter 9].

2.1.4 Manin symbols

If we want to do symbolic calculations with modular symbols, then the above definitions are not quite applicable since the groups of which we take quotients are not finitely generated. The *Manin symbols* enable us to give finite presentations for the spaces of modular symbols.

First we need some definitions and lemmas. For a positive integer N we define a set

$$E_N := \left\{(c,d) \in (\mathbb{Z}/N\mathbb{Z})^2 : \gcd(N,c,d) = 1\right\}.$$

Define the following equivalence relation on E_N:

$$(c,d) \sim (c',d') \quad \stackrel{\text{def}}{\iff} \quad \text{there is an } a \in (\mathbb{Z}/N\mathbb{Z})^\times \text{ such that } (c,d) = (ac',ad')$$

and the denote the quotient by P_N:

$$P_N := E_N/\sim. \tag{2.4}$$

The following lemma is easily verified:

Lemma 2.1. *Let N be a positive integer. Then the maps*

$$\Gamma_1(N)\setminus \mathrm{SL}_2(\mathbb{Z}) \to E_N : \overline{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \mapsto (\overline{c},\overline{d}) \quad \textit{and}$$
$$\Gamma_0(N)\setminus \mathrm{SL}_2(\mathbb{Z}) \to P_N : \overline{\begin{pmatrix} a & b \\ c & d \end{pmatrix}} \mapsto \overline{(c,d)}$$

are well-defined and bijective.

This lemma enables us to write down an explicit set of coset representatives for the orbit spaces $\Gamma_1(N)\setminus \mathrm{SL}_2(\mathbb{Z})$ and $\Gamma_0(N)\setminus \mathrm{SL}_2(\mathbb{Z})$. The following lemma provides us a first step in reducing the set of generators for the spaces of modular symbols:

Lemma 2.2. *Each space $\mathbb{M}_2(\Gamma_1(N))$ or $\mathbb{M}_2(N,\varepsilon)$ is generated by the symbols $\{a/c, b/d\}$ with $a,b,c,d \in \mathbb{Z}$ and $ad-bc=1$, where in this notation a fraction with denominator equal to zero denotes the cusp at infinity.*

Calculating the continued fraction expansion at each cusp in $\mathbb{Q}$ gives us immediately an algorithm to write a given element of $\mathbb{M}_2$ in terms of the generators in the lemma. Furthermore, note that

$$\left\{\frac{a}{c},\frac{b}{d}\right\} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\{\infty,0\},$$

so that we can write each element of $\mathbb{M}_2$ as a sum of $\gamma\{\infty,0\}$ with $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

Let's consider the space $\mathbb{M}_2(\Gamma_1(N))$. As we saw, it is generated by the elements $\gamma\{\infty,0\}$ where γ runs through $\mathrm{SL}_2(\mathbb{Z})$. Now, two matrices γ define the same element this way if they are in the same coset of the quotient $\Gamma_1(N)\setminus \mathrm{SL}_2(\mathbb{Z})$. According to Lemma 2.1 such a coset can be uniquely identified with a pair $(c,d) \in (\mathbb{Z}/N\mathbb{Z})^2$. The corresponding element in $\mathbb{M}_2(\Gamma_1(N))$ is also denoted by (c,d). This element (c,d) is called a *Manin symbol*. Clearly, there are only a finite number of Manin symbols so we now know a finite set of generators for $\mathbb{M}_2(\Gamma_1(N))$.

For arbitrary k we define the Manin symbols in $\mathbb{M}_k(\Gamma_1(N))$ as the symbols of the form $P\otimes(c,d)$ where P is a monomial in $\mathbb{Z}[x,y]_{k-2}$ and (c,d) a Manin symbol in $\mathbb{M}_2(\Gamma_1(N))$. In this case as well there are finitely many Manin symbols and they generate the whole space.

In the modular symbols spaces with character ε, we have $\gamma(\alpha) = \varepsilon(\alpha)$ for $\gamma \in \Gamma_0(N)$. Now for each element of P_N we choose according to Lemma 2.1 a corresponding element $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and hence an element in $\mathbb{M}_2(N,\varepsilon)$, which we call again a Manin symbol. Note that this Manin symbol depends on the choice of γ, but because of the relation $\gamma(x) = \varepsilon(x)$ these chosen Manin symbols always form a finite set of generators for $\mathbb{M}_2(N,\varepsilon)$ as a $\mathbb{Z}[\varepsilon]$-module. Likewise, $\mathbb{M}_k(N,\varepsilon)$ is generated by elements $P\otimes(c,d)$ with P a monomial in $\mathbb{Z}[x,y]_{k-2}$ and (c,d) a Manin symbol in $\mathbb{M}_2(N,\varepsilon)$.

If we want to do symbolic calculations, then besides generators we also need to know the relations between the Manin symbols. For $\mathbb{M}_k(\Gamma_1(N))$ one can do the following.

Proposition 2.1. *Let N be a positive integer and let A be the free abelian group on the Manin symbols of the space $\mathbb{M}_k(\Gamma_1(N))$. Let $I \subset A$ be the subgroup generated by the following elements:*

$$\begin{aligned}
&P(x,y)\otimes(c,d) + P(-y,x)\otimes(-d,-c),\\
&P(x,y)\otimes(c,d) + P(-y,x-y)\otimes(-d,-c-d) + P(-x+y,-x)\otimes(-c-d,-c),\\
&P(x,y)\otimes(c,d) - P(-x,-y)\otimes(c,d),
\end{aligned}$$

where $P(x,y)\otimes(c,d)$ runs through all Manin symbols. Then $\mathbb{M}_k(\Gamma_1(N))$ is naturally isomorphic to the quotient of A/I by its torsion subgroup.

For the modular symbols spaces $\mathbb{M}_k(N,\varepsilon)$ we have a similar proposition.

Proposition 2.2. *Let N and ε be given. Let A be the free $\mathbb{Z}[\varepsilon]$-module on the Manin symbols of $\mathbb{M}_k(N,\varepsilon)$. Let $I \subset A$ be the submodule generated by the elements given in Proposition 2.1 plus for each $n \in (\mathbb{Z}/N\mathbb{Z})^\times$ the elements*

$$P(x,y)\otimes\overline{(nc,nd)} - \varepsilon(n)P(x,y)\otimes\overline{(c,d)}.$$

Then $\mathbb{M}_k(N,\varepsilon)$ is naturally isomorphic to the quotient of A/I by its torsion submodule.

These presentations enable us to perform symbolic calculations very efficiently.

A remark on the computation of Hecke operators is in order here. The formula (2.2) does not express the Hecke action on Manin symbols in terms of Manin symbols. Instead, one uses other formulas to compute Hecke operators. The following theorem, due to Merel, allows us to express Hecke operators more directly in terms of Manin symbols:

Theorem 2.1 (see [53, Theorem 2])**.** *On the spaces $\mathbb{M}_k(\Gamma_1(N))$ and $\mathbb{M}_k(N,\varepsilon)$ the Hecke operator T_n satisfies the following relation:*

$$T_n(P(x,y)\otimes(u,v)) = \sum_{\substack{a>b\geq 0\\ d>c\geq 0\\ ad-bc=n}}{}' P(ax+by,cx+dy)\otimes(au+cv,bu+dv),$$

where the prime in the sum notation means that terms with $\gcd(N,au+cv,bu+dv)\neq 1$ *have to be omitted.*

One would also like to express $\mathbb{S}_k(\Gamma_1(N))$ and $\mathbb{S}_k(N,\varepsilon)$ in terms of the Manin symbols. The following proposition will help us.

Proposition 2.3 (See [53, Proposition 4])**.** *Let integers $N\geq 1$ and $k\geq 2$ be given. Define an equivalence relation on the vector space $\mathbb{Q}[\Gamma_1(N)\setminus\mathbb{Q}^2]$ by*

$$[\overline{\lambda x}]\sim \operatorname{sign}(\lambda)^k[\overline{x}] \quad \textit{for } \lambda\in\mathbb{Q}^\times \textit{ and } x\in\mathbb{Q}^2.$$

Then the map

$$\mu:\mathbb{B}_k(\Gamma_1(N))\to\mathbb{Q}[\Gamma_1(N)\setminus\mathbb{Q}^2]/\sim$$

given by

$$\mu: P\otimes\left\{\frac{a}{b}\right\}\mapsto P(a,b)\left[\overline{\binom{a}{b}}\right] \quad \textit{(a,b coprime integers)}$$

is well-defined and injective.

The vector space $\mathbb{Q}[\Gamma_1(N)\setminus\mathbb{Q}^2]/\sim$ is finite dimensional. The above proposition shows that $\mathbb{S}_k(\Gamma_1(N))$ is the kernel of $\mu\delta$, which is a map that can be computed in terms of Manin symbols. The computation of $\mathbb{S}_k(N,\varepsilon)$ can be done in a similar way, see [79, Section 8.4].

2.2 Basic numerical evaluations

In this section we will describe how to perform basic numerical evaluations, such as the evaluation of a cusp form at a point in $\mathfrak{H}$ and the evaluation of an integral of a cusp form between to points in $\mathfrak{H}^*$. Again, the paradigm will be performing actual computations.

2.2.1 Period integrals: the direct method

In this subsection we will stick to the case $k = 2$, referring to [79, Chapter 10] for a more general approach (see also [18, Section 2.10] for a treatment of $\Gamma_0(N)$). So fix a positive integer N and an $f \in S_2(\Gamma_1(N))$. Our goal is to efficiently evaluate $\langle x, f\rangle$ for $x \in \mathbb{S}_2(\Gamma_1(N))$.

Let us indicate why it suffices to look at newforms f. Because of Theorem 1.5, it suffices to look at $f = \alpha_d(f')$ with $f' \in S_k(\Gamma_1(M))$ a newform for some $M \mid N$ and $d \mid N/M$. By [72, Theorem 4.3] we have

$$\langle x, f\rangle = \langle x, \alpha_d(f')\rangle = d^{1-k}\left\langle \begin{pmatrix} d & 0 \\ 0 & 1 \end{pmatrix} x, f' \right\rangle$$

so that computing period integrals for f reduces to computing period integrals of the newform f'.

Let us now make the important remark that for each $z \in \mathfrak{H}$ we can numerically compute $\int_\infty^z f dq/q$ by formally integrating the q-expansion of f:

$$\int_\infty^z f\frac{dq}{q} = \sum_{n\geq 1} \frac{a_n(f)}{n} q^n \quad \text{where } q = \exp(2\pi i z). \tag{2.5}$$

The radius of convergence of this series is 1 and the coefficients are small (that is, estimated by $\tilde{O}(n^{(k-3)/2})$). So if $\Im z \gg 0$ then we have $|q| \ll 1$ and the series converges rapidly. To be more concrete, for $\Im z > M$ we have $|q^n| < \exp(-2\pi M n)$ so if we want to compute $\int_\infty^z f dq/q$ to a precision of p decimals, we need to compute about $\frac{p\log 10}{2\pi M} \approx 0.37\frac{p}{M}$ terms of the series.

To compute a period integral we remark that for any $\gamma \in \Gamma_1(N)$ and any $z \in \mathfrak{H}^*$ any continuous, piecewise smooth path δ in $\mathfrak{H}^*$ from z to γz, the homology class of δ pushed forward to $X_1(N)(\mathbb{C})$ depends only on γ [51, Proposition 1.4]. Let us denote this homology class by

$$\{\infty, \gamma\infty\} \in \mathbb{S}_2(\Gamma_1(N)) \cong H_1(X_1(N)(\mathbb{C}), \mathbb{Z})$$

and remark that all elements of $H_1(X_1(N)(\mathbb{C}), \mathbb{Z})$ can be written in this way. As we also have $S_2(\Gamma_1(N)) \cong H^0(X_1(N)_{\mathbb{C}}, \Omega^1)$, this means we can calculate $\int_{\{\infty,\gamma\infty\}} f\frac{dq}{q}$ by choosing a smart path in $\mathfrak{H}^*$:

$$\int_\infty^{\gamma\infty} f\frac{dq}{q} = \int_z^{\gamma z} f\frac{dq}{q} = \int_\infty^{\gamma z} f\frac{dq}{q} - \int_\infty^z f\frac{dq}{q}.$$

If we write $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ then a good choice for z is

$$z = -\frac{d}{c} + \frac{i}{|c|}.$$

In this case we have $\Im z = \Im \gamma z = 1/|c|$ so in view of (2.5), to compute the integral to a precision of p decimals we need about $\frac{pc\log 10}{2\pi} \approx 0.37pc$ terms of the series.

Another thing we can use is the Hecke compatibility from (2.3). Put

$$W_f := \big(\mathbb{S}_2(\Gamma_1(N))/I_f\mathbb{S}_2(\Gamma_1(N))\big)\otimes\mathbb{Q},$$

where I_f is the Hecke ideal belonging to f. The space W_f has the structure of a vector space over $\mathbb{T}/I_f \cong K_f$ of dimension 2. This means that computing any period integral of f, we only need to precompute 2 period integrals. So one tries to find a K_f-basis of W_f consisting of elements $\{\infty,\gamma\infty\}$ where $\gamma\in\Gamma_1(N)$ has a very small c-entry. In practice it turns out that we do not need to search very far.

2.2.2 Period integrals: the twisted method

In this subsection we have the same set-up as in the previous subsection. There is another way of computing period integrals for $f\in S_2(\Gamma_1(N))$ which sometimes beats the method described in the previous subsection. The method described in this subsection is similar to [18, Section 2.11] and makes use of winding elements and twists.

The *winding element* of $\mathbb{M}_2(\Gamma_1(N))$ is simply defined as the element $\{\infty,0\}$ (some authors define it as $\{0,\infty\}$ but this is only a matter of sign convention). Integration over this element is easy to perform because we can break up the path in a very neat way:

$$\begin{aligned}\int_\infty^0 f\frac{dq}{q} &= \int_\infty^{i/\sqrt{N}} f\frac{dq}{q} + \int_{i/\sqrt{N}}^0 f\frac{dq}{q} = \int_\infty^{i/\sqrt{N}} f\frac{dq}{q} + \int_{i/\sqrt{N}}^\infty W_N(f)\frac{dq}{q}\\ &= \int_\infty^{i/\sqrt{N}} (f-W_N(f))\frac{dq}{q}.\end{aligned}$$

Now, choose an odd prime $\ell\nmid N$ and a primitive Dirichlet character $\chi:\mathbb{Z}\to\mathbb{C}$ of conductor ℓ. If $f\in S_k(\Gamma_1(N))$ is a newform then $f\otimes\chi$ is a newform in $S_k(\Gamma_1(N\ell^2))$, where

$$f\otimes\chi = \sum_{n\geq 1} a_n(f)\chi(n)q^n.$$

The following formula to express χ as a linear combination of additive characters is well-known:

$$\chi(n) = \frac{g(\chi)}{\ell}\sum_{\nu=1}^{\ell-1}\overline{\chi}(-\nu)\exp\left(\frac{2\pi i\nu n}{\ell}\right),$$

where $g(\chi)$ is the Gauss sum of χ (see (1.7)). It follows now immediately that

$$f\otimes\chi = \frac{g(\chi)}{\ell}\sum_{\nu=1}^{\ell-1}\chi(-\nu)f\left(z+\frac{\nu}{\ell}\right) = \frac{g(\chi)}{\ell}\sum_{\nu=1}^{\ell-1}\chi(-\nu)f\left|\begin{pmatrix}\ell & \nu\\ 0 & \ell\end{pmatrix}\right.. \tag{2.6}$$

For $f\in S_2(\Gamma_1(N))$ we now get the following useful formula for free:

$$\langle\{\infty,0\},f\otimes\chi\rangle = \frac{g(\chi)}{\ell}\left\langle\sum_{\nu=0}^{l-1}\chi(-\nu)\left\{\infty,\frac{\nu}{\ell}\right\},f\right\rangle. \tag{2.7}$$

The element $\sum_{v=0}^{l-1}\chi(-v)\left\{\infty,\frac{v}{\ell}\right\}$ of $\mathbb{M}_k(\Gamma_1(N))\otimes\mathbb{Z}[\chi]$ or of some other modular symbols space where it is well-defined is called a *twisted winding element* or, more precisely the χ*-twisted winding element*. Because of formula (2.7), we can calculate the pairings of newforms in $S_2(\Gamma_1(N))$ with twisted winding elements quite efficiently as well.

We can describe the action of the Atkin-Lehner operator $W_{N\ell^2}$ on $f\otimes\chi$:

$$W_{N\ell^2}(f\otimes\chi)=\frac{g(\chi)}{g(\overline{\chi})}\varepsilon(\ell)\chi(-N)\lambda_N(f)\tilde{f}\otimes\overline{\chi},$$

where $\tilde{f}=\sum_{n\geq 1}\overline{a_n(f)}q^n$ (see for example [3, Section 3]). So in particular we have the following integral formula for a newform $f\in S_2(N,\varepsilon)$:

$$\begin{aligned}\int_\infty^0 f\otimes\chi\frac{dq}{q}&=\int_\infty^{i/(\ell\sqrt{N})}(f\otimes\chi-W_{N\ell^2}(f\otimes\chi))\frac{dq}{q}\\&=\int_\infty^{i/(\ell\sqrt{N})}\left(f\otimes\chi-\frac{g(\chi)}{g(\overline{\chi})}\chi(-N)\varepsilon(\ell)\lambda_N(f)\tilde{f}\otimes\overline{\chi}\right)\frac{dq}{q}.\end{aligned}\tag{2.8}$$

So to calculate

$$\left\langle\sum_{v=0}^{l-1}\chi(-v)\left\{\infty,\frac{v}{\ell}\right\},f\right\rangle$$

we need to evaluate the series (2.5) at z with $\Im z=1/(\ell\sqrt{N})$ which means that for a precision of p decimals we need about $\frac{p\ell\sqrt{N}\log 10}{2\pi}\approx 0.37p\ell\sqrt{N}$ terms of the series. In the spirit of the previous subsection, we try several ℓ and χ, as well as the untwisted winding element $\{\infty,0\}$, until we can make a K_f-basis for W_f. It follows from [71, Theorems 1 and 3] that we can always find such a basis. Also here, it turns out that in practice we do not need to search very far.

2.2.3 Computation of q-expansions at various cusps

The upper half plane $\mathfrak{H}$ is covered by neighbourhoods of the cusps. If we want to evaluate a cusp form $f\in S_k(\Gamma_1(N))$ or an integral of a cusp form at a point in such a neighbourhood then it is useful to be able to calculate the q-expansion of f at the corresponding cusp. We shall mean by this the following: A cusp a/c can be written as $\gamma\infty$ with $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\mathrm{SL}_2(\mathbb{Z})$. Then a q-expansion of f at a/c is simply the q-expansion of $f|_k\gamma$. This notation is abusive, since it depends on the choice of γ. The q-expansion will be an element of the power series ring $\mathbb{C}[[q^{1/w}]]$ where w is the width of the cusp a/c and $q^{1/w}=\exp(2\pi iz/w)$.

If the level N is square-free this can be done symbolically. However for general N it is not known how to do this but we shall give some attempts that do at least give numerical computations of q-expansions. We use that we can compute the q-expansions of newforms in $S_k(\Gamma_1(N))$ at ∞ using modular symbols methods.

The case of square-free N

The method we present here is due to Asai [2]. Let N be square-free and let $f \in S_k(\Gamma_1(N))$ be a newform of character ε. The main reason for being able to compute q-expansions at all cusps in this case is because the group generated by $\Gamma_0(N)$ and all w_Q (see (1.18)) acts transitively on the cusps.

So let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ be given. Put

$$c' = \frac{c}{\gcd(N,c)}, \quad \text{and} \quad Q = \frac{N}{\gcd(N,c)}.$$

Let $r \in \mathbb{Z}$ be such that $d \equiv cr \bmod Q$ and define $b', d' \in \mathbb{Z}$ by

$$Qd' = d - cr \quad \text{and} \quad b' = b - ar.$$

Then we have

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} Qa & b' \\ Nc' & Qd' \end{pmatrix} \begin{pmatrix} Q^{-1} & rQ^{-1} \\ 0 & 1 \end{pmatrix}.$$

We know how $\left(\begin{smallmatrix} Qa & b' \\ Nc' & Qd' \end{smallmatrix}\right)$ acts on q-expansions by Theorems 1.6 and 1.8. The action of $\left(\begin{smallmatrix} Q^{-1} & rQ^{-1} \\ 0 & 1 \end{smallmatrix}\right)$ on q-expansions is simply

$$\sum_{n\geq 1} a_n q^n \mapsto Q^{1-k} \sum_{n\geq 1} a_n \zeta_Q^{rn} q^{n/Q} \quad \text{with } \zeta_Q = \exp(\tfrac{2\pi i}{Q}).$$

This shows how the q-expansion of $f|_k\gamma$ can be derived from the q-expansion of f.

Let us now explain how to do it for oldforms as well. By induction and Theorem 1.5 we may suppose $f = \alpha_p(f')$ with $p \mid N$ prime, $f' \in S_k(\Gamma_1(N/p))$ and that we know how to compute the q-expansions of f' at all the cusps. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be given. Then we have

$$f|_k\gamma = p^{1-k} f'\big|_k \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \gamma = p^{1-k} f'\big|_k \begin{pmatrix} pa & pb \\ c & d \end{pmatrix}.$$

We will now distinguish on two cases: $p \mid c$ and $p \nmid c$. If $p \mid c$ then we have a decomposition

$$\begin{pmatrix} pa & pb \\ c & d \end{pmatrix} = \begin{pmatrix} a & pb \\ c/p & d \end{pmatrix} \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$$

and we know how both matrices on the right hand side act on q-expansions. If $p \nmid c$, choose b', d' with $pad' - b'c = 1$. Then we have

$$\begin{pmatrix} pa & pb \\ c & d \end{pmatrix} = \begin{pmatrix} pa & b' \\ c & d' \end{pmatrix} \beta$$

with $\beta \in \mathrm{GL}_2^+(\mathbb{Q})$ upper triangular, so also in this case we know how both matrices on the right hand side act on q-expansions.

The general case

In a discussion with Peter Bruin, the author figured out an attempt to drop the assumption that N be square-free and compute q-expansions of cusp forms numerically in this case. The idea is to generalise the W_Q operators from Subsection 1.1.7.

So let N be given. Let Q be a divisor of N and put $R = \gcd(Q, N/Q)$. Let w_Q be any matrix of the form

$$w_Q = \begin{pmatrix} RQa & b \\ RNc & Qd \end{pmatrix} \quad \text{with } a,b,c,d \in \mathbb{Z}$$

such that $\det w_Q = QR^2$ (the conditions guarantee us that such matrices do exist). One can then verify

$$\Gamma_1(NR^2) < w_Q^{-1}\Gamma_1(N)w_Q,$$

so that slashing with w_Q defines a linear map

$$S_k(\Gamma_1(N)) \oplus \overline{S}_k(\Gamma_1(N)) \xrightarrow{|w_Q} S_k(\Gamma_1(NR^2)) \oplus \overline{S}_k(\Gamma_1(NR^2))$$

which is injective since the slash operator defines a group action on the space of all functions $\mathfrak{H} \to \mathbb{C}$.

On the other hand, w_Q defines an operation on $\mathbb{M}_k$ which can be shown to induce a linear map

$$w_Q : \mathbb{S}_k(\Gamma_1(NR^2)) \otimes \mathbb{Q} \to \mathbb{S}_k(\Gamma_1(N)) \otimes \mathbb{Q}$$

that satisfies the following compatibility with respect to the integration pairing between modular symbols and cusp forms (see [72, Theorem 4.3]):

$$\langle w_Q x, f \rangle = \langle x, f|_k w_Q \rangle. \tag{2.9}$$

Let $(x_1, \ldots, x_r)$ and $(y_1, \ldots, y_s)$ be bases of $\mathbb{S}_k(\Gamma_1(N)) \otimes \mathbb{Q}$ and $\mathbb{S}_k(\Gamma_1(NR^2)) \otimes \mathbb{Q}$ respectively. Then one can write down a matrix A in terms of these basis that describes the map w_Q since we can express any symbol $P \otimes \{\alpha, \beta\}$ in terms of Manin symbols. The matrix A^t then defines the action of w_Q in terms of the bases of the cusp forms spaces that are dual to $(x_1, \ldots x_r)$ and $(y_1, \ldots, y_s)$.

Now, let $(f_1, \ldots, f_r)$ be a basis of $S_k(\Gamma_1(N)) \oplus \overline{S}_k(\Gamma_1(N))$ and let $(g_1, \ldots, g_s)$ be a basis of $S_k(\Gamma_1(NR^2)) \oplus \overline{S}_k(\Gamma_1(NR^2))$ (for instance we could take bases consisting of eigenforms for the Hecke operators away from N). Define matrices

$$B := \left(\langle x_i, f_j \rangle\right)_{i,j} \quad \text{and} \quad C := \left(\langle y_i, g_j \rangle\right)_{i,j}.$$

These can be computed numerically as the entries are period integrals. Then the matrix $C^{-1}A^tB$ describes the map $\cdot|_k w_Q$ in terms of the bases $(f_1, \ldots, f_r)$ and $(g_1, \ldots, g_s)$. Hence if we can invert C efficiently, then we can numerically compute the q-expansion of $f|_k w_Q$ with $f \in S_k(\Gamma_1(N))$.

Let now a matrix $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ be given. Put

$$c' := \gcd(N, c) \quad \text{and} \quad Q := N/c'.$$

Because of $\gcd(c/c', Q) = 1$ we can find $\alpha \in (\mathbb{Z}/Q\mathbb{Z})^\times$ with $\alpha c/c' \equiv 1 \bmod Q$. If we lift α to $(\mathbb{Z}/N\mathbb{Z})^\times$ then we have $\alpha c \equiv c' \bmod N$. Let now $d' \in \mathbb{Z}$ be a lift of αd. We have $\gcd(c', d') = \gcd(c', d', N) = 1$ so we can find $a', b' \in \mathbb{Z}$ that satisfy $a'd' - b'c' = 1$. According to Lemma 2.1, we have

$$\gamma = \gamma_0 \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \quad \text{with } \gamma_0 \in \Gamma_0(N).$$

Put $R = \gcd(c', Q)$. Then we have $\gcd(NR, Q^2 R a') = QR \gcd(c', Qa') = QR^2$ so there exist $b'', d'' \in \mathbb{Z}$ with

$$w_Q := \begin{pmatrix} QRa' & b'' \\ NR & Qd'' \end{pmatrix}$$

having determinant QR^2. One can now verify that we have $\left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) = w_Q \beta$ with $\beta \in \mathrm{GL}_2^+(\mathbb{Q})$ upper triangular. So in the decomposition

$$\gamma = \gamma_0 w_Q \beta$$

we can compute the slash action of all three matrices on the right hand side in terms of q-expansions, hence also of γ.

In conclusion we see that in this method we have to increase the level and go to $S_k(\Gamma_1(NR^2))$ for the square divisors R^2 of N to compute q-expansions of cusp forms in $S_k(\Gamma_1(N))$ at arbitrary cusps.

2.2.4 Numerical evaluation of cusp forms

For $f \in S_k(\Gamma_1(N))$ and a point $P \in \mathfrak{H}$ we wish to compute $f(P)$ to a high numerical precision. Before we do this let us say some words on how P should be represented. Looking at Figure 1.1 on page 2 we convince ourselves that representing P as $x + iy$ with $x, y \in \mathbb{R}$ is not a good idea, as this would be numerically very unstable when P is close to the real line. Instead, we represent P as

$$P = \gamma z \quad \text{with } \gamma \in \mathrm{SL}_2(\mathbb{Z}),\ z = x + iy,\ x \ll \infty \text{ and } y \gg 0. \tag{2.10}$$

For instance, one could demand $z \in \mathscr{F}$, although this is not strictly necessary.

So let $P = \gamma z$ be given, with $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and $\Im z > M$, say. Let $w = w(\gamma)$ be the width of the cusp $\gamma\infty$ with respect to $\Gamma_1(N)$. To compute $f(P)$ we make use of a q-expansion of f at $\gamma\infty$:

$$f(P) = (cz + d)^k (f|_k \gamma)(z) = (cz + d)^k \sum_{n \geq 1} a_n q^{n/w} \quad \text{where } q^{1/w} = \exp(2\pi i z/w).$$

The radius of convergence is 1 and the coefficients are small (estimated by $\tilde{O}(n^{(k-1)/2})$). So to compute $f(P)$ to a precision of p decimals we need about $\frac{pw\log 10}{2\pi M} \approx 0.37\frac{pw}{M}$ terms of the q-expansion of $f|_k\gamma$.

Of course, we have some freedom in choosing γ and z to write down P. We want to find γ such that $P = \gamma z$ with $\Im z/w(\gamma)$ as large as possible. In general, one can always write $P = \gamma z$ with $z \in \mathscr{F}$ so one obtains

$$\max_{\gamma\in \mathrm{SL}_2(\mathbb{Z})} \frac{\Im \gamma^{-1}P}{w(\gamma)} \geq \frac{\sqrt{3}}{2N}. \tag{2.11}$$

We see that in order to calculate $f(P)$ to a precision of p decimals it suffices to use about $\frac{pN\log 10}{\sqrt{3}\pi} \approx 0.42pN$ terms of the q-expansions at each cusp. Although for most points P there is a better way of writing it as γz in this respect than taking $z \in \mathscr{F}$, it seems hard to improve the bound $\frac{\sqrt{3}}{2N}$ in general.

We wish to adjust the representation sometimes from $P = \gamma z$ to $P = \gamma' z'$ where $\gamma' \in \mathrm{SL}_2(\mathbb{Z})$ is another matrix, for instance because during our calculations $\Re z$ has become too large or $\Im z$ has become too small (but still within reasonable bounds). We can make $\Re z$ smaller by putting $z' := z - n$ for appropriate $n \in \mathbb{Z}$ and putting $\gamma' := \gamma \left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right)$. Making $\Im z$ larger is very easy as well. We want to find $\gamma'' = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$\Im \gamma'' z = \frac{\Im z}{|cz+d|^2}$$

is large. But this simply means that we have to find a small vector $cz+d$ in the lattice $\mathbb{Z}z+\mathbb{Z}$, something which can be done easily if $\Re z \ll \infty$ and $\Im z \gg 0$. If c and d are not coprime we can divide both by their greatest common divisor to obtain a smaller vector. The matrix γ'' can now be completed and we put $z' := \gamma'' z$ and $\gamma' := (\gamma'')^{-1}$.

2.2.5 Numerical evaluation of integrals of cusp forms

In this subsection we will describe for $f \in S_2(\Gamma_1(N))$ and $P \in \mathfrak{H}$ how to evaluate the integral $\int_\infty^P f dq/q$. As in the previous subsection, we assume P to be given by means of (2.10). The path of integration will be broken into two parts: first we go from ∞ to a cusp α near P and then we go from α to P.

Integrals over paths between cusps

The pairing (2.1) gives a map

$$\Theta : \mathbb{M}_2(\Gamma_1(N)) \to \mathrm{Hom}_{\mathbb{C}}\left(S_2(\Gamma_1(N)), \mathbb{C}\right),$$

which is injective when restricted to $\mathbb{S}_2(\Gamma_1(N))$. The image of Θ is a lattice of full rank, hence the induced map

$$\mathbb{S}_2(\Gamma_1(N)) \otimes \mathbb{R} \to \mathrm{Hom}_{\mathbb{C}}\left(S_2(\Gamma_1(N)), \mathbb{C}\right)$$

is an isomorphism. In particular we obtain a map

$$\Phi : \mathbb{M}_2(\Gamma_1(N)) \to \mathbb{S}_2(\Gamma_1(N)) \otimes \mathbb{R},$$

which is an interesting map to compute if we want to calculate integrals of cusp forms along paths between cusps. The map Φ is called a *period mapping*.

The Manin-Drinfel'd theorem (see [51, Corollary 3.6] and [26, Theorem 1]) tells us that $\mathrm{im}(\Phi) \subset \mathbb{S}_2(\Gamma_1(N)) \otimes \mathbb{Q}$. This is equivalent to saying that each degree 0 divisor of $X_1(N)$ which is supported on cusps is a torsion point of $J_1(N)$. The proof given in [26] already indicates how to compute Φ with symbolic methods: let p be a prime that is 1 mod N. Then the operator $p+1-T_p$ on $\mathbb{M}_2(\Gamma_1(N))$ has its image in $\mathbb{S}_2(\Gamma_1(N))$. The same operator is invertible on $\mathbb{S}_2(\Gamma_1(N)) \otimes \mathbb{Q}$. So we simply have

$$\Phi = (p+1-T_p)^{-1}(p+1-T_p),$$

where the rightmost $p+1-T_p$ denotes the map $\mathbb{M}_2(\Gamma_1(N)) \to \mathbb{S}_2(\Gamma_1(N))$ and the leftmost $p+1-T_p$ denotes the invertible operator on $\mathbb{S}_2(\Gamma_1(N)) \otimes \mathbb{Q}$. For other methods to compute Φ, see [79, Section 10.6]. So we can express the integral of $f dq/q$ between any two cusps α and β in terms of period integrals, which we have already seen how to compute:

$$\int_\alpha^\beta f \frac{dq}{q} = \langle \Phi(\{\alpha, \beta\}), f \rangle.$$

Integrals over general paths

We can imitate the previous subsection pretty much. Write $P \in \mathfrak{H}$ as $P = \gamma z$ with $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\Im z/w(\gamma\infty)$ is as large as possible. Then we have

$$\int_\infty^P f \frac{dq}{q} = \int_\infty^{\gamma\infty} f \frac{dq}{q} + \int_{\gamma\infty}^{\gamma z} f \frac{dq}{q} = \int_\infty^{\gamma\infty} f \frac{dq}{q} + \int_\infty^z (f|_2\gamma) \frac{dq}{q}. \tag{2.12}$$

The integral $\int_\infty^{\gamma\infty} f \frac{dq}{q}$ is over a path between two cusps so we can compute it by the above discussion and the integral $\int_\infty^z (f|_2\gamma)\frac{dq}{q}$ can be computed using the q-expansion of $f|_2\gamma$:

$$\int_\infty^z (f|_2\gamma) \frac{dq}{q} = w \sum_{n \geq 1} \frac{a_n}{n} q^{n/w},$$

where $w = w(\gamma)$, $q^{1/w} = \exp(2\pi i z/w)$ and $f|_2\gamma = \sum a_n q^{n/w}$. Because of (2.11), computing about $\frac{pN \log 10}{\sqrt{3}\pi} \approx 0.42pN$ terms of the series should suffice to compute $\int_\infty^P f \frac{dq}{q}$ for any $P \in \mathfrak{H}$.

Note also that we can use formula (2.12) to compute the pseudo-eigenvalue $\lambda_Q(f)$ by plugging in $\gamma = w_Q$ and a z for which both $\mathrm{im}\, z$ and $\mathrm{im}\, w_Q z$ are high and for which $\int_\infty^z W_q(f) dq/q$ is not too close to zero.

2.3 Computation of modular Galois representations

In this section, we will give a short overview of the project [28] to which the research of this thesis belongs. Here we omit many details which can be found in [28]. However, we will not give precise references to sections or theorems, since at the time of writing the present section, the paper [28] is undergoing a huge revision. In the first few subsections we will explain the theoretical ideas and in Subsection 2.3.3 we will discuss how to perform actual computations.

A motivational question is: how fast can the q-coefficients of a modular form be computed? Our main example here will be the Ramanujan tau function, but we remark that most techniques that we discuss here can be generalised.

From the recurrence properties on page 6 it follows that we can compute $\tau(n)$ if we can factor n and compute $\tau(p)$ for all prime factors $p \mid n$. Also, in [4] it was shown that we can factor numbers $n = pq$ where p and q are distinct unknown primes if we can compute $\tau(n)$ and $\tau(n^2)$, provided at least one of these numbers is non-zero. The idea is as follows: put $\alpha = \tau(p)/p^{11}$ and $\beta = \tau(q)/q^{11}$. We can compute α and β because their product is $\tau(n)/n^{11}$ and their sum is $(\tau(n)^2 - \tau(n^2) - n^{11})/n^{11}$. The primes p and q can now be obtained by looking at the denominators of α and β.

Because of the above discussion, it seems reasonable to focus on computing $\tau(p)$ for p prime. A strategy for this is computing $\tau(p) \bmod \ell$ for many small primes ℓ. If the product of all these primes ℓ exceeds $4p^{11/2}$ then by the bound $|\tau(p)| \leq 2p^{11/2}$ we know exactly what $\tau(p)$ is. The main theorem of [28] is the following:

Theorem 2.2. *There exists a probabilistic algorithm that on input two prime numbers p and ℓ with $p \neq \ell$ can compute $\tau(p) \bmod \ell$ in expected time polynomial in $\log p$ and ℓ.*

Corollary 2.1. *There exists a probabilistic algorithm that on input a prime number p can compute $\tau(p)$ in expected time polynomial in $\log p$.*

2.3.1 Computing representations for $\tau(p) \bmod \ell$

We saw in Subsection 1.1.2 that for some values of ℓ, called exceptional primes, there exist simple formulas for $\tau(p) \bmod \ell$. So assume from now that ℓ is non-exceptional. We can work with the residual representations $\overline{\rho}_\ell := \overline{\rho}_{\Delta,\ell}$, see Subsections 1.3.4 and 1.3.5. For $p \neq \ell$ we have

$$\tau(p) \equiv \operatorname{tr}(\overline{\rho}(\mathrm{Frob}_p)) \bmod \ell.$$

If we put $K_\ell := \overline{\mathbb{Q}}^{\ker(\overline{\rho}_\ell)}$ then $\overline{\rho}_\ell$ factors through $\mathrm{Gal}(K_\ell/\mathbb{Q})$. Our main task is to give a polynomial P_ℓ whose splitting field is K_ℓ. Since $\operatorname{im}\rho_\ell$ acts faithfully and transitively on $\mathbb{F}_\ell^2 - \{0\}$ (remember that ℓ is non-exceptional), we will demand that P_ℓ has degree $\ell^2 - 1$ and that the number field K'_ℓ defined by P_ℓ is the subfield of K_ℓ that is fixed by the stabiliser of a point in

$\mathbb{F}_\ell^2 - \{0\}$.

We can find ρ_ℓ inside the Jacobian of $X_1(\ell)$. If $\mathbb{T} \subset \mathrm{End}(J_1(\ell))$ is the algebra generated by the diamond and Hecke operators acting on $J_1(\ell)$ then we have a homomorphism

$$\theta = \theta_{\Delta,\ell} : \mathbb{T} \to \mathbb{F}_\ell, \quad \theta : \langle d \rangle \mapsto d^{10} \bmod \ell, \quad \theta : T_n \mapsto \tau(n) \bmod \ell.$$

If $I \subset \mathbb{T}$ denotes the kernel of θ, then ρ_ℓ can be defined as $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acting on $V_\ell := J_1(\ell)(\overline{\mathbb{Q}})[I]$, which is a 2-dimensional $\mathbb{F}_\ell$-linear subspace of $J_1(\ell)(\overline{\mathbb{Q}})[\ell]$. One can express this space in terms of modular symbols since we have isomorphisms

$$J_1(\ell)(\mathbb{C})[\ell] \cong H_1(X_1(\ell)(\mathbb{C}), \mathbb{F}_\ell) \cong \mathbb{S}_2(\Gamma_1(\ell)) \otimes \mathbb{F}_\ell$$

and the action of $\mathbb{T}$ on $\mathbb{S}_2(\Gamma_1(\ell))$ can be computed.

Let g be the genus of $X_1(\ell)$. If we choose an effective divisor D of degree g on $X_1(\ell)$ then we have a morphism

$$\phi : X_1(\ell)^g \to J_1(\ell), \quad (Q_1, \ldots, Q_g) \mapsto \sum_{i=1}^{g} Q_i - D$$

which induces a birational morphism

$$\phi' : \mathrm{Sym}^g X_1(\ell) \to J_1(\ell). \tag{2.13}$$

Suppose that D is such that ϕ is étale over V_ℓ. Take a function $f \in \mathbb{Q}(X_1(\ell))$ such that for any $(Q_1, \ldots, Q_g) \in \phi^{-1}(V_\ell - \{0\})$ it has no poles at the Q_i and such that the induced map $f_* : \mathrm{Sym}^g(X_1(\ell)) \to \mathrm{Sym}^g(\mathbb{P}^1_{\mathbb{Q}})$ is injective on $\phi'^{-1}(V_\ell - \{0\})$.

The field K'_ℓ is the field of definition of a point $P \in V_\ell - \{0\}$. Put $\phi'^{-1}(P) = (Q_1, \ldots, Q_g)$. Then certainly K'_ℓ contains $e_i(f(Q_1), \ldots, f(Q_g))$ for all i, where e_i is the i-th elementary symmetric polynomial in g variables. But in fact we have an equality

$$K'_\ell = \mathbb{Q}\big(e_1(f(Q_1), \ldots, f(Q_g)), \ldots, e_g(f(Q_1), \ldots, f(Q_g))\big).$$

This can be seen as follows: the field on the right hand side, say L, is the field of definition of $f_*(\phi'^{-1}(P))$. The group $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$ acts on $\mathrm{Sym}^g \mathbb{P}^1(\overline{\mathbb{Q}})$ and fixes $f_*(\phi'^{-1}(P))$. But f_* is injective on $\phi'^{-1}(V_\ell - \{0\})$ so $\mathrm{Gal}(\overline{\mathbb{Q}}/L)$ fixes P as well. So L contains, hence is equal to, K'_ℓ.

In practice, it often suffices to take $D = g \cdot [0]$ (remember from Subsection 1.2.3 that the cusp 0 is defined over $\mathbb{Q}$) and any non-constant f. The field K'_ℓ will almost always be equal to $\mathbb{Q}(f(Q_1) + \cdots + f(Q_g))$. If we assume that all of this is correct, then P_ℓ will be equal to

$$P_\ell = \prod_{P \in V_\ell - \{0\}} \big(x - \sum_i f(Q_i)\big) \quad \text{where } (Q_1, \ldots, Q_g) = \phi'^{-1}(P). \tag{2.14}$$

In theory however, to show that a good divisor D and a good function f can be found, one has to work with $X_1(5\ell)_{\mathbb{Q}(\zeta_\ell)}$ instead of $X_1(\ell)$. In this thesis, we will ignore these theoretical

complications. The main reasons for this are that we want to compute actual polynomials and we want to explain ideas rather than technical details.

To compute the polynomial P_ℓ we will use numerical methods. The idea is to approximate the coefficients of P_ℓ. This could be done in several ways, for instance approximating them p-adically for one or more primes p or approximating them in $\mathbb{R}$. In [17] and [40] one can find methods to compute with modular curves over $\mathbb{F}_p$ which can be used to compute $P_\ell \bmod p$ for primes p. Note that this is a special case of p-adically approximating P_ℓ. In Subsection 2.3.3 we will describe how to approximate P_λ over the reals, in a way that is practically convenient.

Heights

If the used precision for the approximation of P_ℓ is high enough, we can compute the exact coefficients in $\mathbb{Q}$. To know how high this precision should actually be, we use *height bounds*.

Definition 2.3. Let K be a number field and take $\alpha \in K$. Then the (logarithmic) field height of α is defined as

$$\mathrm{ht}_K(\alpha) := \sum_v [K_v : \mathbb{Q}_v] \log \max(1, |\alpha|_v).$$

Here, the sum is taken over all places of K and the absolute value is normalised by demanding $|p|_v = 1/p$ for v finite lying above p and $|x|_v = |\sigma(x)|$ for v infinite belonging to the embedding $\sigma : K \hookrightarrow \mathbb{C}$. The absolute (logarithmic) height of α is defined as

$$\mathrm{ht}(\alpha) := \frac{\mathrm{ht}_K(\alpha)}{[K : \mathbb{Q}]}.$$

The absolute height of an algebraic number is independent of the number field we put around it. Also note that for a rational number p/q written in lowest terms we have $\mathrm{ht}(p/q) = \log \max(|p|, |q|)$.

Definition 2.4. Let K be a number field and consider a point $P = (\alpha_0 : \ldots : \alpha_n) \in \mathbb{P}^n(K)$. Then the (logarithmic) field height of P is defined as

$$\mathrm{ht}_K(P) := \sum_v [K_v : \mathbb{Q}_v] \log \max_i |\alpha_i|_v,$$

using the same conventions for valuations as in Definition 2.3. The absolute (logarithmic) height of P is defined as

$$\mathrm{ht}(P) := \frac{\mathrm{ht}_K(P)}{[K : \mathbb{Q}]}.$$

It is a fact that this definition is consistent in the sense that the height does not depend on the scaling of projective coordinates. Again, the absolute height of $P \in \mathbb{P}^n(\overline{\mathbb{Q}})$ does not depend on the chosen number field. If we write $P \in \mathbb{P}^n(\mathbb{Q})$ as $(p_0 : \ldots : p_n)$ with p_i coprime integers, then $\mathrm{ht(P)} = \log \max_i |p_i|$.

For $P = a_n x^n + \cdots + a_0 \in K[x]$ with K a number field we define the height of P as the height of $(a_0 : \ldots : a_n) \in \mathbb{P}^n(K)$. If $P \in \mathbb{Q}[x]$ is an irreducible polynomial of degree d and $\alpha \in \overline{\mathbb{Q}}$ is a root of P then we have the following estimations between the height of P and the field height of α in $\mathbb{Q}(\alpha)$:

$$\mathrm{ht}(P) - d\log 2 \leq d\,\mathrm{ht}(\alpha) \leq \mathrm{ht}\,P + \log(d+1)/2.$$

This means that bounding the height of P_λ is equivalent with bounding the height of its roots. One can embed $X_1(\ell)$ into projective space. Bounding the roots of P_λ boils then down to bounding the Q_i occurring in formula (2.14), or rather the version of this formula that can be proven to be correct. Using a vast amount of highly non-trivial Arakelov geometry, Bas Edixhoven and Robin de Jong succeeded in bounding the Q_i and using this to show that $\mathrm{ht}(P_\ell)$ is bounded polynomially in ℓ.

Their method relies on the fact that Δ is a modular form of level one. In fact, this method works for any newform of level one. At the time of writing this section, it is not known how to produce bounds for more general levels but some progress on this is expected to be made soon.

Suppose now that a height bound for a rational number $x = p/q$ (written in lowest terms with $q > 0$) is known, say $\mathrm{ht}(x) < C$. Using non-archimedean local approximations of x one can find a large integer $M > 0$ with $\gcd(q,M) = 1$ and with $x \bmod M$ congruent to a given number a. Using real approximations, one can find a small $\varepsilon > 0$ and a ξ such that $|x - \xi| \leq \varepsilon |x| < \varepsilon \frac{\exp C}{q}$. If one doesn't use non-archimedian approximations, one can take $M = 1$ and if one doesn't use real approximations one can put $\xi = 0$ and $\varepsilon = 1$. If the approximations are close enough to satisfy $\log \frac{M}{2\varepsilon} > 2C$ then they determine the number x: suppose that $x' = p'/q'$ is another rational number satisfying the same approximation conditions as x. Then we have

$$2\varepsilon \frac{\exp(2C)}{qq'} > \varepsilon \exp(C)\left(\frac{1}{q} + \frac{1}{q'}\right) > |x-\xi| + |x'-\xi| \geq |x - x'| \geq \frac{M}{|qq'|}, \tag{2.15}$$

leading to a contradiction with $\log \frac{M}{2\varepsilon} > 2C$.

We want to actually compute x from its approximations and height bound. Note that the above reasoning is still valid if we weaken the condition $p/q \equiv a \bmod M$ to $p \equiv qa \bmod M$, dropping the assumption $\gcd(q,M) = 1$. We will change our notation a bit and assume that the approximation ξ is given in terms of a rational number $\xi = m/n$ with $n > 0$ (so typically n will be a power of 2 or 10). We thus assume

$$\left| \frac{p}{q} - \frac{m}{n} \right| < \frac{1}{2n} \tag{2.16}$$

and the condition that we need to determine p/q uniquely is

$$\log \frac{Mn}{q} > C.$$

We can use the extended Euclidean algorithm [73, Section 4.2] with $(na-m, nM)$ as input to generate a sequence of triples (q_i, r_i, s_i) satisfying $(na-m)q_i + nMr_i = s_i$ with $|s_i|$ decreasing and $|q_i r_{i+1} - q_{i+1} r_i| = 1$ for all i. Put $r = (p-qa)/M$ and $s = pn - qm$. From (2.16) it follows that the triple (q,r,s) satisfies $(na-m)q + nMr = s$ with $|s| < \frac{q}{2} < \frac{Mn}{2\exp(C)}$. By [73, Theorem 4.9] the first index i for which the bound $|s_i| \leq \lceil \frac{Mn}{2\exp(C)} \rceil - 1$ holds satisfies $|q_i| \leq \lceil \exp(C) \rceil - 1$ and $r_i/q_i = r/q$, thus also $p/q = (q_i a + Mr_i)/q_i$.

2.3.2 Computing $\tau(p) \bmod \ell$ from P_ℓ

The image of $\overline{\rho}_\ell$ is a group G between $\mathrm{SL}_2(\mathbb{F}_\ell)$ and $\mathrm{GL}_2(\mathbb{F}_\ell)$. The stabiliser subgroup of a basis of $\mathbb{F}_\ell^2$ in G is trivial, so K_ℓ can be obtained by adjoining two roots of P_ℓ; make sure that the second root is not in the field generated by the first root. There are methods to compute this [45, Corollary 6]. Also, we have obtained P_ℓ from approximations in $J_1(\ell)$. From this we can deduce a bijection between the roots of P_ℓ and $V_\ell - \{0\}$ that induces an isomorphism $\mathrm{Gal}(K_\ell/\mathbb{Q}) \cong G$ which defines $\overline{\rho}_\ell$.

Let p be a prime different from ℓ. We want to compute the conjugacy class $[\mathrm{Frob}_p]$ inside $\mathrm{Gal}(K_\ell/\mathbb{Q})$. This would give us $\overline{\rho}_\ell(\mathrm{Frob}_p)$ and thus $\tau(p) \bmod \ell$. To do this, one first computes the maximal order $\mathscr{O}_{K_\ell}$ of K_ℓ [11, Theorem 1.4]. For a prime $\mathfrak{p}$ of K_ℓ above p we have that $\mathrm{Frob}_{\mathfrak{p}/p}$ is equal to the unique $\sigma \in G$ that satisfies $\sigma(\mathfrak{p}) = \mathfrak{p}$ and $\sigma(x) \equiv x^p \bmod \mathfrak{p}$ for all $x \in \mathscr{O}_{K_\ell}$. We have a decomposition

$$\mathscr{O}_{K_\ell}/(p) \cong \prod_{\mathfrak{p}|p} \mathscr{O}_{K_\ell}/\mathfrak{p}.$$

So $\mathrm{Frob}_{\mathfrak{p}/p}$ is the element that fixes $\mathscr{O}_{K_\ell}/\mathfrak{p}$ in this decomposition and acts there as $x \mapsto x^p$. To check whether $\sigma \in G$ is equal to $\mathrm{Frob}_{\mathfrak{p}/p}$ for at least one $\mathfrak{p} \mid p$ we do the following. Both σ and $x \to x^p$ are $\mathbb{F}_p$-linear maps from $\mathscr{O}_{K_\ell}/(p)$ to itself; compute them. We have $\sigma = \mathrm{Frob}_{\mathfrak{p}/p}$ if and only if the image of the map $\sigma - (x \to x^p)$ is contained in $\mathfrak{p}$. From this it follows that σ is equal to at least one of the $\mathrm{Frob}_{\mathfrak{p}/p}$ if and only if the ideal in $\mathscr{O}_{K_\ell}/(p)$ generated by the image of $\sigma - (x \to x^p)$ is not the unit ideal. So given p, we can obtain $[\mathrm{Frob}_p]$ by checking the above for all $\sigma \in G$.

2.3.3 Explicit numerical computations

Let now an arbitrary positive integer N be given and let $f \in S_2(\Gamma_1(N))$ be a newform with character ε (remember from Subsection 1.3.4 that we can reduce to the weight 2 case). Also, let ℓ be a prime number and let λ be a prime of K_f lying above ℓ. Assume that the representation $\overline{\rho}_{f,\lambda}$ is absolutely irreducible and let $\mathbb{T}$ be the Hecke algebra acting on $J_1(N)$. In Subsection 1.3.4 we saw that there is a subspace V_λ of $J_1(N)(\overline{\mathbb{Q}})[\ell]$ on which both $\mathbb{T}$ and $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ act, such that the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ defines $\overline{\rho}_{f,\lambda}$.

Approximation of torsion points

The Jacobian $J_1(N)_{\mathbb{C}}$ can be described as follows. Pick a basis $f_1,\ldots,f_g$ of $S_2(\Gamma_1(N))$. Put

$$\Lambda := \left\{ \int_\gamma (f_1,\ldots,f_g)\frac{dq}{q} : [\gamma] \in H_1(X_1(N)(\mathbb{C}),\mathbb{Z}) \right\} \subset \mathbb{C}^g.$$

This is a lattice in $\mathbb{C}^g$ of full rank. By the Abel-Jacobi theorem we have an isomorphism

$$J_1(N)(\mathbb{C}) \xrightarrow{\sim} \mathbb{C}^g/\Lambda, \quad \big[\sum_i ([Q_i]-[R_i])\big] \mapsto \sum_i \int_{R_i}^{Q_i} (f_1,\ldots,f_g)\frac{dq}{q}.$$

Let again a divisor $D = \sum_{i=1}^g [R_i]$ on $X_1(N)$ be given. Identifying $J_1(N)(\mathbb{C})$ with $\mathbb{C}^g/\Lambda$ in this way, the map (2.13) becomes a birational morphism

$$\phi' : \mathrm{Sym}^g X_1(N)(\mathbb{C}) \to \mathbb{C}^g/\Lambda, \quad (Q_1,\ldots,Q_g) \mapsto \sum_{i=1}^g \int_{R_i}^{Q_i} (f_1,\ldots,f_g)\frac{dq}{q}.$$

The homology group $H_1(X_1(N)(\mathbb{C}),\mathbb{Z})$ is canonically isomorphic to the modular symbols space $\mathbb{S}_2(\Gamma_1(N))$. The period lattice Λ can thus be computed numerically using the methods from Subsections 2.2.1 and 2.2.2. Since we can compute the action of $\mathbb{T}$ on $\mathbb{S}_2(\Gamma_1(N)) \cong \Lambda$, we can write down the points in $\frac{1}{\ell}\Lambda/\Lambda \subset \mathbb{C}^g/\Lambda$ that correspond to the points of V_λ. The aim is now to compute the divisors on $X_1(N)_{\mathbb{C}}$ that map to these points along ϕ'. In our computations, we assume without proof that ϕ is étale above V_λ.

We start calculating with a small precision. Let $P \in V_\lambda(\mathbb{C}) \subset \mathbb{C}^g/\Lambda$ be given. First we try out a lot of random points $Q = (Q_1,\ldots,Q_g) \in X_1(N)(\mathbb{C})$. Here, each Q_i will be written as $Q_i = \gamma_i w_i$, with γ_i in a set of representatives for $\Gamma_1(N)\setminus \mathrm{SL}_2(\mathbb{Z})$ and $w_i \in \mathscr{F}$. We can compute $\phi'(Q)$ using methods from Subsection 2.2.5. We work with the point Q for which $\phi'(Q)$ is closest to P. If we in fact already know some points Q with $\phi'(Q)$ approximately equal to a point in $V_\lambda(\mathbb{C})$, then we could also take one of those points as a starting point Q to work with.

The next thing to do is adjust Q so that $\phi'(Q)$ comes closer to P. We'll make use of the Newton-Raphson approximation method. Let $\phi'' : \mathfrak{H}^g \to \mathbb{C}^g/\Lambda$ be the function defined by

$$\phi''(z_1,\ldots z_g) = \phi'(\gamma_1 z_1,\ldots \gamma_g z_g).$$

We observe that for a small vector $h = (h_1,\ldots h_g) \in \mathbb{C}^g$ we have

$$\phi''(w_1+h_1,\ldots,w_g+h_g) = \phi'(Q) + hD + O(\|h\|^2)$$

with

$$D = \left(\begin{array}{ccc} \frac{\partial \phi_1''}{\partial z_1} & \cdots & \frac{\partial \phi_g''}{\partial z_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial \phi_1''}{\partial z_g} & \cdots & \frac{\partial \phi_g''}{\partial z_g} \end{array} \right)\Bigg|_{(w_1,\ldots,w_g)}.$$

From the definition of ϕ' we can immediately deduce

$$\frac{\partial \phi_i''}{\partial z_j}(w_1,\ldots,w_g) = 2\pi i (f_i|_2\gamma_j)(w_j),$$

where we apologise for the ambiguous i. We can thus compute the matrix D using the methods of Subsection 2.2.4. Now choose a small vector $v = (v_1,\ldots,v_g) \in \mathbb{C}^g$ such that $\phi'(Q)+v$ is closer to P than $\phi'(Q)$ is. For example, v can be chosen among all vectors of a bounded length so that $\phi'(Q)+v$ is closest to P. If we write

$$h = vD^{-1},$$

then we expect $\phi''(w_1+h_1,\ldots,w_g+h_g)$ to be approximately equal to $\phi'(Q)+v$. If this is not the case, then we try the same thing with a smaller v. It could be that this still fails, for instance because we are too close to the non-étale locus of the map ϕ. In that case, we start with a new random point Q.

We repeat the above adjustments until we are (almost) as close as we can get, considering our calculation precision. It might happen that the w_i become too wild, i.e. $|\Re w_i|$ becomes too large or $\Im w_i$ becomes too small. If this is the case we adjust the way we write Q_i as $\gamma_i w_i$ using the method described in Subsection 2.2.4. We can always replace the γ_i then by a small matrix in the same coset of $\Gamma_1(N)\setminus \mathrm{SL}_2(\mathbb{Z})$.

Once we have for each $P \in V_\lambda$ a point Q such that $\phi'(Q)$ is approximately equal to P, we can start increasing the precision. We double our calculation precision and repeat the above adjustments ($\phi'(Q)+v$ will in this case be equal to P). We repeat this a few times until we have very good approximations.

Computation of polynomials

Now, we have to choose a function in $h \in \mathbb{Q}(X_1(N))$. Since h multiplies heights of points roughly by $\deg(h)$, we want to find a function of small degree. Take any k and a basis $h_1,\ldots,h_n$ of $S_k(\Gamma_1(N))$ such that the q-expansions of the h_i lie in $\mathbb{Z}[[q]]$ and such that the exponents of the first non-zero terms of these q-expansions form a strictly increasing sequence. We propose to use $h = W_N(h_{n-1})/W_N(h_n)$ as a function to use (assuming $n \geq 2$). Remember from Subsection 1.2.4 that $S_k(\Gamma_1(N))$ is the space of global sections of the line bundle $\mathscr{L} = \omega^{\otimes k}(-\mathrm{cusps})$ on $X_1(N)$, base changed to $\mathbb{C}$. Remember also that the cusp ∞ is not defined over $\mathbb{Q}$, but the cusp 0 is. Since we demand the q-expansions to have rational coefficients, the sections $W_N(h_1),\ldots,W_N(h_n)$ are defined over $\mathbb{Q}$ and they have increasing order at 0. One can now verify that for $h = W_N(h_{n-1})/W_N(h_n)$ we have

$$\deg(h) \leq \deg(\mathscr{L}) - \mathrm{ord}_\infty(h_{n-1}) \leq \deg(\mathscr{L}) - \dim H^0(\mathscr{L}) + 2 \leq g+1.$$

For $k=2$ and $g \geq 2$ we have $\mathscr{L} \cong \Omega^1(X_1(N))$ and we get g as an upper bound for $\deg(h)$. Using methods from Subsection 2.2.4, we can evaluate h numerically. The author is not aware of a sophisticated method for finding a function $h \in \mathbb{Q}(X_1(N))$ of minimal degree in

general; this minimal degree is called the *gonality* of the curve $X_1(N)$. Published results on these matters seem to either be limited to $X_0(N)$ or to concern only *lower* bounds for the gonality of modular curves, see for example [1], [5, Chapter 3] or [60].

Now put

$$\alpha_P = \sum_{i=1}^{g} h(Q_i), \quad \text{for } P \in V_\lambda(\mathbb{C}) - \{0\} \text{ and where } \phi'(Q_1, \ldots, Q_g) = P.$$

We work out the product in

$$P_\lambda(x) := \prod_{P \in V_\lambda(\mathbb{C}) - \{0\}} (x - \alpha_i) = \sum_{k=0}^{n} a_k x^k, \quad \text{where } n = \deg P_\lambda.$$

The coefficients a_k are rational numbers that we have computed numerically. Since the height of P_λ is expected to be not too large, the denominators of the a_k should have a relative small common denominator. The LLL algorithm can be used to compute integers $p_0, \ldots, p_{n-1}, q$ such that $|p_k - a_k q|$ is small for all k, see [49, Proposition 1.39]. If the sequence (a_k) is arbitrary, then we'll be able to find p_k and q such that $|p_k - a_k q|$ is roughly of order $q^{-1/n}$ for each k, but not much better than that. So if it happens that we find p_k and q with $|p_k - a_k q|$ much smaller than $q^{-1/n}$ for all k, then we guess that a_k is equal to p_k/q. If we cannot find such p_k and q then we will double the precision and repeat all the calculations described above.

Heuristically, the calculation precision that is needed to find the true value of a_k is about $(1 + 1/n)\,\mathrm{ht}(P_\lambda)/\log(10)$ decimals. Another way of finding rational approximations of the a_k is by approximating them using continued fractions. For this method, the precision needed to find the true value of a_k would be about $2\,\mathrm{ht}(P_\lambda)/\log(10)$ decimals.

Since the degree of P_λ will be quite large, we won't be able to do many further calculations with it. In particular it may be hard to verify whether all the guesses we made were indeed correct. Instead, we will look at the following variant. If $\mathfrak{m}$ is the Hecke ideal of f mod λ, then V_λ is a vector space over $\mathbb{T}/\mathfrak{m}$. The representation $\overline{\rho}_{f,\lambda}$ induces an action $\tilde{\rho}_\lambda$ of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the set $\mathbb{P}(V_\lambda)$ of lines in V_λ. We can attach a polynomial $\tilde{P}_\lambda$ to this projectivised representation $\tilde{\rho}_\lambda$, analogously to the way this was done for $\overline{\rho}$. This polynomial will have smaller a degree than P_λ. We put

$$\tilde{P}_\lambda(x) = \prod_{L \in \mathbb{P}(V_\ell)} \big(x - \sum_{P \in L - \{0\}} \alpha_P\big) = \sum_{k=0}^{m} b_k x^k, \quad \text{where } m = \deg \tilde{P}_\lambda.$$

As above, if the calculation precision is sufficient we can use lattice reduction algorithms to compute the exact values of the b_k.

Reduction of polynomials

Although the polynomial $\tilde{P}$ will not have a very huge height, its height is still too large to do any useful computations with it. The first step in making a polynomial of smaller height

defining the same number field is computing the maximal order of that number field. Let q be the common denominator of the coefficients and put $p_k = b_k q$. Consider the polynomial

$$Q(x) = q \cdot \tilde{P}_\lambda(x) = qx^m + p_{m-1}x^{m-1} + \cdots + p_0.$$

We make ourselves confident that we correctly computed $Q(x)$ (although we won't prove anything at this point yet). For instance, we verify that $Q(x)$ is irreducible and that its discriminant has the prime factors of $N\ell$ in it. We can also compute for several primes p not dividing $\mathrm{Disc}(Q(x))$ the decomposition type of $Q(x) \bmod p$ and verify that it could be equal to the cycle type of $\tilde{\rho}(\mathrm{Frob}_p)$. If not, we again double the precision and repeat the above calculations.

Let now α be a root of $\tilde{P}_\lambda(x)$ and write down the order

$$\mathscr{O} := \mathbb{Z} + \sum_{k=1}^{m-1} \left(\mathbb{Z} \cdot \sum_{j=0}^{k-1} a_{m-j} \alpha^{k-j} \right),$$

which is an order that is closer to the maximal order than $\mathbb{Z}[q\alpha]$ (see [48, Subsection 2.10]). Being confident in the correctness of $Q(x)$, we know where the number field K defined by it ramifies and thus we can compute its maximal order (see [11, Section 6 and Theorems 1.1 and 1.4]). Having done this, we embed $\mathscr{O}_K$ as a lattice into $\mathbb{C}^m$ in the usual way and we use the LLL algorithm to compute a basis of small vectors in $\mathscr{O}_K$. We can then search for an element of small length in $\mathscr{O}_K$ that generates K over $\mathbb{Q}$. Its defining polynomial $\tilde{P}'_\lambda$ will have small coefficients. See also [16].

In the computation of the polynomials P_λ and $\tilde{P}'_\lambda$ we made several guesses and assumptions that we cannot prove to be correct. In Chapters 3 and 4, we work out in special cases how we can use established parts of Serre's conjecture to prove afterwards for polynomials of the style $\tilde{P}'_\lambda$ that they indeed belong to the modular Galois representations that we claim they belong to. In the unlikely case that such tests may fail we can of course make adjustments like choosing another function h or another divisor D.

Further refinements

The Jacobian $J_1(N)$ has large dimension (for N prime it is $(N-5)(N-7)/24$). It could be that our newform f is an element of $S_2(\Gamma)$ with $\Gamma_1(N) \lneq \Gamma < \Gamma_0(N)$. In that case we work with the curve X_Γ, which is given its $\mathbb{Q}$-structure by defining it as a quotient of $X_1(N)$. The Jacobian J_Γ of X_Γ is isogenous to an abelian subvariety of $J_1(N)$ that contains V_λ, so this works perfectly well.

In the case $\Gamma = \Gamma_0(N)$ we can sometimes go a step further. The operator W_N on $X_0(N)$ is defined over $\mathbb{Q}$. If f is invariant under W_N, one can work with the curve $X_0^+(N) := X_0(N)/\langle W_N \rangle$. Its Jacobian $J_0^+(N)$ is isogenous to an abelian subvariety of $J_1(N)$ that contains V_λ, so also here it works. Some words on the computation of the homology of $X_0^+(N)$ are in order. The action of W_N on $X_0(N)$ induces an action on $H_1(X_0(N)(\mathbb{C}), \mathbb{Z})$ and on

$H_1(X_0(N)(\mathbb{C}), \text{cusps}, \mathbb{Z})$. Since paths between cusps on $X_0^+(N)(\mathbb{C})$ lift to paths between cusps on $X_0(N)(\mathbb{C})$ we have a surjection

$$H_1(X_0(N), \text{cusps}, \mathbb{Z}) \twoheadrightarrow H_1(X_0^+(N)(\mathbb{C}), \text{cusps}, \mathbb{Z}).$$

The kernel of this surjection consists of the elements $[\gamma] \in H_1(X_0(N), \text{cusps}, \mathbb{Z})$ satisfying $W_N([\gamma]) = -[\gamma]$. So modular symbols methods allow us to compute $H_1(X_0^+(N)(\mathbb{C}), \text{cusps}, \mathbb{Z})$ as a quotient of $\mathbb{M}_2(\Gamma_0(N))$. Let $\mathbb{B}_2^+(\Gamma_0(N))$ be the free abelian group on the cusps of $X_0^+(N)(\mathbb{C})$ and define

$$\delta : H_1(X_0^+(N)(\mathbb{C}), \text{cusps}, \mathbb{Z}) \to \mathbb{B}_2^+(\Gamma_0(N)), \quad \{\alpha, \beta\} \mapsto \{\beta\} - \{\alpha\}.$$

Then $H_1(X_0^+(N)(\mathbb{C})) = \ker(\delta)$.

Chapter 3

A polynomial with Galois group $\mathrm{SL}_2(\mathbb{F}_{16})$

This chapter consists of an article that has been published as [7], with some slight lay-out modifications.

Abstract. In this paper we display an explicit polynomial having Galois group $\mathrm{SL}_2(\mathbb{F}_{16})$, filling in a gap in the tables of Jürgen Klüners and Gunter Malle. Furthermore, the polynomial has small Galois root discriminant; this fact answers a question of John Jones and David Roberts. The computation of this polynomial uses modular forms and their Galois representations.

3.1 Introduction

It is a computational challenge to construct polynomials with a prescribed Galois group; see [44] for methods and examples. Here, by the Galois group of a polynomial $f \in \mathbb{Q}[x]$ we mean the Galois group of a splitting field of f over $\mathbb{Q}$ together with its natural action on the roots of f in this splitting field. Jürgen Klüners informed me about an interesting group for which a polynomial had not been found yet, namely $\mathrm{SL}_2(\mathbb{F}_{16})$ with its natural action on $\mathbb{P}^1(\mathbb{F}_{16})$. This action is faithful because of $\mathrm{char}(\mathbb{F}_{16}) = 2$. It must be noted that the existence of such a polynomial was already known to Mestre (unpublished). In this paper we will give an explicit example.

Proposition 3.1. *The polynomial*

$$\begin{aligned} P(x) := {} & x^{17} - 5x^{16} + 12x^{15} - 28x^{14} + 72x^{13} - 132x^{12} + 116x^{11} - 74x^9 \\ & + 90x^8 - 28x^7 - 12x^6 + 24x^5 - 12x^4 - 4x^3 - 3x - 1 \in \mathbb{Q}[x] \end{aligned}$$

has Galois group isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$ *with its natural action on* $\mathbb{P}^1(\mathbb{F}_{16})$.

What is still unknown is whether there exists a regular extension of $\mathbb{Q}(T)$ with Galois group isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$; regular here means that it contains no algebraic elements over $\mathbb{Q}$ apart from $\mathbb{Q}$ itself. In Section 3.2 we will say some words about the calculation of the polynomial and the connection with modular forms. We'll indicate how one can verify that it

has the claimed Galois group in Section 3.3 using computational Galois theory. We will show in Section 3.4 that this polynomial gives a Galois representation associated to an explicitly given modular form.

3.1.1 Further remarks

In algebraic number theory, the root discriminant of a number field K is defined as $d(K) := |\mathrm{Disc}(\mathscr{O}_K)|^{1/[K:\mathbb{Q}]}$. This way of measuring number fields appears to be very useful in asymptotic analysis on the set of all number fields (inside a fixed algebraic closure of $\mathbb{Q}$, say). An excellent survey paper on this material is [57]. Let us mention some interesting results here as well. For example it is known that the bounds

$$22.38 \approx 4\pi e^{\gamma} \leq \liminf_K d(K) \leq 82.11$$

hold; see [59, Section 7] for the lower bound and [30, Section 3.2] for the upper bound. Under the assumption of the Generalised Riemann Hypothesis we even have

$$\liminf_K d(K) \geq \Omega := 8\pi e^{\gamma} \approx 44.76,$$

see [69]. In view of this lower bound, root discriminants below Ω are called *small* and it is interesting to construct number fields that have small root discriminant. A paper focusing on the construction of Galois number fields with small root discriminant is [33]. A question asked in that paper is whether there exists such a field of which the Galois group contains a subgroup isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$ (see [33, Section 13]). The splitting field of the polynomial in Proposition 3.1 has root discriminant $2^{15/8} \cdot 137^{1/2} \approx 42.93$ and thus answers this question affirmatively.

The example given in Proposition 3.1 is not the only polynomial that the author could produce. Here are the other examples of polynomials having Galois group $\mathrm{SL}_2(\mathbb{F}_{16})$ computed so far:

$$\begin{aligned} &x^{17} + x^{16} - 4x^{15} - 2x^{14} + 54x^{13} + 6x^{12} - 36x^{11} - 16x^{10} + 714x^9 \\ &- 1238x^8 + 484x^7 + 764x^6 - 1084x^5 - 520x^4 + 668x^3 + 776x^2 + 382x + 74 \end{aligned}$$

and

$$\begin{aligned} &x^{17} + x^{16} + 18x^{15} + 10x^{14} + 194x^{13} + 250x^{12} + 442x^{11} + 1006x^{10} + 1176x^9 \\ &- 392x^8 + 1178x^7 + 4490x^6 + 4790x^5 + 1606x^4 + 286x^3 + 38x^2 + 25x + 1. \end{aligned}$$

The former polynomial defines a number field that ramifies above 2 and 173 and the number field defined by the latter polynomial ramifies above 2 and 199. The root discriminants of their splitting fields are not small, as they are equal to $2^{15/8}173^{1/2} \approx 48.25$ and $2^{15/8}199^{1/2} \approx 51.74$ respectively.

3.2 Computation of the polynomial

In this section we will briefly indicate how one can find a polynomial like the one in Proposition 3.1. We will make use of modular forms. For an overview as well as many further references on this subject the reader is referred to [24].

Let N be a positive integer and consider the space $S_2(\Gamma_0(N))$ of holomorphic cusp forms of weight 2 for $\Gamma_0(N)$. A newform $f \in S_2(\Gamma_0(N))$ has a q-expansion $f = \sum a_n q^n$ where the coefficients a_n are in a number field. The smallest number field containing all the coefficients is denoted by K_f. To a given prime number ℓ and a place λ of K_f above ℓ one can attach a semi-simple Galois representation $\overline{\rho}_f = \overline{\rho}_{f,\lambda} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_\lambda)$ unramified outside $N\ell$ satisfying the following property: for each prime $p \nmid N\ell$ and any Frobenius element Frob_p in $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ attached to p we have

$$\mathrm{tr}(\overline{\rho}_f(\mathrm{Frob}_p)) \equiv a_p \bmod \lambda \quad \text{and} \quad \det(\overline{\rho}_f(\mathrm{Frob}_p)) \equiv p \bmod \lambda. \tag{3.1}$$

The representation $\overline{\rho}_f$ is unique up to isomorphism. The fixed field of $\ker(\overline{\rho}_f)$ in $\overline{\mathbb{Q}}$ is Galois over $\mathbb{Q}$ with Galois group isomorphic to $\mathrm{im}(\overline{\rho}_f)$. For $\ell = 2$ and any λ above ℓ equation (3.1) together with Chebotarev's density theorem imply that $\mathrm{im}(\overline{\rho}_f)$ is contained in $\mathrm{SL}_2(\mathbb{F}_\lambda)$. So to show that there is an extension of $\mathbb{Q}$ with Galois group isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$ it suffices to find an N and a newform $f \in S_2(\Gamma_0(N))$ such that there is a prime λ of degree 4 above 2 in K_f and $\mathrm{im}(\overline{\rho}_f)$ is the full group $\mathrm{SL}_2(\mathbb{F}_\lambda)$. Using modular symbols we can calculate the coefficients of f, hence traces of matrices that occur in the image of $\overline{\rho}_f$. For a survey paper on how this works, see [80]. A subgroup Γ of $\mathrm{SL}_2(\mathbb{F}_{16})$ contains elements of every trace if and only if Γ equals $\mathrm{SL}_2(\mathbb{F}_{16})$; this can be shown in several ways, either by a direct calculation or by invoking a more general classification result like [82, Theorem III.6.25]. With this in mind, after a small computer search in which we check the occurring values of $\mathrm{tr}(\overline{\rho}_f(\mathrm{Frob}_p))$ up to some moderate bound of p, one finds that a suitable modular form f exists in $S_2(\Gamma_0(137))$. It turns out that we have $K_f \cong \mathbb{Q}(\alpha)$ with the minimal polynomial of α equal to $x^4 + 3x^3 - 4x - 1$ and that f is the form whose q-expansion starts with

$$f = q + \alpha q^2 + (\alpha^3 + \alpha^2 - 3\alpha - 2)q^3 + (\alpha^2 - 2)q^4 + \cdots.$$

Now the next question comes in: knowing this modular form, how does one produce a polynomial? In general, one can use the Jacobian $J_0(N)$ to construct $\overline{\rho}_f$. In this particular case we can do that in the following way. We observe that K_f is of degree 4 and that the prime 2 is inert in it. Furthermore we can verify that the subspace of $S_2(\Gamma_0(137))$ fixed by the Atkin-Lehner operator w_{137} is exactly the subspace generated by all the complex conjugates of f. These observations imply that $\overline{\rho}_f$ is isomorphic to the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $\mathrm{Jac}(X_0(137)/\langle w_{137}\rangle)[2]$, where we give this latter space an $\mathbb{F}_{16}$-vector space structure via the action of the Hecke operators. Note that $\mathrm{im}(\overline{\rho}_f) = \mathrm{SL}_2(\mathbb{F}_{16})$ implies surjectivity of the natural map $\mathbb{T} \to \mathscr{O}_{K,f}/(2) \cong \mathbb{F}_{16}$, where $\mathbb{T}$ is the Hecke algebra attached to $S_2(\Gamma_0(N))$. The methods described in [28, Sections 11 & 24] allow us now to give complex approximations

of the 2-torsion points of $\mathrm{Jac}(X_0(137)/\langle w_{137}\rangle)$ to a high precision. This part of the calculation took by far the most effort; the author will write more details about how this works in a future paper (or thesis). We use this to give a real approximation of a polynomial with Galois group isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$. The results from [28, Sections 14 to 19] do, at least implicitly, give a theoretical upper bound for the height of the coefficients of the polynomial hence an upper bound for the calculation precision to get an exact result. Though this upper bound is small in the sense that it leads to a polynomial time algorithm, it is still far too high to be of use in practice. However it turns out that we can use a much smaller precision to obtain our polynomial, the only drawback being that this does not give us a proof of its correctness, so we have to verify this afterwards.

The polynomial P' obtained in this way has coefficients of about 200 digits so we want to find a polynomial of smaller height defining the same number field K. To do this, we first compute the ring of integers $\mathcal{O}_K$ of K. In [11, Section 6] an algorithm to do this is described, provided that one knows the square-free factorisation of $\mathrm{Disc}(f)$ [11, Theorem 1.4] and even if we don't know the square-free factorisation of the discriminant, the algorithm produces a 'good' order in K (see [11, Theorem 1.1]). Assuming that our polynomial P' is correct we know that K is unramified outside $2 \cdot 137$ so we can easily calculate the square-free factorisation of $\mathrm{Disc}(f)$ and hence apply the algorithm. Having done this we obtain an order in K with a discriminant small enough to be able to factor and hence we know that this is indeed the maximal order $\mathcal{O}_K$. Explicitly, the discriminant is equal to

$$\mathrm{Disc}(\mathcal{O}_K) = 2^{30} \cdot 137^8. \tag{3.2}$$

We embed $\mathcal{O}_K$ as a lattice into $\mathbb{C}^{[K:\mathbb{Q}]}$ in the natural way and use lattice basis reduction, see [49, (1.15)], to compute a short vector $\alpha \in \mathcal{O}_K - \mathbb{Z}$. The minimal polynomial of α has small coefficients. In our particular case $[K : \mathbb{Q}]$ is equal to 17, which is a prime number, hence this new polynomial must define the full field K. This method gives us also a way of expressing α as an element of $\mathbb{Q}(x)/(P'(x))$.

3.3 Verification of the Galois group

Now that we have computed a polynomial $P(x)$, we want to verify that its Galois group $\mathrm{Gal}(P)$ is really isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$ and that we can identify the set $\Omega(P)$ of roots of P with $\mathbb{P}^1(\mathbb{F}_{16})$ in such a way that the action of $\mathrm{Gal}(P)$ on $\Omega(P)$ is identified with the action of $\mathrm{SL}_2(\mathbb{F}_{16})$ on $\mathbb{P}^1(\mathbb{F}_{16})$.

For completeness let us remark that it is easy to verify that $P(x)$ is irreducible since it is irreducible modulo 5. The irreducibility of P implies that $\mathrm{Gal}(P)$ is a transitive permutation group of degree 17. The transitive permutation groups of degree 17 have been classified, see for example [75, Section 5]. It follows from [82, Theorem III.6.25] that up to conjugacy there is only one subgroup of index 17 in $\mathrm{SL}_2(\mathbb{F}_{16})$, namely the group of upper triangular matrices. This implies that up to conjugacy there is exactly one transitive $G < S_{17}$ that is isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$. Hence if $\mathrm{Gal}(P) \cong \mathrm{SL}_2(\mathbb{F}_{16})$ is an isomorphism of groups then there

is an identification of $\Omega(P)$ with $\mathbb{P}^1(\mathbb{F}_{16})$ such that the group actions become compatible.

It follows from the classification in [75, Section 5] that if the order of a transitive $G < S_{17}$ is divisible by 5, then G contains a transitive subgroup isomorphic to $\mathrm{SL}_2(\mathbb{F}_{16})$. To show $5 \mid \#\mathrm{Gal}(P)$ we use the fact that for a prime $p \nmid \mathrm{Disc}(P)$ the decomposition type of P modulo p is equal to the cycle type of any Frobenius element in $\mathrm{Gal}(P)$ attached to p. One can verify that modulo 7 the polynomial P has an irreducible factor of degree 15, showing that indeed $5 \mid \#\mathrm{Gal}(P)$ holds, hence $\mathrm{Gal}(P)$ contains $\mathrm{SL}_2(\mathbb{F}_{16})$ as a subgroup.

To show that $\mathrm{Gal}(P)$ cannot be bigger than $\mathrm{SL}_2(\mathbb{F}_{16})$ it seems inevitable to use heavy computer calculations. We will use ideas from [29], in particular we will use [29, Algorithm 6.1], which combines the absolute resolvent method from [76] with an improved version of the relative resolvent method from [77]. It would be interesting to see how $\mathrm{Gal}(P) \cong \mathrm{SL}_2(\mathbb{F}_{16})$ can be proven without using heavy calculations.

Note that the action of $\mathrm{SL}_2(\mathbb{F}_{16})$ on $\mathbb{P}^1(\mathbb{F}_{16})$ is sharply 3-transitive. So first we show that $\mathrm{Gal}(P)$ is not 4-transitive to prove that it does not contain A_{17}. To do this we start with calculating the polynomial

$$Q(x) := \prod_{\{\alpha_1,\alpha_2,\alpha_3,\alpha_4\} \subset \Omega(P)} (X - \alpha_1 - \alpha_2 - \alpha_3 - \alpha_4), \tag{3.3}$$

where the product runs over all subsets of $\{1,\ldots,17\}$ consisting of exactly 4 elements. This implies $\deg(Q) = 2380$. One can calculate $Q(x)$ using symbolic methods [15, Section 2.1]. Suppose that $\mathrm{Gal}(P)$ acting on $\Omega(P)$ is 4-transitive. Then the action on $\Omega(Q)$ is transitive hence if $Q(x)$ is square-free it is irreducible. So if we can show that $Q(x)$ is reducible and square-free, we have shown that $\mathrm{Gal}(P)$ is not 4-transitive.

We have two ways to find a nontrivial factor of $Q(x)$: the first way is use a factorisation algorithm and the second way is to produce a candidate factor ourselves. An algorithm that works very well for our type of polynomial is Van Hoeij's algorithm [31, Section 2.2]. One finds that $Q(x)$ is the product of 3 distinct irreducible polynomials of degrees 340, 1020 and 1020 respectively. A more direct way to produce a candidate factorisation is as follows. The calculation of the 2-torsion in the Jacobian mentioned in Section 3.2 gives a bijection between the set of complex roots of P' and the set $\mathbb{P}^1(\mathbb{F}_{16})$ such that the action of $\mathrm{Gal}(P')$ on $\Omega(P')$ corresponds to the action of $\mathrm{SL}_2(\mathbb{F}_{16})$ on $\mathbb{P}^1(\mathbb{F}_{16})$, assuming the outcome is correct. From the previous section we know how to express the roots of P as rational expressions in the roots of P hence this gives us a bijection between $\Omega(P)$ and $\mathbb{P}^1(\mathbb{F}_{16})$, conjecturally compatible with the group actions of $\mathrm{Gal}(P)$ and $\mathrm{SL}_2(\mathbb{F}_{16})$ respectively. A calculation shows that the action of $\mathrm{SL}_2(\mathbb{F}_{16})$ on the set of unordered four-tuples of elements of $\mathbb{P}^1(\mathbb{F}_{16})$ has 3 orbits, of size 340, 1020 and 1020 respectively. Using approximations to a high precision of the roots, we use these orbits to produce sub-products of (3.3), round off the coefficients to the nearest integer and verify afterwards that the obtained polynomials are indeed factors of $Q(x)$.

Let us remark that the group $\mathrm{SL}_2(\mathbb{F}_{16}).4 := \mathrm{SL}_2(\mathbb{F}_{16}) \rtimes \mathrm{Aut}(\mathbb{F}_{16})$ with its natural action on $\mathbb{P}^1(\mathbb{F}_{16})$ is a transitive permutation group of degree 17, and the same holds for its normal subgroup $\mathrm{SL}_2(\mathbb{F}_{16}).2 := \mathrm{SL}_2(\mathbb{F}_{16}) \rtimes \langle \mathrm{Frob}_2^2 \rangle$. Furthermore, it is well-known that $\mathrm{SL}_2(\mathbb{F}_{16}).4$ is isomorphic to $\mathrm{Aut}(\mathrm{SL}_2(\mathbb{F}_{16}))$ (where $\mathrm{SL}_2(\mathbb{F}_{16})$ acts by conjugation and $\mathrm{Aut}(\mathbb{F}_{16})$ acts on matrix entries) and actually inside S_{17} this group is the normaliser of both $\mathrm{SL}_2(\mathbb{F}_{16})$ and itself. According to the classification of transitive permutation groups of degree 17 in [75, Section 5] these two groups are the only ones that lie strictly between $\mathrm{SL}_2(\mathbb{F}_{16})$ and A_{17}. Once we have fixed $\mathrm{SL}_2(\mathbb{F}_{16})$ inside S_{17}, these two groups are actually unique subgroups of S_{17}, not just up to conjugacy.

From $A_{17} \not< \mathrm{Gal}(P)$ we can thus conclude $\mathrm{Gal}(P) < \mathrm{SL}_2(\mathbb{F}_{16}).4$. To proceed we consult [29, Theorem 2.17], which gives a good computational method to move down over small steps in a lattice of transitive permutation groups. Using this method we can easily go from $\mathrm{Gal}(P) < \mathrm{SL}_2(\mathbb{F}_{16}).4$ to $\mathrm{Gal}(P) < \mathrm{SL}_2(\mathbb{F}_{16}).2$ and from there to $\mathrm{Gal}(P) < \mathrm{SL}_2(\mathbb{F}_{16})$. So indeed we have $\mathrm{Gal}(P) \cong \mathrm{SL}_2(\mathbb{F}_{16})$.

3.4 Does P indeed define $\overline{\rho}_f$?

So now that we have shown $\mathrm{Gal}(P) \cong \mathrm{SL}_2(\mathbb{F}_{16})$ we can wonder whether we can prove that P comes from the modular form f we used to construct it with. Once an isomorphism of $\mathrm{Gal}(P)$ with $\mathrm{SL}_2(\mathbb{F}_{16})$ is given, P defines a representation $\overline{\rho}_P : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{SL}_2(\mathbb{F}_{16})$. Above we mentioned that that $\mathrm{Out}(\mathrm{SL}_2(\mathbb{F}_{16}))$ is isomorphic to $\mathrm{Aut}(\mathbb{F}_{16})$ acting on matrix entries. Hence, up to an automorphism of $\mathbb{F}_{16}$, the map sending $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ to the characteristic polynomial of $\overline{\rho}_P$ in $\mathbb{F}_{16}[x]$ is determined by P and in fact the isomorphism class of $\overline{\rho}_P$ is well-defined up to an automorphism of $\mathbb{F}_{16}$. More concretely, we have to show that the splitting field of P, which we will denote by L, is the fixed field of $\ker(\overline{\rho}_f)$.

A continuous representation $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ has a *level*, denoted by $N(\overline{\rho})$, and a *weight*, denoted by $k(\overline{\rho})$. Instead of repeating the full definitions here, which are lengthy (at least for the weight) and can be found in [70, Sections 1.2 and 2] (see also [27, Section 4] for a discussion on the definition of the weight), we will just say that they are defined in terms of the local representations $\overline{\rho}_p : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ obtained from $\overline{\rho}$. The level is defined in terms of the representations $\overline{\rho}_p$ with $p \neq \ell$ and the weight is defined in terms of $\overline{\rho}_\ell$. The following conjecture is due to Serre:

Conjecture 3.1 (Serre's strong conjecture, [70, Conjecture 3.2.4])**.** *Let ℓ be a prime and let $\overline{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ be a continuous odd irreducible Galois representation (a representation is called odd if the image of a complex conjugation has determinant -1). Then there exists a modular form f of level $N(\overline{\rho})$ and weight $k(\overline{\rho})$ which is a normalised eigenform and a prime $\lambda \mid \ell$ of K_f such that $\overline{\rho}$ and $\overline{\rho}_{f,\lambda}$ become isomorphic after a suitable embedding of $\mathbb{F}_\lambda$ into $\overline{\mathbb{F}}_\ell$.*

In 2006, Khare and Wintenberger proved the following part of Serre's strong conjecture:

Theorem 3.1 (Khare & Wintenberger, [39, Theorem 1.2])**.** *Conjecture 3.1 holds in each of the following cases:*

- $N(\overline{\rho})$ *is odd and* $\ell > 2$.
- $\ell = 2$ *and* $k(\overline{\rho}) = 2$.

With Theorem 3.1 in mind it is sufficient to prove that a representation $\overline{\rho} = \overline{\rho}_P$ attached to P has level 137 and weight 2, which are the level and weight of the modular form f we used to construct it with and that of all eigenforms in $S_2(\Gamma_1(137))$, the form f is one which gives rise to $\overline{\rho}_P$. Therefore, in the remainder of this section we will verify the following proposition.

Proposition 3.2. *Let f be the cusp form from Section 3.2. Up to an automorphism of $\mathbb{F}_{16}$, the representations $\overline{\rho}_P$ and $\overline{\rho}_{f,(2)}$ are isomorphic. In particular, the representation $\overline{\rho}_P$ has Serre-level* 137 *and Serre-weight* 2.

Let us argue that it is not clear how to prove the modularity of $\overline{\rho}_P$ using only results that are older than Theorem 3.1. The older results deal with cases that are 'small' in some sense. For example, [55, Thms 1 & 2] deal with $\overline{\rho}$ that satisfy $N(\overline{\rho}) = 1$ or $k(\overline{\rho}) = 1$ and focus on proving *non-existence* of Galois representations. Also, the group $\mathrm{SL}_2(\mathbb{F}_{16})$ is too big to apply other results. It is a non-solvable group and in that case there are some old results dealing with $\mathrm{im}\,\overline{\rho} \subset \mathrm{GL}_2(\mathbb{F}_q)$ for $q \in \{2^2, 3^2, 5, 7\}$, but not for $q = 16$ (see [37, Section 1.3] for a survey). Neither is it clear how to do a computer search of whichever kind that will eliminate the possibility that $\overline{\rho}_P$ is not isomorphic to $\overline{\rho}_{f,(2)}$, as the group $\mathrm{SL}_2(\mathbb{F}_{16})$ and the degree 17 are simply too big.

3.4.1 Verification of the level

The level is the easiest of the two to verify. Here we have to do local computations in p-adic fields with $p \neq 2$. According to the definition of $N(\overline{\rho})$ in [70, Section 1.2] it suffices to verify that $\overline{\rho}$ is unramified outside 2 and 137, tamely ramified at 137 and the local inertia subgroup I at 137 leaves exactly one line of $\mathbb{F}_{16}^2$ point-wise fixed. That $\overline{\rho}_P$ is unramified outside 2 and 137 follows immediately from (3.2).

From (3.2) and the fact that $137^8 \| \mathrm{Disc}(P)$ it follows that the monogenous order defined by P is maximal at 137. Modulo 137, the polynomial P factors as

$$\overline{P} = (x+14)(x^2+6x+101)^2(x^2+88x+97)^2(x^2+106x+112)^2(x^2+133x+110)^2$$

into irreducibles. Let v be any prime above 137 in L. From the above factorisation it follows that the prime 137 decomposes in K as a product of 5 primes; one of them has its inertial and ramification degree equal to 1 and the other four ones have their inertial and ramification degrees equal to 2. Thus $\deg(v)$ is a power of 2, as L is obtained by successively adjoining roots of P and in each step the relative inertial and ramification degrees of the prime below v are both at most 2. In particular, $\mathrm{Gal}(L_v/\mathbb{Q}_{137})$ is a subgroup of $\mathrm{SL}_2(\mathbb{F}_{16})$ whose order is a power of 2. Now, $\{\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)\}$ is a Sylow 2-subgroup of $\mathrm{SL}_2(\mathbb{F}_{16})$, so $\mathrm{Gal}(L_v/\mathbb{Q}_{137})$ is, up to

conjugacy, a subgroup of $\{\left(\begin{smallmatrix}1 & *\\ 0 & 1\end{smallmatrix}\right)\}$. Hence I is also conjugate to a subgroup of $\{\left(\begin{smallmatrix}1 & *\\ 0 & 1\end{smallmatrix}\right)\}$ and it is actually nontrivial because 137 ramifies in L (so I is of order 2 since the tame inertia group of any finite Galois extension of local fields is cyclic).

It is immediate that $\overline{\rho}$ is tamely ramified at 137 as no power of 2 is divisible by 137. Also, it is clear that I leaves exactly one line of $\mathbb{F}_{16}^2$ point-wise fixed since $\{\binom{*}{0}\}$ is the only point-wise fixed line of any nontrivial element of $\{\left(\begin{smallmatrix}1 & *\\ 0 & 1\end{smallmatrix}\right)\}$. This establishes the verification of $N(\overline{\rho}) = 137$.

3.4.2 Verification of the weight

Because the weight is defined in terms of the induced local representation $\overline{\rho}_2$, we will try to compute some relevant properties of the splitting field L_v of P over $\mathbb{Q}_2$, where v is any place of L above 2. In p-adic fields one can only do calculations with a certain precision, but this does not give any problems since practically all properties one needs to know can be verified rigorously using a bounded precision calculation and the error bounds in the calculations can be kept track of exactly.

The polynomial P does not define an order which is maximal at the prime 2. Instead we use the polynomial

$$\begin{aligned}R = {} & x^{17} - 11x^{16} + 64x^{15} - 322x^{14} + 916x^{13} + 276x^{12} - 5380x^{11} + 2748x^{10}\\ & + 6904x^9 - 23320x^8 + 131500x^7 - 140744x^6 - 16288x^5 - 39752x^4\\ & - 48840x^3 + 102352x^2 + 234466x - 1518,\end{aligned}$$

which is the minimal polynomial of

$$\begin{aligned}&\big(36863 + 22144\alpha + 123236\alpha^2 + 154875\alpha^3 - 416913\alpha^4 + 436074\alpha^5\\ &+ 229905\alpha^6 - 1698406\alpha^7 + 1857625\alpha^8 - 467748\alpha^9 - 2289954\alpha^{10}\\ &+ 2838473\alpha^{11} - 1565993\alpha^{12} + 605054\alpha^{13} - 263133\alpha^{14} + 112104\alpha^{15}\\ &- 22586\alpha^{16}\big)/8844,\end{aligned}$$

where α is a root of P. We can factor R over $\mathbb{Q}_2$ and see that it has one root in $\mathbb{Q}_2$ which happens to be odd, and an Eisenstein factor of degree 16, which we will call E. This type of decomposition can be read off from the Newton polygon of R and it also shows that the order defined by R is indeed maximal at 2. From the oddness of the root and (3.2) we see

$$v_2(\mathrm{Disc}(E)) = 30. \tag{3.4}$$

For the action of $\mathrm{Gal}(\overline{\mathbb{Q}}_2/\mathbb{Q}_2)$ on $\mathbb{P}^1(\mathbb{F}_{16})$ the factorisation means that there is one fixed point and one orbit of degree 16. If we adjoin a root β of E to $\mathbb{Q}_2$ and factor E over $\mathbb{Q}_2(\beta)$ then we see that it has an irreducible factor of degree 15; in [14, Section 6] one can find methods for factorisation and irreducibility testing that can be used to verify this. This means that

$[L_v : \mathbb{Q}_2]$ is at least 240.

A subgroup of $\mathrm{SL}_2(\mathbb{F}_{16})$ that fixes a point of $\mathbb{P}^1(\mathbb{F}_{16})$ has to be conjugate to a subgroup of the group

$$H := \left\{ \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \right\} \subset \mathrm{SL}_2(\mathbb{F}_{16}),$$

which is the stabiliser subgroup of $[\binom{*}{0}]$. But we have $\#H = 240$ so $\mathrm{Gal}(L_v/\mathbb{Q}_2)$ is isomorphic to H and from now on we will identify these two groups with each other. We can filter H by normal subgroups:

$$H \supset I \supset I_2 \supset \{e\},$$

where I is the inertia subgroup and I_2 is the wild ramification subgroup, which is the unique Sylow 2-subgroup of I. We wish to determine the groups I and I_2. Let $k(v)$ be the residue class field of L_v. The group H/I is isomorphic to $\mathrm{Gal}(k(v)/\mathbb{F}_2)$ and I/I_2 is isomorphic to a subgroup of $k(v)^*$. In particular $[I : I_2] \mid (2^{[H:I]} - 1)$ follows. The group H has the nice property

$$[H,H] = \left\{ \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \right\} \cong \mathbb{F}_{16},$$

which is its unique Sylow 2-subgroup. As H/I is abelian, we see that $[H,H] \subset I$. We conclude that $I_2 = [H,H]$, since above we remarked that I_2 is the unique Sylow 2-subgroup of I. The restriction $[I : I_2] \mid (2^{[H:I]} - 1)$ leaves only one possibility for I, namely $I = I_2$.

Let L_v' be the subextension of $L_v/\mathbb{Q}_2$ fixed by I. Then L_v' is the maximal unramified subextension as well as the maximal tamely ramified subextension. It is in fact isomorphic to $\mathbb{Q}_{2^{15}}$, the unique unramified extension of $\mathbb{Q}_2$ of degree 15 and the Eisenstein polynomial E from above, being irreducible over any unramified extension of $\mathbb{Q}_2$, is a defining polynomial for the extension $L_v/\mathbb{Q}_{2^{15}}$. According to [55, Theorem 3] we can relate the discriminant of L_v to $k(\overline{\rho})$ as follows:

$$v_2(\mathrm{Disc}(L_v)) = \begin{cases} 240 \cdot \frac{15}{8} = 450 & \text{if } k(\overline{\rho}) = 2 \\ 240 \cdot \frac{19}{8} = 570 & \text{if } k(\overline{\rho}) \neq 2 \end{cases}$$

It follows from (3.4) that $v_2(\mathrm{Disc}(L_v/\mathbb{Q}_2)) = 30 \cdot 15 = 450$, so indeed $k(\overline{\rho}) = 2$.

3.4.3 Verification of the form f

Now we know $N(\overline{\rho}_P) = 137$ and $k(\overline{\rho}_P) = 2$, Theorem 3.1 shows that there is an eigenform $g \in S_2(\Gamma_1(137))$ giving rise to $\overline{\rho}_P$. Using [12, Corollary 2.7] we see that if such a g exists, then there actually exists such a g of trivial Nebentypus, i.e. $g \in S_2(\Gamma_0(137))$ (as $\mathrm{SL}_2(\mathbb{F}_{16})$ is non-solvable $\overline{\rho}_P$ cannot be an induced Hecke character from $\mathbb{Q}(i)$).

A modular symbols calculation shows that there exist two Galois orbits of newforms in $S_2(\Gamma_0(137))$: the form f we used for our calculations and another form, g say. The prime 2 decomposes in K_g as a product $\lambda^3\mu$, where λ has inertial degree 1 and μ has inertial degree 4. So it could be that $g \bmod \mu$ gives rise to $\overline{\rho}_P$. We will show now that $f \bmod (2)$ and $g \bmod \mu$ actually give the same representation. The completions of $\mathscr{O}_{K_f}$ and $\mathscr{O}_{K_g}$ at the

primes (2) and μ respectively are both isomorphic to $\mathbb{Z}_{16}$, the unramified extension of $\mathbb{Z}_2$ of degree 4. After a choice of embeddings of $\mathcal{O}_{K_f}$ and $\mathcal{O}_{K_g}$ into $\mathbb{Z}_{16}$ we obtain two modular forms f' and g' with coefficients in $\mathbb{Z}_{16}$ and we wish to show that a suitable choice of embeddings exists such that they are congruent modulo 2. According to [81, Theorem 1], it suffices to check there is a suitable choice of embeddings that gives $a_n(f') \equiv a_n(g') \bmod 2$ for all $n \leq [\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(137)]/6 = 23$ (in [81] this theorem is formulated for modular forms with coefficients in the ring of integers of a number field, but the proof also works for p-adic rings). Using a modular symbols calculation, this can be easily verified. The bound on the indices up to which one has to check such a congruence is usually referred to as the *Sturm bound* or *Hecke bound*.

3.5 MAGMA code used for computations

All the calculations were done using MAGMA (see [6]); for most of them the author used the MEDICIS cluster (`http://medicis.polytechnique.fr`). The MAGMA code used for the computation of the polynomials, together with a short instruction on how to use it, has been included as an add-on to this paper and may be found at

`http://www.lms.ac.uk/jcm/10/lms2007-024/appendix-a`

Acknowledgements. I would like to thank Jürgen Klüners for proposing this computational challenge and explaining some computational Galois theory to me. Furthermore I want to thank Bas Edixhoven for teaching me about modular forms and the calculation of their coefficients. Thanks also go to David Roberts, for making me aware of the small root discriminant problem and the fact that this polynomial provides an example for it. For being able to make use of the MEDICIS cluster I want to thank Marc Giusti and Pierre Lafon.

Chapter 4

Some polynomials for level one forms

The contents of this chapter will appear in the final version of the manuscript [28] that will eventually be published as a volume of the Annals of Mathematics Studies.

4.1 Introduction

In this chapter we explicitly compute mod ℓ Galois representations associated to modular forms. To be precise, we look at cases with $\ell \leq 23$ and the modular forms considered will be cusp forms of level 1 and weight up to 22. We present the result in terms of polynomials associated to the projectivised representations. As an application, we will improve a known result on Lehmer's non-vanishing conjecture for Ramanujan's tau function (see [47, p. 429]) .

To fix a notation, for any $k \in \mathbb{Z}$ satisfying $\dim S_k(\mathrm{SL}_2(\mathbb{Z})) = 1$ we will denote the unique normalised cusp form in $S_k(\mathrm{SL}_2(\mathbb{Z}))$ by Δ_k. We will denote the coefficients of the q-expansion of Δ_k by $\tau_k(n)$:

$$\Delta_k(z) = \sum_{n \geq 1} \tau_k(n) q^n \in S_k(\mathrm{SL}_2(\mathbb{Z})).$$

From $\dim S_k(\mathrm{SL}_2(\mathbb{Z})) = 1$ it follows that the numbers $\tau_k(n)$ are integers. For every Δ_k and every prime ℓ there is a continuous representation

$$\rho_{\Delta_k,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\mathbb{F}_\ell)$$

such that for every prime $p \neq \ell$ we have that the characteristic polynomial of $\rho_{\Delta_k,\ell}(\mathrm{Frob}_p)$ is congruent to $X^2 - \tau_k(p)X + p^{k-1}$ mod ℓ. For a summary on the exceptional representations $\rho_{\Delta_k,\ell}$ and the corresponding congruences for $\tau_k(n)$, see [83].

4.1.1 Notational conventions

Throughout this chapter, for every field K we will fix an algebraic closure $\overline{K}$ and all algebraic extension fields of K will be regarded as subfields of $\overline{K}$. Furthermore, for each prime number p we will fix an embedding $\overline{\mathbb{Q}} \hookrightarrow \overline{\mathbb{Q}}_p$ and hence an embedding $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \hookrightarrow \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$,

whose image we call D_p. We will use I_p to denote the inertia subgroup of $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)$.

For any field K, a linear representation $\rho : G \to \mathrm{GL}_n(K)$ defines a projective representation $\tilde{\rho} : G \to \mathrm{PGL}_n(K)$ via the canonical map $\mathrm{GL}_n(K) \to \mathrm{PGL}_n(K)$. We say that a projective representation $\tilde{\rho} : G \to \mathrm{PGL}_n(K)$ is *irreducible* if the induced action of G on $\mathbb{P}^{n-1}(K)$ fixes no proper subspace. So for $n = 2$ this means that every point of $\mathbb{P}^1(K)$ has its stabiliser subgroup not equal to G. Representations are assumed to be continuous.

4.1.2 Statement of results

Theorem 4.1. *For every pair (k,ℓ) occurring in Table 4.1 on page 87, let the polynomial $P_{k,\ell}$ be defined as in that same table. Then the splitting field of each $P_{k,\ell}$ is the fixed field of* $\mathrm{Ker}(\tilde{\rho}_{\Delta_k,\ell})$ *and has Galois group* $\mathrm{PGL}_2(\mathbb{F}_\ell)$. *Furthermore, if $\alpha \in \overline{\mathbb{Q}}$ is a root of $P_{k,\ell}$ then the subgroup of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ *fixing α corresponds via $\tilde{\rho}_{\Delta_k,\ell}$ to a subgroup of* $\mathrm{PGL}_2(\mathbb{F}_\ell)$ *fixing a point of* $\mathbb{P}^1(\mathbb{F}_\ell)$.

For completeness we also included the pairs (k,ℓ) for which $\rho_{k,\ell}$ is isomorphic to the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the ℓ-torsion of an elliptic curve. These are the pairs in Table 4.1 with $\ell = k-1$, as there the representation is the ℓ-torsion of $J_0(\ell)$, which happens to be an elliptic curve for $\ell \in \{11,17,19\}$. A simple calculation with division polynomials [46, Chapter II] can be used to treat these cases. In the general case, one has to work in the more complicated Jacobian variety $J_1(\ell)$, which has dimension 12 for $\ell = 23$ for instance.

We can apply Theorem 4.1 to verify the following result.

Corollary 4.1. *The non-vanishing of $\tau(n)$ holds for all*

$$n < 22798241520242687999 \approx 2 \cdot 10^{19}.$$

In [34], the non-vanishing of $\tau(n)$ was verified for all

$$n < 22689242781695999 \approx 2 \cdot 10^{16}.$$

To compute the polynomials, the author used a weakened version of algorithms described elsewhere in this book. After a suggestion of Couveignes, Complex approximations were used. We worked directly in $X_1(\ell)$ rather than $X_1(5\ell)_{\mathbb{Q}(\zeta_\ell)}$ and we guessed the rational coefficients of our polynomials using lattice reduction techniques [49, Proposition 1.39]. instead of computing the height first. Also reduction techniques were used to make the coefficients smaller [16]; after the initial computations some of the polynomials had coefficients of almost 2000 digits. The used algorithms do not give a proven output, so we have to concentrate on the verification. We will show how to verify the correctness of the polynomials in Section 4.3 after setting up some preliminaries about Galois representations in Section 4.2. In Section 4.4 we will point out how to use Theorem 4.1 in a calculation that verifies Corollary 4.1. All the calculations were performed using MAGMA (see [6]).

4.2 Galois representations

This section will be used to state some results on Galois representations that we will need in the proof of Theorem 4.1.

4.2.1 Liftings of projective representations

Let G be a topological group, let K be a topological field and let $\tilde{\rho} : G \to \mathrm{PGL}_n(K)$ be a projective representation. Let L be an extension field of K. By a *lifting* of $\tilde{\rho}$ over L we shall mean a representation $\rho : G \to \mathrm{GL}_n(L)$ that makes the following diagram commute:

$$\begin{array}{ccc} G & \xrightarrow{\tilde{\rho}} & \mathrm{PGL}_n(K) \\ {\scriptstyle\rho}\downarrow & & \downarrow \\ \mathrm{GL}_n(L) & \twoheadrightarrow & \mathrm{PGL}_n(L) \end{array}$$

where the maps on the bottom and the right are the canonical ones. If the field L is not specified then by a lifting of $\tilde{\rho}$ we shall mean a lifting over $\overline{K}$.

An important theorem of Tate arises in the context of liftings. For the proof we refer to [66, Section 6]. Note that in the reference representations over $\mathbb{C}$ are considered, but the proof works for representations over arbitrary algebraically closed fields.

Theorem 4.2 (Tate)**.** *Let K be a field and let $\tilde{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_n(K)$ be a projective representation. For each prime number p, there exists a lifting $\rho'_p : D_p \to \mathrm{GL}_n(\overline{K})$ of $\tilde{\rho}|_{D_p}$. Assume that these liftings ρ'_p have been chosen so that all but finitely many of them are unramified. Then there is a unique lifting $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_n(\overline{K})$ such that for all primes p we have*

$$\rho|_{I_p} = \rho'_p|_{I_p}.$$

Lemma 4.1. *Let p be a prime number and let K be a field. Suppose that we are given a projective representation $\tilde{\rho}_p : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathrm{PGL}_n(K)$ that is unramified. Then there exists a lifting $\rho_p : \mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathrm{GL}_n(\overline{K})$ of $\tilde{\rho}_p$ that is unramified as well.*

Proof. Since $\tilde{\rho}$ is unramified, it factors through $\mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p) \cong \hat{\mathbb{Z}}$ and is determined by the image of $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p)$. By continuity, this image is an element of $\mathrm{PGL}_n(K)$ of finite order, say of order m. If we take any lift F of $\tilde{\rho}(\mathrm{Frob}_p)$ to $\mathrm{GL}_n(K)$ then we have $F^m = a$ for some $a \in K^\times$. So $F' := \alpha^{-1}F$, where $\alpha \in \overline{K}$ is any m-th root of a, has order m in $\mathrm{GL}_n(\overline{K})$. Hence the homomorphism $\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \to \mathrm{GL}_n(\overline{K})$ obtained by the composition

$$\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p) \twoheadrightarrow \mathrm{Gal}(\overline{\mathbb{F}}_p/\mathbb{F}_p) \xrightarrow{\sim} \hat{\mathbb{Z}} \twoheadrightarrow \mathbb{Z}/m\mathbb{Z} \xrightarrow{1\mapsto F'} \mathrm{GL}_n(\overline{K})$$

lifts $\tilde{\rho}$ and is continuous as well as unramified. □

4.2.2 Serre invariants and Serre's conjecture

Let ℓ be a prime. A Galois representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ has a *level* $N(\rho)$ and a *weight* $k(\rho)$. The definitions were introduced by Serre (see [70, Sections 1.2 & 2]). Later on, Edixhoven found an improved definition for the weight, which is the one we will use, see [27, Section 4]. The level $N(\rho)$ is defined as the prime-to-ℓ part of the Artin conductor of ρ and equals 1 if ρ is unramified outside ℓ. The weight is defined in terms of the local representation $\rho|_{D_\ell}$; its definition is rather lengthy so we will not write it out here. When we need results about the weight we will just state them. Let us for now mention that one can consider the weights of the twists $\rho \otimes \chi$ of a representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ by a character $\chi : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \overline{\mathbb{F}}_\ell^\times$. If one chooses χ so that $k(\rho \otimes \chi)$ is minimal, then we always have $1 \leq k(\rho \otimes \chi) \leq \ell + 1$ and we can in fact choose our χ to be a power of the mod ℓ cyclotomic character.

Serre conjectured [70, Conjecture 3.2.4] that if ρ is irreducible and odd, then ρ belongs to a modular form of level $N(\rho)$ and weight $k(\rho)$. Oddness here means that the image of a complex conjugation has determinant -1. A proof of this conjecture in the case $N(\rho) = 1$ has been published by Khare and Wintenberger:

Theorem 4.3 (Khare & Wintenberger, [38, Theorem 1.1])**.** *Let ℓ be a prime number and let $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ be an odd irreducible representation of level $N(\rho) = 1$. Then there exists a modular form f of level 1 and weight $k(\rho)$ which is a normalised eigenform and a prime $\lambda \mid \ell$ of K_f such that ρ and $\rho_{f,\lambda}$ become isomorphic after a suitable embedding of $\mathbb{F}_\lambda$ into $\overline{\mathbb{F}}_\ell$.*

4.2.3 Weights and discriminants

If a representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ is wildly ramified at ℓ it is possible to relate the weight to discriminants of certain number fields. In this subsection we will present a theorem of Moon and Taguchi on this matter and derive some results from it that are of use to us.

Theorem 4.4 (Moon & Taguchi, [55, Theorem 3])**.** *Consider a wildly ramified representation $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$. Let $\alpha \in \mathbb{Z}$ be such that $k(\rho \otimes \chi_\ell^{-\alpha})$ is minimal where $\chi_\ell : \mathrm{Gal}(\overline{\mathbb{Q}}_\ell/\mathbb{Q}_\ell) \to \mathbb{F}_\ell^\times$ is the mod ℓ cyclotomic character. Put $\tilde{k} = k(\rho \otimes \chi_\ell^{-\alpha})$, put $d = \gcd(\alpha, \tilde{k}-1, \ell-1)$ and put $K = \overline{\mathbb{Q}}_\ell^{\mathrm{Ker}(\rho)}$. Define $m \in \mathbb{Z}$ by letting ℓ^m be the wild ramification degree of K over $\mathbb{Q}_\ell$. Then we have*

$$v_\ell(\mathscr{D}_{K/\mathbb{Q}_\ell}) = \begin{cases} 1 + \frac{\tilde{k}-1}{\ell-1} - \frac{\tilde{k}-1+d}{(\ell-1)\ell^m} & \text{if } 2 \leq \tilde{k} \leq \ell, \\ 2 + \frac{1}{(\ell-1)\ell} - \frac{2}{(\ell-1)\ell^m} & \text{if } \tilde{k} = \ell+1, \end{cases}$$

where $\mathscr{D}_{K/\mathbb{Q}_\ell}$ denotes the different of K over $\mathbb{Q}_\ell$ and v_ℓ is normalised by $v_\ell(\ell) = 1$.

We can simplify this formula to one which is useful in our case. In the proof of the following corollaries, v_ℓ denotes a valuation at a prime above ℓ normalised by $v_\ell(\ell) = 1$.

Corollary 4.2. *Let $\tilde{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_2(\mathbb{F}_\ell)$ be an irreducible projective representation that is wildly ramified at ℓ. Take a point in $\mathbb{P}^1(\mathbb{F}_\ell)$, let $H \subset \mathrm{PGL}_2(\mathbb{F}_\ell)$ be its stabiliser subgroup and let K be the number field defined as*

$$K = \overline{\mathbb{Q}}^{\tilde{\rho}^{-1}(H)}.$$

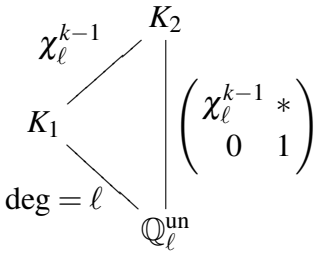
Then the ℓ-primary part of $\mathrm{Disc}(K/\mathbb{Q})$ is related to the minimal weight k of the liftings of $\tilde{\rho}$ by the following formula:

$$v_\ell(\mathrm{Disc}(K/\mathbb{Q})) = k + \ell - 2.$$

Proof. Let ρ be a lifting of $\tilde{\rho}$ of minimal weight. Since ρ is wildly ramified, after a suitable conjugation in $\mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ we may assume

$$\rho|_{I_\ell} = \begin{pmatrix} \chi_\ell^{k-1} & * \\ 0 & 1 \end{pmatrix}, \tag{4.1}$$

where $\chi_\ell : I_\ell \to \mathbb{F}_\ell^\times$ denotes the mod ℓ cyclotomic character; this follows from the definition of weight. The canonical map $\mathrm{GL}_2(\overline{\mathbb{F}}_\ell) \to \mathrm{PGL}_2(\overline{\mathbb{F}}_\ell)$ is injective on the subgroup $\left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right)$, so the subfields of $\overline{\mathbb{Q}}_\ell$ cut out by $\rho|_{I_\ell}$ and $\tilde{\rho}|_{I_\ell}$ are equal, call them K_2. Also, let $K_1 \subset K_2$ be the fixed field of the diagonal matrices in $\mathrm{Im}\,\rho|_{I_\ell}$. We see from (4.1) that in the notation of Theorem 4.4 we can put $\alpha = 0$, $m = 1$ and $d = \gcd(\ell-1, k-1)$. So we have the following diagram of field extensions:



The extension K_2/K_1 is tamely ramified of degree $(\ell-1)/d$ hence we have

$$v_\ell(\mathscr{D}_{K_2/K_1}) = \frac{(\ell-1)/d - 1}{(\ell-1)\ell/d} = \frac{\ell - 1 - d}{(\ell-1)\ell}.$$

Consulting Theorem 4.4 for the case $2 \le k \le \ell$ now yields

$$\begin{aligned} v_\ell(\mathscr{D}_{K_1/\mathbb{Q}_\ell^{\mathrm{un}}}) &= v_\ell(\mathscr{D}_{K_2/\mathbb{Q}_\ell^{\mathrm{un}}}) - v_\ell(\mathscr{D}_{K_2/K_1}) \\ &= 1 + \frac{k-1}{\ell-1} - \frac{k-1+d}{(\ell-1)\ell} - \frac{\ell-1-d}{(\ell-1)\ell} = \frac{k+\ell-2}{\ell} \end{aligned}$$

and also in the case $k = \ell + 1$ we get

$$v_\ell(\mathscr{D}_{K_1/\mathbb{Q}_\ell^{\mathrm{un}}}) = 2 + \frac{1}{(\ell-1)\ell} - \frac{2}{(\ell-1)\ell} - \frac{\ell-2}{(\ell-1)\ell} = \frac{k+\ell-2}{\ell}.$$

Let L be the number field $\overline{\mathbb{Q}}^{\mathrm{Ker}(\tilde{\rho})}$. From the irreducibility of $\tilde{\rho}$ and the fact that $\mathrm{Im}\,\tilde{\rho}$ has an element of order ℓ it follows that the induced action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on $\mathbb{P}^1(\mathbb{F}_\ell)$ is transitive

and hence that L is the normal closure of K in $\overline{\mathbb{Q}}$. This in particular implies that $K/\mathbb{Q}$ is wildly ramified. Now from $[K:\mathbb{Q}] = \ell+1$ it follows that there are two primes in K above ℓ: one is unramified and the other has inertia degree 1 and ramification degree ℓ. From the considerations above it now follows that any ramification subgroup of $\mathrm{Gal}(L/\mathbb{Q})$ at ℓ is isomorphic to a subgroup of $\left(\begin{smallmatrix} * & * \\ 0 & 1 \end{smallmatrix}\right) \subset \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ of order $(\ell-1)\ell/d$ with $d \mid \ell-1$. Up to conjugacy, the only subgroup of index ℓ is the subgroup of diagonal matrices. Hence K_1 and $K_{\lambda_2}^{\mathrm{un}}$ are isomorphic field extensions of $\mathbb{Q}_\ell^{\mathrm{un}}$, from which

$$v_\ell(\mathrm{Disc}(K/\mathbb{Q})) = v_\ell(\mathrm{Disc}(K_1/\mathbb{Q}_\ell^{\mathrm{un}})) = \ell \cdot v_\ell(\mathscr{D}_{K_1/\mathbb{Q}_\ell^{\mathrm{un}}}) = k+\ell-2.$$

follows. □

Corollary 4.3. *Let $\tilde{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_2(\mathbb{F}_\ell)$ be an irreducible projective representation and let ρ be a lifting of $\tilde{\rho}$ of minimal weight. Let K be the number field belonging to a point of $\mathbb{P}^1(\mathbb{F}_\ell)$, as in the notation of Corollary 4.2. If $k \geq 3$ is such that*

$$v_\ell(\mathrm{Disc}(K/\mathbb{Q})) = k+\ell-2$$

holds, then we have $k(\rho) = k$.

Proof. From $v_\ell(\mathrm{Disc}(K/\mathbb{Q})) = k+\ell-2 \geq \ell+1$ it follows that $\tilde{\rho}$ is wildly ramified at ℓ so we can apply Corollary 4.2. □

4.3 Proof of the theorem

To prove Theorem 4.1 we need to do several verifications. We will derive representations from the polynomials $P_{k,\ell}$ and verify that they satisfy the conditions of Theorem 4.3. Then we know there are modular forms attached to them that have the right level and weight and uniqueness follows then easily.

First we we will verify that the polynomials $P_{k,\ell}$ from Table 4.1 have the right Galois group. The algorithm described in [29, Algorithm 6.1] can be used perfectly to do this verification; proving $A_{\ell+1} \not\leq \mathrm{Gal}(P_{k,\ell})$ is the most time-consuming part of the calculation here. It turns out that in all cases we have

$$\mathrm{Gal}(P_{k,\ell}) \cong \mathrm{PGL}_2(\mathbb{F}_\ell). \tag{4.2}$$

That the action of $\mathrm{Gal}(P_{k,\ell})$ on the roots of $P_{k,\ell}$ is compatible with the action of $\mathrm{PGL}_2(\mathbb{F}_\ell)$ follows from the following well-known lemma:

Lemma 4.2. *Let ℓ be a prime and let G be a subgroup of $\mathrm{PGL}_2(\mathbb{F}_\ell)$ of index $\ell+1$. Then G is the stabiliser subgroup of a point in $\mathbb{P}^1(\mathbb{F}_\ell)$. In particular any transitive permutation representation of $\mathrm{PGL}_2(\mathbb{F}_\ell)$ of degree $\ell+1$ is isomorphic to the standard action on $\mathbb{P}^1(\mathbb{F}_\ell)$.*

Proof. This follows from [82, Proof of Theorem 6.25]. □

So now we have shown that the second assertion in Theorem 4.1 follows from the first one.

Next we will verify that we can obtain representations from this that have the right Serre invariants. Let us first note that the group $\mathrm{PGL}_2(\mathbb{F}_\ell)$ has no outer automorphisms. This implies that for every $P_{k,\ell}$, two isomorphisms as in (4.2) define isomorphic representations $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_2(\mathbb{F}_\ell)$ via composition with the canonical map $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \twoheadrightarrow \mathrm{Gal}(P_{k,\ell})$. In other words, every $P_{k,\ell}$ gives a projective representation $\tilde{\rho} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{PGL}_2(\mathbb{F}_\ell)$ that is well-defined up to isomorphism.

Now, for each (k,ℓ) in Table 4.1, the polynomial $P_{k,\ell}$ is irreducible and hence defines a number field

$$K_{k,\ell} := \mathbb{Q}[x]/(P_{k,\ell}),$$

whose ring of integers we will denote by $\mathscr{O}_{k,\ell}$. It is possible to compute $\mathscr{O}_{k,\ell}$ using the algorithm from [11, Section 6] (see also [11, Theorems 1.1 & 1.4]), since we know what kind of ramification behaviour to expect. In all cases it turns out that we have

$$\mathrm{Disc}(K_{k,\ell}/\mathbb{Q}) = (-1)^{(\ell-1)/2}\ell^{k+\ell-2}.$$

We see that for each (k,ℓ) the representation $\tilde{\rho}_{k,\ell}$ is unramified outside ℓ. From Lemma 4.1 it follows that for each $p \neq \ell$, the representation $\tilde{\rho}_{k,\ell}|_{\mathrm{Gal}(\overline{\mathbb{Q}}_p/\mathbb{Q}_p)}$ has an unramified lifting. Above we saw that via $\tilde{\rho}_{k,\ell}$ the action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the set of roots of $P_{k,\ell}$ is compatible with the action of $\mathrm{PGL}_2(\mathbb{F}_\ell)$ on $\mathbb{P}^1(\mathbb{F}_\ell)$, hence we can apply Corollary 4.3 to show that the minimal weight of a lifting of $\tilde{\rho}_{k,\ell}$ equals k. Theorem 4.2 now shows that every $\tilde{\rho}_{k,\ell}$ has a lifting $\rho_{k,\ell}$ that has level 1 and weight k. From $\mathrm{Im}\,\tilde{\rho}_{k,\ell} = \mathrm{PGL}_2(\mathbb{F}_\ell)$ it follows that each $\rho_{k,\ell}$ is absolutely irreducible.

To apply Theorem 4.3 we should still verify that $\rho_{k,\ell}$ is odd. Let (k,ℓ) be given and suppose $\rho_{k,\ell}$ is even. Then a complex conjugation $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is sent to a matrix $M \in \mathrm{GL}_2(\overline{\mathbb{F}}_\ell)$ of determinant 1 and of order 2. Because ℓ is odd, this means $M = \pm 1$ so the image of M in $\mathrm{PGL}_2(\mathbb{F}_\ell)$ is the identity. It follows now that $K_{k,\ell}$ is totally real. One could arrive at a contradiction by approximating the roots of $P_{k,\ell}$ to a high precision, but to get a proof one should use only symbolic calculations. The fields $K_{k,\ell}$ with $\ell \equiv 3 \bmod 4$ have negative discriminant hence cannot be totally real. Now suppose that a polynomial $P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ has only real roots. Then $a_{n-1}^2 - 2a_{n-2}$, being the sum of the squares of the roots, is non-negative and for a similar reason $a_1^2 - 2a_0a_2$ is non-negative as well. One can verify immediately that each of the polynomials $P_{k,\ell}$ with $\ell \equiv 1 \bmod 4$ fails at least one of these two criteria, hence none of the fields $K_{k,\ell}$ is totally real. This proves the oddness of the representations $\rho_{k,\ell}$. Of course, this can also be checked with more general methods, like considering the trace pairing on $K_{k,\ell}$ or invoking Sturm's theorem [32, Theorem 5.4].

So now that we have verified all the conditions of Theorem 4.3 we remark as a final step that all spaces of modular forms $S_k(\mathrm{SL}_2(\mathbb{Z}))$ involved here are 1-dimensional. So the modularity of each $\rho_{k,\ell}$ implies immediately the isomorphism $\rho_{k,\ell} \cong \rho_{\Delta_k,\ell}$, hence also $\tilde{\rho}_{k,\ell} \cong \tilde{\rho}_{\Delta_k,\ell}$, which completes the proof of Theorem 4.1.

4.4 Proof of the corollary

If τ vanishes somewhere, then the smallest positive integer n for which $\tau(n)$ is zero is a prime (see [47, Theorem 2]). Using results on the exceptional representations for $\tau(p)$, Serre pointed out [68, Section 3.3] that if p is a prime number with $\tau(p) = 0$ then p can be written as

$$p = hM - 1$$

with

$$M = 2^{14}3^7 5^3 691 = 3094972416000,$$
$$\left(\frac{h+1}{23}\right) = 1 \quad \text{and} \quad h \equiv 0, 30 \text{ or } 48 \bmod 49.$$

In fact p is of this form if and only if $\tau(p) \equiv 0 \bmod 23 \cdot 49 \cdot M$ holds. Knowing this, we will do a computer search on these primes p and verify whether $\tau(p) \equiv 0 \bmod \ell$ for $\ell \in \{11, 13, 17, 19\}$. To do that we will use the following lemma.

Lemma 4.3. *Let K be a field of characteristic not equal to 2. Then the following conditions on $M \in \mathrm{GL}_2(K)$ are equivalent:*

(1) $\mathrm{tr}\, M = 0$.

(2) For the action of M on $\mathbb{P}^1(K)$, there are 0 or 2 orbits of length 1 and all other orbits have length 2.

(3) The action of M on $\mathbb{P}^1(K)$ has an orbit of length 2.

Proof. We begin with verifying (1)$\Rightarrow$(2). Suppose $\mathrm{tr}\, M = 0$. Matrices of trace 0 in $\mathrm{GL}_2(K)$ have distinct eigenvalues in $\overline{K}$ because of $\mathrm{char}(K) \neq 2$. It follows that two such matrices are conjugate if and only if their characteristic polynomials coincide. Hence M and $M' := \left(\begin{smallmatrix} 0 & 1 \\ -\det M & 0 \end{smallmatrix}\right)$ are conjugate so without loss of generality we assume $M = M'$. Since M^2 is a scalar matrix, all the orbits of M on $\mathbb{P}^1(K)$ have length 1 or 2. If there are at least 3 orbits of length 1 then K^2 itself is an eigenspace of M hence M is scalar, which is not the case. If there is exactly one orbit of length 1 then M has a non-scalar Jordan block in its Jordan decomposition, which contradicts the fact that the eigenvalues are distinct.

The implication (2)$\Rightarrow$(3) is trivial so that leaves proving (3)$\Rightarrow$(1). Suppose that M has an orbit of length 2 in $\mathbb{P}^1(K)$. After a suitable conjugation, we may assume that this orbit is $\{[\binom{1}{0}], [\binom{0}{1}]\}$. But this means that $M \sim \left(\begin{smallmatrix} 0 & a \\ b & 0 \end{smallmatrix}\right)$ for certain $a, b \in K$ hence $\mathrm{tr}\, M = 0$. □

Combining this lemma with Theorem 4.1 one sees that for $\ell \in \{11, 13, 17, 19\}$ and $p \neq \ell$ we have $\tau(p) \equiv 0 \bmod \ell$ if and only if the prime p decomposes in the number field $\mathbb{Q}[x]/(P_{12,\ell})$ as a product of primes of degree 1 and 2, with degree 2 occurring at least once. For $p \nmid \mathrm{Disc}(P_{12,\ell})$, which is a property that all primes p satisfying Serre's criteria possess, we can verify this condition by checking whether $P_{12,\ell}$ has an irreducible factor of degree 2 over $\mathbb{F}_p$. This can be easily checked by verifying

$$\bar{x}^{p^2} = \bar{x} \quad \text{and} \quad \bar{x}^p \neq \bar{x} \quad \text{in} \quad \mathbb{F}_p[x]/(\overline{P}_{12,\ell}).$$

Having done a computer search, it turns out that the first few primes satisfying Serre's criteria as well as $\tau(p) \equiv 0 \bmod 11 \cdot 13 \cdot 17 \cdot 19$ are

$$22798241520242687999,\ 60707199950936063999,\ 93433753964906495999.$$

Remark. The unpublished paper [34] in which Bruce Jordan and Blair Kelly obtained the previous bound for the verification of Lehmer's conjecture seems to be unfindable. Kevin Buzzard asked me the question what method they could have used. If we weaken the above search to using only the prime $\ell = 11$ we obtain the same bound as Jordan and Kelly did. So our speculation is that they searched for primes p satisfying Serre's criteria as well as $\tau(p) \equiv 0 \bmod 11$. This congruence can be verified using an elliptic curve computation, as was already remarked in Subsection 4.1.2.

4.5 The table of polynomials

In this section we present the table of polynomials that is referred to throughout this chapter.

Table 4.1: Polynomials belonging to projective modular representations

(k,ℓ)	$P_{k,\ell}$
$(12,11)$	$x^{12}-4x^{11}+55x^9-165x^8+264x^7-341x^6+330x^5$ $-165x^4-55x^3+99x^2-41x-111$
$(12,13)$	$x^{14}+7x^{13}+26x^{12}+78x^{11}+169x^{10}+52x^9-702x^8-1248x^7$ $+494x^6+2561x^5+312x^4-2223x^3+169x^2+506x-215$
$(12,17)$	$x^{18}-9x^{17}+51x^{16}-170x^{15}+374x^{14}-578x^{13}+493x^{12}$ $-901x^{11}+578x^{10}-51x^9+986x^8+1105x^7+476x^6+510x^5$ $+119x^4+68x^3+306x^2+273x+76$
$(12,19)$,	$x^{20}-7x^{19}+76x^{17}-38x^{16}-380x^{15}+114x^{14}+1121x^{13}-798x^{12}$ $-1425x^{11}+6517x^{10}+152x^9-19266x^8-11096x^7+16340x^6$ $+37240x^5+30020x^4-17841x^3-47443x^2-31323x-8055$
$(16,17)$	$x^{18}-2x^{17}-17x^{15}+204x^{14}-1904x^{13}+3655x^{12}+5950x^{11}$ $-3672x^{10}-38794x^9+19465x^8+95982x^7-280041x^6-206074x^5$ $+455804x^4+946288x^3-1315239x^2+606768x-378241$
$(16,19)$	$x^{20}+x^{19}+57x^{18}+38x^{17}+950x^{16}+4389x^{15}+20444x^{14}$ $+84018x^{13}+130359x^{12}-4902x^{11}-93252x^{10}+75848x^9$ $-1041219x^8-1219781x^7+3225611x^6+1074203x^5$ $-3129300x^4-2826364x^3+2406692x^2+6555150x-5271039$

Continued on next page

Table 4.1 – continued from previous page

(k,ℓ)	$P_{k,\ell}$
$(16,23)$	$x^{24}+9x^{23}+46x^{22}+115x^{21}-138x^{20}-1886x^{19}+1058x^{18}$ $+59639x^{17}+255599x^{16}+308798x^{15}-1208328x^{14}$ $-6156732x^{13}-10740931x^{12}+2669403x^{11}+52203054x^{10}+106722024x^{9}$ $+60172945x^{8}-158103380x^{7}-397878081x^{6}-357303183x^{5}$ $+41851168x^{4}+438371490x^{3}+484510019x^{2}+252536071x+55431347$
$(18,17)$	$x^{18}-7x^{17}+17x^{16}+17x^{15}-935x^{14}+799x^{13}+9231x^{12}-41463x^{11}$ $+192780x^{10}+291686x^{9}-390014x^{8}+6132223x^{7}-3955645x^{6}+2916112x^{5}$ $+45030739x^{4}-94452714x^{3}+184016925x^{2}-141466230x+113422599$
$(18,19)$	$x^{20}+10x^{19}+57x^{18}+228x^{17}-361x^{16}-3420x^{15}+23446x^{14}+88749x^{13}$ $-333526x^{12}-1138233x^{11}+1629212x^{10}+13416014x^{9}+7667184x^{8}$ $-208954438x^{7}+95548948x^{6}+593881632x^{5}-1508120801x^{4}$ $-1823516526x^{3}+2205335301x^{2}+1251488657x-8632629109$
$(18,23)$	$x^{24}+23x^{22}-69x^{21}-345x^{20}-483x^{19}-6739x^{18}+18262x^{17}$ $+96715x^{16}-349853x^{15}+2196684x^{14}-7507476x^{13}+59547x^{12}$ $+57434887x^{11}-194471417x^{10}+545807411x^{9}+596464566x^{8}$ $-9923877597x^{7}+33911401963x^{6}-92316759105x^{5}+157585411007x^{4}$ $-171471034142x^{3}+237109280887x^{2}-93742087853x+97228856961$
$(20,19)$	$x^{20}-5x^{19}+76x^{18}-247x^{17}+1197x^{16}-8474x^{15}+15561x^{14}-112347x^{13}$ $+325793x^{12}-787322x^{11}+3851661x^{10}-5756183x^{9}+20865344x^{8}$ $-48001353x^{7}+45895165x^{6}-245996344x^{5}+8889264x^{4}$ $-588303992x^{3}-54940704x^{2}-538817408x+31141888$
$(20,23)$	$x^{24}-x^{23}-23x^{22}-184x^{21}-667x^{20}-5543x^{19}-22448x^{18}$ $+96508x^{17}+1855180x^{16}+13281488x^{15}+66851616x^{14}$ $+282546237x^{13}+1087723107x^{12}+3479009049x^{11}+8319918708x^{10}$ $+8576048755x^{9}-19169464149x^{8}-111605931055x^{7}-227855922888x^{6}$ $-193255204370x^{5}+176888550627x^{4}+1139040818642x^{3}$ $+1055509532423x^{2}+1500432519809x+314072259618$
$(22,23)$	$x^{24}-2x^{23}+115x^{22}+23x^{21}+1909x^{20}+22218x^{19}+9223x^{18}+121141x^{17}$ $+1837654x^{16}-800032x^{15}+9856374x^{14}+52362168x^{13}-32040725x^{12}$ $+279370098x^{11}+1464085056x^{10}+1129229689x^{9}+3299556862x^{8}$ $+14586202192x^{7}+29414918270x^{6}+45332850431x^{5}-6437110763x^{4}$ $-111429920358x^{3}-12449542097x^{2}+93960798341x-31890957224$

Bibliography

[1] D. Abramovich, *A linear lower bound on the gonality of modular curves*, Internat. Math. Res. Notices **30** (1996) 1005–1011.

[2] T. Asai, *On the Fourier coefficients of at various cusps and some applications to Rankin's involution*, J. Math. Soc. Japan **28** (1976) no.1, 48–61.

[3] A. O. L. Atkin and W. W. Li, *Twists of newforms and pseudo-eigenvalues of W-operators*, Invent. Math. **48** (1978) 221–243.

[4] E. Bach and D. Charles, *The hardness of computing an eigenform*, arXiv reference 0708.1192.

[5] M. H. Baker, Torsion points on modular curves, Ph.D. thesis, University of California, Berkeley (1999).

[6] W. Bosma, J. J. Cannon, C. E. Playoust, *The magma algebra system I: the user language*, J. Symbolic Comput. **24** (1997) no. 3/4, 235–265.

[7] J. G. Bosman, *A polynomial with Galois group* $\mathrm{SL}_2(\mathbb{F}_{16})$, LMS J. Comput. Math. **10** (2007) 378–388.

[8] J. G. Bosman, *Modulaire vormen en berekeningen in Galoistheorie*, Nieuw Arch. Wiskd. (5) **2** (2008) no. 3, 184–187.

[9] N. Boston, H. W. Lenstra, K. A. Ribet, *Quotients of group rings arising from two-dimensional representations*, C. R. Acad. Sci. Paris, Série I **312** (1991) 323–328.

[10] N. Bourbaki, Algèbre, chapitre 8: modules et anneaux semi-simples, Actualités scientifiques et industrielles **1261**, Hermann, Paris, 1958.

[11] J. A. Buchmann and H. W. Lenstra, Jr., *Approximating rings of integers in number fields*, J. Théor. Nombres Bordeaux **6** (1994) no. 2, 221–260.

[12] K. Buzzard, *On level-lowering for mod* 2 *representations*, Math. Research Letters **7** (2000) 95–110.

[13] K. Buzzard, *A mod* ℓ *multiplicity one result*, appendix to [62].

[14] D. G. Cantor and D. M. Gordon, *Factoring polynomials over p-adic fields*, Proceedings of the 4th International Symposium on Algorithmic Number Theory, 2000, 185–208.

[15] D. Casperson and J. McKay, *Symmetric functions, m-sets, and Galois groups*, Math. Comp. **63** (1994) 749–757.

[16] H. Cohen and F. Diaz y Diaz, *A polynomial reduction algorithm*, Sém Th. Nombres Bordeaux (Série 2) **3** (1991) 351–360.

[17] J.-M. Couveignes, *Linearizing torsion classes in the Picard group of algebraic curves over finite fields*, to appear in J. Algebra.

[18] J. E. Cremona, Algorithms for modular elliptic curves, Cambridge University Press, 1992.

[19] C. Curtis and I. Reiner, Representation theory of finite groups and associative algebras, Pure and applied mathematics **11**, Interscience, New York, 1962.

[20] H. Darmon, *Serre's conjectures*, Seminar on Fermat's Last Theorem (Toronto, ON, 1993-1994) CMS Conf. Proc., **17**, Amer. Math. Soc., Providence, RI, 1995, 135–153.

[21] P. Deligne, *Formes modulaires et représentations ℓ-adiques*, Lecture Notes in Mathematics **179** (1971) 139–172.

[22] P. Deligne and M. Rapoport, *Les schémas de modules de courbes elliptiques*, Lecture Notes in Mathematics **349** (1973) 143–316.

[23] P. Deligne and J-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. Ec. Norm. Sup. **7** (1974) 507–530.

[24] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem (Toronto, ON, 1993-1994) CMS Conf. Proc., **17**, Amer. Math. Soc., Providence, RI, 1995, 39–133.

[25] F. Diamond and J. Shurman, A first course in modular forms, Graduate Texts in Mathematics **228**, Springer-Verlag, New York, 2005.

[26] V. G. Drinfel'd, *Two theorems on modular curves*, Functional Analysis and Its Applications **7** (1973) no. 2, 155–156.

[27] S. J. Edixhoven, *The weight in Serre's conjectures on modular forms*, Invent. Math. **109** (1992) no. 3, 563–594.

[28] S. J. Edixhoven, J.-M. Couveignes, R. S. de Jong, F. Merkl, J. G. Bosman, *On the computation of coefficients of a modular form*, eprint, 2006, arXiv reference math.NT/0605244v1.

[29] K. Geissler and J. Klüners, *Galois group computation for rational polynomials*, J. Symbolic Comput. **30** (2000) 653–674.

[30] F. Hajir and C. Maire, *Tamely ramified towers and discriminant bounds for number fields II*, J. Symbolic Comput. **33** (2002) 415–423.

[31] M. van Hoeij, *Factoring polynomials and the knapsack problem*, J. Number Theory **95** (2002) 167–189.

[32] N. Jacobson, Basic algebra I, Freeman and Company, San Francisco, 1974.

[33] J. W. Jones and D. P. Roberts, *Galois number fields with small root discriminant*, J. Number Theory **122** (2007) 379–407.

[34] B. Jordan and B. Kelly, *The vanishing of the Ramanujan Tau function*, preprint, 1999.

[35] N. M. Katz, *p-adic properties of modular schemes and modular forms*, Lecture Notes in Mathematics **350** (1973) 69–170.

[36] N. M. Katz and B. Mazur Arithmetic moduli of elliptic curves, Ann. Math. Studies **108**, Princeton Univ. Press, Princeton, 1985.

[37] C. Khare, *Serre's modularity conjecture: a survey of the level one case*, to appear in *L*-functions and Galois representations (Durham, U.K., 2004).

[38] C. Khare and J.-P. Wintenberger, *Serre's modularity conjecture: the level one case*, to appear in Duke Math. J.

[39] C. Khare and J.-P. Wintenberger, *Serre's modularity conjecture: the odd conductor case (I, II)*, preprint, 2006, available at `http://www.math.utah.edu/ shekhar/papers.html`

[40] K. Khuri-Makdisi, *Asymptotically fast group operations on Jacobians of general curves*, Math. Comp. **76** (2007) 2213–2239.

[41] L. J. P. Kilford and G. Wiese, *On the failure of the Gorenstein property for Hecke algebras of prime weight*, to appear in Exp. Math.

[42] M. Kisin, *Modularity of* 2*-dimensional Galois representations*, Current Developments in Mathematics 2005, 191–230.

[43] M. Kisin, *Modularity of* 2*-adic Barsotti-Tate representations*, preprint, 2007.

[44] J. Klüners and G. Malle, *Explicit Galois realization of transitive groups of degree up to 15*, J. Symbolic Comput. **30** (2000) no. 6, 675–716.

[45] S. Landau, *Factoring polynomials over algebraic number fields*, SIAM J. Comput **14** (1985) 184–195.

[46] S. Lang, Elliptic curves: Diophantine analysis, Grundlehren der mathematischen Wissenschaften **231**, Springer-Verlag, New York, 1978.

[47] D. H. Lehmer, *The vanishing of Ramanujan's function* $\tau(n)$, Duke Math. J. **10** (1947) 429–433.

[48] H. W. Lenstra, Jr., *Algorithms in algebraic number theory*, Bull. Amer. Math. Soc. (N.S.) **26** (1992) no. 2, 211–244.

[49] A. K. Lenstra, H. W. Lenstra, Jr., L. Lovász, *Factoring polynomials with rational coefficients*, *Math. Ann.* **261** (1982) 515–534.

[50] W. W. Li, *Newforms and functional equations*, Math. Ann. **212** (1975) 285–315.

[51] Y. Manin, *Parabolic points and zeta functions of modular curves*, Math. USSR Izvestija **36** (1972) 19–66.

[52] B. Mazur, *Modular curves and the Eisenstein ideal*, Publ. Math. I.H.E.S. **47** (1977) 33–186.

[53] L. Merel, *Universal fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations (Berlin), Springer, 1994, 59–94.

[54] J. Milne, *Modular functions and modular forms*, course notes, `http://www.jmilne.org/math/CourseNotes/math678.html`

[55] H. Moon and Y. Taguchi, *Refinement of Tate's discriminant bound and non-existence theorems for mod p Galois representations*, Documenta Math. Extra Volume Kato (2003) 641–654.

[56] J. Neukirch, Algebraic number theory, Grundlehren der mathematischen Wissenschaften **322**, Springer-Verlag, Berlin and London, 1999.

[57] A. M. Odlyzko, *Bounds for discriminants and related estimates for class numbers, regulators and zeros of zeta functions: a survey or recent results*, Sém. Théor. Nombres, Bordeaux **2** (1990) 119–141.

[58] A. Ogg, *On the eigenvalues of Hecke operators*, Math. Ann. **179** (1969) 101–108.

[59] G. Poitou, *Minoration de discriminants (d'après A. M. Odlyzko)*, Lecture Notes in Mathematics **567** (1977) 136–153.

[60] B. Poonen, *Gonality of modular curves in characteristic p*, Math. Res. Lett. **14** (2007) no. 4, 691–701.

[61] K. A. Ribet, *Report on mod ℓ representations of* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Motives (Seattle, WA, 1991), Amer. Math. Soc., Providence, RI, 1994, 639–676.

[62] K. A. Ribet and W. A. Stein, *Lectures on Serre's conjectures*, Arithmetic algebraic geometry (Park City, UT, 1999), Amer. Math. Soc., Providence, RI, 2001, 143–232.

[63] R. Schoof, *Elliptic curves over finite fields and the computation of square roots mod p*, Math. Comp. **44** (1985) 483–494.

[64] J-P. Serre, *Une interprétation des congruences relatives à la fonction τ de Ramanujan*, Séminaire Delange-Pisot-Poitou, 1968, no. 14.

[65] J-P. Serre, Linear representations of finite groups, Graduate Texts in Mathematics **42**, Springer-Verlag, New York, 1977.

[66] J-P. Serre, *Modular forms of weight one and Galois representations*, Algebraic number fields: *L*-functions and Galois properties (A. Frölich, ed.), Academic Press, London, 1977, 193–268.

[67] J-P. Serre, Local fields, Graduate Texts in Mathematics **67**, Springer-Verlag, New York, 1979.

[68] J-P. Serre, *Sur la lacunarité des puissances de* η, Glasgow Math. J. **27** (1985) 203–221.

[69] J-P. Serre, *Minoration de discriminants*, note of October 1975, published in Œuvres, Vol. III (Springer, 1986) 240–243.

[70] J-P. Serre, *Sur les représentations modulaires de degré* 2 *de* $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, Duke Math. J. **54** (1987) no. 1, 179–230.

[71] G. Shimura, *On the periods of modular forms*, Math. Ann. **229** (1977) no. 3, 211–221.

[72] V. V. Shokurov, *Shimura integrals of cusp forms*, Math. USSR Izvestija **16** (1981) no. 3, 603–646.

[73] V. Shoup, A computational introduction to number theory and algebra, 2nd edition, to appear in Cambridge University Press.

[74] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Graduate Texts in Mathematics **151**, Springer-Verlag, New York, 1994.

[75] C. C. Sims, *Computational methods for permutation groups*, Computational problems in abstract algebra (J. Leech, ed.), (Pergamon, Elmsforth, N.Y., 1970) 169–184.

[76] L. Soicher and J. McKay, *Computing Galois groups over the rationals*, J. Number Theory **20** (1985) 273–281.

[77] R. P. Stauduhar, *The determination of Galois groups*, Math. Comp. **27** (1973) 981–996.

[78] W. A. Stein, *Explicit approaches to modular abelian varieties*, Ph.D. thesis, University of California, Berkeley (2000).

[79] W. A. Stein, *Modular forms, a computational approach*, Graduate Studies in Mathematics **79**, Amer. Math. Soc., Providence, RI, 2007.

[80] W. A. Stein, *An introduction to computing modular forms using modular symbols*, Algorithmic Number Theory, MSRI Publications **44** (2008), 641–652.

[81] J. Sturm, *On the congruence of modular forms*, Lecture Notes in Mathematics **1240** (1987) 275–280.

[82] M. Suzuki, Group Theory I, Grundlehren der mathematischen Wissenschaften **247**, Springer-Verlag, New York, 1982.

[83] H. P. F. Swinnerton-Dyer, *On ℓ-adic representations and congruences for modular forms*, Lecture Notes in Mathematics **350** (1973) 1–55.

[84] H. P. F. Swinnerton-Dyer, *On ℓ-adic representations and congruences for modular forms (II)*, Lecture Notes in Mathematics **601** (1977) 64–90.

[85] R. Taylor, *Galois representations*, Ann. Fac. Sci. Toulouse **13** (2004) 73–119.

[86] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. Math. **141** (1995) no. 3, 553–572.

[87] G. Wiese, *Multiplicities of Galois representations of weight one*, Algebra and Number Theory **1** (2007) no. 1, 67–85.

Samenvatting

Expliciete berekeningen met modulaire Galoisrepresentaties

De tekst van deze samenvatting is gebaseerd op het door de auteur geschreven populairwetenschappelijke artikel [8].

Galoistheorie

Op de middelbare school leert iedereen een kwadratische vergelijking $ax^2+bx+c=0$ oplossen met behulp van de *abc*-formule:

$$x=\frac{-b\pm\sqrt{b^2-4ac}}{2a}.$$

Voor vergelijkingen van graad 3 bestaat er een soortgelijke formule, in 1545 gepubliceerd door Cardano, na het gestolen te hebben van Tartaglia: de nulpunten van het polynoom ax^3+bx^2+cx+d zijn gelijk aan

$$x=\sqrt[3]{C+\sqrt{D}}+\sqrt[3]{C-\sqrt{D}}-\frac{b}{3a},$$

waarbij

$$C=\frac{-b^3}{27a^3}+\frac{bc}{6a^2}-\frac{d}{2a},\quad D=C^2+\left(\frac{c}{3a}-\frac{b^2}{9a^2}\right)^3$$

en de derdemachtswortels geschikt gekozen dienen te worden. We zien dat de nulpunten van tweedegraads- en derdegraadspolynomen gegeven kunnen worden als uitdrukkingen in de coëfficiënten, waarbij we de operaties $+$, $-$, $\cdot$, $/$ en $\sqrt[n]{}$ gebruiken. We zullen in zo'n geval zeggen dat het polynoom *oplosbaar* is. We kunnen ons afvragen of dit ook geldt voor polynomen van willekeurige graad. Ferrari, een student van Cardano, had in 1540 al aangetoond dat vierdegraadsvergelijkingen oplosbaar zijn, onder de voorwaarde dat derdegraadsvergelijkingen oplosbaar zijn.

Naar een formule voor de nulpunten van polynomen van graad 5 en hoger heeft men sindsdien eeuwenlang tevergeefs gezocht. In 1799 vond de Italiaanse wiskundige Ruffini zelfs een bewijs dat zo'n formule in het algemeen niet bestaat! Niemand geloofde hem echter,

totdat Abel in 1826 eveneens een bewijs vond. Zelfs vandaag de dag zijn er nog ongelovige thomassen die, uiteraard zonder succes, formules voor oplossingen van vijfdegraadsvergelijkingen proberen te vinden. Laten we hierbij wel opmerken dat het niet zo is dat geen enkele vergelijking van graad 5 of hoger opgelost kan worden. De nulpunten van $x^5 - x - 1$ kun je weliswaar niet uitdrukken in elementaire formules, maar die van $x^5 - 2$ wel: dat zijn alle waarden van $\sqrt[5]{2}$.

In 1832 vond Galois een nieuw bewijs voor het feit dat vergelijkingen vanaf graad 5 niet op te lossen zijn met $+$, $-$, $\cdot$, $/$ en $\sqrt[n]{}$. Het bewijs van Galois is zeer interessant omdat het veel meer inzicht en structuur aan een polynoom geeft dan alleen 'ja, het kan' of 'nee, het kan niet'.

Laten we eens kijken hoe Galois het deed. Kies je favoriete polynoom

$$P(x) = a_n x^n + \cdots + a_0 \in \mathbb{Q}[x] \quad \text{met nulpunten } \alpha_1, \ldots, \alpha_n \in \mathbb{C}.$$

We zullen veronderstellen dat de nulpunten *verschillend* zijn; dit is geen grote belemmering want we kunnen meervoudige factoren makkelijk vinden. Er zijn allerlei relaties tussen de nulpunten. Zo kunnen we het product uitwerken in de identiteit

$$a_n(x - \alpha_1) \cdots (x - \alpha_n) = a_n x^n + \cdots + a_0$$

en dan vinden we bij elke coëfficiënt een symmetrische relatie, bijvoorbeeld

$$\alpha_1 + \cdots + \alpha_n = \frac{-a_{n-1}}{a_n} \quad \text{en} \quad \alpha_1 \cdots \alpha_n = \frac{(-1)^n a_0}{a_n}.$$

Afhankelijk van het polynoom kunnen er meerdere relaties tussen de nulpunten zijn dan degenen die je direct uit de symmetrische relaties kunt afleiden. Galois kwam op het idee om de groep van alle permutaties van de nulpunten te bekijken die alle relaties tussen deze nulpunten vasthouden; deze groep heet vandaag de dag de *Galoisgroep* van het polynoom P en noteren we met $\mathrm{Gal}(P)$. Als er niet meer relaties tussen de nulpunten zijn dan degenen die je uit de symmetrische relaties kunt afleiden, dan zal $\mathrm{Gal}(P)$ uit alle mogelijk permutaties tussen de nulpunten bestaan en dus isomorf zijn met S_n, de volledige symmetrische groep van graad n. Als er echter meer relaties zijn, dan leggen deze restricties op de permutaties op en zal $\mathrm{Gal}(P)$ dus kleiner zijn.

De oplosbaarheid van een polynoom P kan nu worden uitgedrukt in abstracte eigenschappen van de Galoisgroep $G = \mathrm{Gal}(P)$. We gaan een rij

$$G = G_1 \supset G_2 \supset \cdots$$

van ondergroepen van G maken aan de hand van het volgende recept: Begin met $G_1 = G$ en neem daarna telkens de *commutatorondergroep*:

$$G_{i+1} = [G_i, G_i] := \langle ghg^{-1}h^{-1} : g, h \in G \rangle.$$

Men kan laten zien dat P oplosbaar is dan en slechts dan als ergens in deze rij de triviale groep voorkomt. We zeggen in zo'n geval ook wel dat de groep G *oplosbaar* is.

Een groep die erg cruciaal is in deze context is $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, de automorfismengroep van het lichaam van algebraïsche getallen. Het is een topologische groep waarin de Galoisgroepen van alle polynomen in $\mathbb{Q}[x]$ gecodeerd zitten. Voor eindige groepen G is het geven van een polynoom met Galoisgroep G (grofweg) equivalent met het geven van een continu surjectief homomorfisme $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to G$. De groep $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ is binnen deze theorie dus een soort allesomvattend object en daarmee ook meteen heel moeilijk te begrijpen.

Modulaire vormen en Galoisrepresentaties

Modulaire vormen spelen een belangrijke rol in de getaltheorie. Grofweg zijn het holomorfe functies op het complexe bovenhalfvlak $\mathfrak{H}$ die aan bepaalde groeivoorwaarden en aan bepaalde symmetrierelaties ten aanzien van transformaties van de vorm $z \mapsto \frac{az+b}{cz+d}$ voldoen.

Een belangrijk voorbeeld van een modulaire vorm die veel wiskundigen heeft beziggehouden is de functie

$$\Delta(z) = q\prod_{n\geq 1}(1-q^n)^{24}, \quad \text{waarbij } q = e^{2\pi iz}. \tag{4.3}$$

De groeivoorwaarde voor deze functie is $\lim_{\Im z\to\infty}\Delta(z) = 0$ en de symmetrierelatie luidt in dit geval dat $\Delta(z)$ voldoet aan

$$\Delta\left(\frac{az+b}{cz+d}\right) = (cz+d)^{12}\Delta(z)$$

voor alle $z \in \mathfrak{H}$ en $a,b,c,d \in \mathbb{Z}$ met $ad-bc = 1$. We kunnen deze transformaties visualiseren in een plaatje dat laat zien hoe het complexe bovenhalfvlak in driehoeken wordt opgedeeld; zie de figuur op bladzijde 2. Als we het product in (4.3) uitwerken dan krijgen we een machtreeks

$$\Delta(z) = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 + \cdots = \sum_{n\geq 1}\tau(n)q^n,$$

met $\tau(n)$ geheel. De functie $\tau : \mathbb{Z}_{>0} \to \mathbb{Z}$ die op deze manier gedefinieerd is heet de *Ramanujan tau-functie*. Ramanujan merkte een aantal merkwaardige eigenschappen van zijn tau-functie op. Onder andere waren daar de volgende drie eigenschappen, die hij niet kon bewijzen:

- Als m en n ondeelbaar zijn dan geldt $\tau(mn) = \tau(m)\tau(n)$.
- Voor priemmachten geldt de recursie $\tau(p^{r+1}) = \tau(p)\tau(p^r) - p^{11}\tau(p^{r-1})$.
- Voor priemgetallen hebben we een ongelijkheid $|\tau(p)| \leq 2p^{11/2}$.

De eerste twee eigenschappen zijn in 1917 door Mordell bewezen, maar de derde is lange tijd onopgelost geweest.

Behalve de bovengenoemde eigenschappen vond Ramanujan ook nog congruenties voor $\tau(n)$ modulo (machten van) de priemgetallen 2, 3, 5, 7, 23 en 691, bijvoorbeeld

$$\tau(n) \equiv 1 + n^{11} \quad \text{voor alle } n.$$

Serre begon zich af te vragen waarom zulke congruenties niet bestaan modulo andere priemgetallen. In 1968 formuleerde hij een vermoeden waarin werd gesteld dat $\tau(p)$ uit te drukken is in termen van 2-dimensionale Galoisrepresentaties, dat wil zeggen continue homomorfismen $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(K)$ waarbij K een zeker lichaam is. Hij bracht op die manier het modulo ℓ gedrag van $\tau(p)$ in verband met de grens $|\tau(p)| \leq 2p^{11/2}$. Het lukte Deligne in 1969 om het bestaan van zulke representaties aan te tonen en in 1974 slaagde hij erin om hiermee $|\tau(p)| \leq 2p^{11/2}$ te bewijzen. Het bewijs van Deligne gebruikt diepe resultaten uit de algebraïsche meetkunde; het totale aantal pagina's dat je krijgt als je alles helemaal vanaf het begin zou uitschrijven wordt geschat op ongeveer 2000.

De vorm Δ is niet uniek hierin. Eigenschappen die vergelijkbaar zijn met die voor de vorm Δ gelden voor veel meer modulaire vormen. De modulaire vormen in kwestie heten *eigenvormen* omdat het eigenvectoren zijn voor bepaalde lineaire operatoren op ruimten van modulaire vormen, de zogenaamde Heckeoperatoren. Bij elke eigenvorm blijken er Galoisrepresentaties gemaakt te kunnen worden.

De afgelopen decennia is het verband tussen eigenvormen en Galoisrepresentaties zeer intensief bestudeerd. Een van de grote resultaten die hieruit voortkwam is Wiles' bewijs voor de Laatste Stelling van Fermat. Een ander groot resultaat, dat sterk in verband staat met het werk van Wiles, is het bewijs voor het Serrevermoeden, gegeven door Khare, Wintenberger en Kisin. Dit Serrevermoeden stelt dat een representatie $\rho : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(K)$ met K een eindig lichaam slechts aan een paar hele milde voorwaarden hoeft te voldoen om al van een eigenvorm afkomstig te zijn.

Het berekenen van $\tau(n)$

Een vraag die René Schoof aan Bas Edixhoven stelde is of het mogelijk is om $\tau(n)$ efficiënt uit te rekenen. Als we $\tau(p)$ kunnen uitrekenen voor priemgetallen p en n kunnen factoriseren in priemgetallen dan kunnen we, wegens de observaties van Ramanujan, $\tau(n)$ uitrekenen. Het zou mooi zijn om $\tau(n)$ snel te kunnen uitrekenen zonder te hoeven factoriseren; in dat geval zou het bekende en veelgebruikte cryptosysteem RSA namelijk gekraakt zijn. Voorlopig is het echter niet duidelijk hoe dit aangepakt zou kunnen worden en kunt u nog veilig internetbankieren.

Als we nu $\tau(p) \bmod \ell$ uitrekenen voor zo veel priemgetallen ℓ dat hun product groter dan $4p^{11/2}$ is, dan ligt, gezien de grens voor $|\tau(p)|$ hiermee $\tau(p)$ zelf vast. Met dit in het achterhoofd is Edixhoven een project gestart waarin hij het probleem tracht aan te pakken door de bijbehorende Galoisrepresentaties uit te rekenen. Dit proefschrift vormt een onderdeel van het project. Het basisidee van de berekeningen komt uit de meetkunde: de Galoisrepresentatie die bij $\tau(p) \bmod \ell$ hoort voor een gegeven ℓ kan worden gerealiseerd in een variëteit

die $J_1(\ell)$ genoemd wordt en dimensie $(\ell-5)(\ell-7)/24$ heeft. Jean-Marc Couveignes had het idee om hierbij numerieke berekeningen te gebruiken. Om deze ideeën hard te maken lijkt het echter onvermijdelijk om *Arakelovmeetkunde* te gebruiken; op dit punt kon Robin de Jong zijn steentje bijdragen aan het project. Hierbij is gebruikgemaakt van een resultaat van Franz Merkl, iemand uit de kansrekening.

Een nadeel van het algoritme van Edixhoven, Couveignes en De Jong is dat het praktisch niet goed werkt. Zo is de rekenprecisie te hoog en moeten we in plaats van $J_1(\ell)$ de variëteit $J_1(5\ell)$ gebruiken, waarvan de dimensie $(\ell-2)^2$ is.

In de praktijk kunnen we deze bezwaren negeren en gewoon gaan rekenen. We krijgen polynomen met coëfficiënten van een hoge precisie (denk hier aan enkele duizenden decimalen). We weten dat de coëfficiënten benaderingen zijn van rationale getallen. Als de benadering sterk genoeg is, dan gokken we dat de rationale getallen waar ze dichtbij liggen de daadwerkelijke coëfficiënten zijn van de polynomen die bij de representaties horen. We moeten dan wel nog achteraf nagaan dat het verkregen polynoom correct is. Dankzij het feit dat het Serrevermoeden nu bewezen is, is dit allemaal goed te doen. Uiteraard geldt ook hier dat we niet tot de tau-functie beperkt zijn. De rekenmethoden werken met eigenvormen in het algemeen.

Dit proefschrift

In Hoofdstuk 1 van dit proefschrift zullen wij de theorie van modulaire vormen behandelen. Voorts zullen wij in Hoofdstuk 2 bespreken hoe er gerekend kan worden aan modulaire vormen en Galoisrepresentaties. In de Hoofdstukken 3 en 4 zullen we enkele resultaten van de berekeningen presenteren die zijn uitgevoerd.

In Hoofdstuk 3 betreft deze berekening de oplossing van een probleem uit de *computationele inverse Galoistheorie* dat Jürgen Klüners, een van de grote wereldexperts op dit gebied, mij had voorgelegd. In de computationele inverse Galoistheorie tracht men voor zo veel mogelijk groepen G een polynoom te vinden waarvan G de Galoisgroep is. De groep in Hoofdstuk 3 betreft $\mathrm{SL}_2(\mathbb{F}_{16})$, de groep van 2 bij 2 matrices met determinant 1 en coëfficiënten in het lichaam van 16 elementen. Verschillende mensen waren naar zo'n polynoom op zoek. Of er ook voor elke groep G een polynoom bestaat met Galoisgroep G is een zeer moeilijk onopgelost probleem in de getaltheorie.

De resultaten van Hoofdstuk 4, betreffen enkele berekeningen aan Galoisrepresentaties voor de tau-functie en daaraan gerelateerde functies. In dat hoofdstuk zullen we projectieve representaties voor deze functies modulo de priemgetallen $\ell \geq 23$ geven. Als toepassing verbeteren we de grens waarvoor het *Lehmervermoeden* geverifiëerd is met meer dan een factor duizend. Dit vermoeden stelt dat de tau-functie nergens de waarde nul aanneemt. Het resultaat van het hoofdstuk is interessant omdat het een toepassing geeft van het Serrevermoeden, een theoretisch resultaat, in een computationele context.

Curriculum vitae

Johan Bosman werd op 5 februari van het jaar 1979 geboren te Wageningen. Hij heeft zijn jeugd doorgebracht in Renkum, een plaats in Gelderland die volgens historische bronnen al voor het jaar 1000 bestond en daarmee ouder is dan Amsterdam.

In 1997 slaagde hij voor zijn VWO-examen aan het Rijn IJssel College te Arnhem. In 2004 studeerde hij aan de Universiteit Utrecht cum laude af als wiskundige. Naast zijn studie heeft hij trainingen begeleid voor deelnemers aan de Internationale Wiskunde Olympiade en heeft hij zich ontwikkeld als een van de beste weggeefschakers ter wereld. Na zijn studie heeft hij de bovengenoemde nevenactiviteiten stopgezet om zich aan het promotieonderzoek te kunnen wijden waarvan resultaten in dit proefschrift zijn verwerkt. Vanaf oktober 2008 is hij werkzaam als software engineer bij ORTEC Logistics.

Hij heeft verschillende prijzen gewonnen voor zijn prestaties op wiskundig gebied, waaronder een zilveren medaille op de Internationale Wiskunde Olympiade in 1997, de Philips Wiskundeprijs voor promovendi in 2007 en hiertussen meerdere malen de eerste prijs in de Universitaire Wiskunde Competitie.

Hij heeft echter nooit traditionele Japanse muziek leren spelen op de shakuhachi.

Index

B_k, 5
$B_{k,\phi}$, 8
$\mathbb{B}_k$, 44
$\langle d \rangle$, 10, 27, 44
$E_k(\Gamma_1(N))$, 9
$f|_k\gamma$, 10
(f,g), 13
GL_2^+, 2
G_k, 5, 6
$G_k^{\psi,\phi}$, 8, 9
$\mathfrak{H}$, 1
ht, 60
$J_1(N)$, 27
K_f, 15
$k(\rho)$, $\tilde{k}(\rho)$, 39, 40
$M_k(\Gamma)$, 3
$\mathbb{M}_k$, 43, 44
$N(\rho)$, 38
$S_k(N,\varepsilon)$, 10
$S_k(\Gamma_1(N))^{\mathrm{new}}$, $S_k(\Gamma_1(N))^{\mathrm{old}}$, 14
$S_k(\Gamma)$, 3
$\mathbb{S}_k$, 45
T_γ, T_p, T_n, 11, 12, 27, 46
$\mathbb{T}, \mathbb{T}'$, 13
W_Q, 16
$X_0(N)$, $X_1(N)$, $X(N)$, 18, 22
X_Γ, 18
$X_\mu(N)$, 22
$Y_0(N)$, $Y_1(N)$, $Y(N)$, 18, 19, 21
Y_Γ, 18
$Y_\mu(N)$, 20
α_d, 14
$\Gamma_0(N)$, $\Gamma_1(N)$, $\Gamma(N)$, 3
$\Delta(z)$, 6
$\rho_{f,\lambda}$, $\overline{\rho}_{f,\lambda}$, 30, 32
ρ_p, 29
ρ^{ss}, 28
$\sigma_{k-1}(n)$, 6
$\sigma_{k-1}^{\psi,\phi}(n)$, 8
$\tau(n)$, 6
χ_ℓ, $\overline{\chi}_\ell$, 29, 30

Atkin-Lehner operator, 16

Brauer-Nesbitt theorem, 28

commensurable subgroups, 11
congruence subgroup, 3
cusp, 1, 22
cusp form, 3, 25
cyclotomic character, 29, 30

degeneracy map, 14
diamond operator, 10, 27, 44
discriminant modular form, 6

Eichler-Shimura relation, 27
eigenform, 14
Eisenstein series, 5, 7

fundamental character, 37
fundamental domain, 2

Galois representation, *see* representation
Gauss sum, 8
generalised Bernoulli number, 8
generalised elliptic curve, 21, 22

Hecke algebra, 13
Hecke bound, *see* Sturm bound
Hecke ideal, 15
Hecke operator, 11, 27, 46
height, 60
hyperbolic measure, 12

level, 3

of representation, 38
of tame character, 37
level structure, 18

Manin symbols, 47
modular curve, 17
modular discriminant, *see* discriminant modular form
modular form, 3, 24
of level one, 4
with character, 10
modular symbols, 44
boundary, 44
cuspidal, 45
modular transformation property, 3, 5, 10, 25
multiplicity one, 15, 33

Néron polygon, 21
newform, 15

period mapping, 57
Petersson inner product, 12
peu/très ramifiée, 38
pseudo-eigenvalue, 17

q-expansion, 4, 26
at cusp, 52

Ramanujan tau function, 6, 34
reduced weight, 40
representation, 28
associated to newform, 30, 32
exceptional, 35
irreducible, 28
multiplicity of, 33
odd, 30
reduction of, 32
semi-simple, 28
unramified, 29
root discriminant, 70

semi-simplification, 29
Serre invariants, 35
Serre's conjecture, 35, 41
slash operator, 10
star involution, 46
Sturm bound, 78

tame character, 37
tame ramification, 36
Tate curve, 23
Tate module, 31
très ramifiée, *see* peu/très ramifiée

upper half plane, 1

weight, 3
of representation, 39
width of cusp, 2
wild ramification, 36
winding element, 51, 52