Explicit computations with modular Galois representations

Bosman, J.G.

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

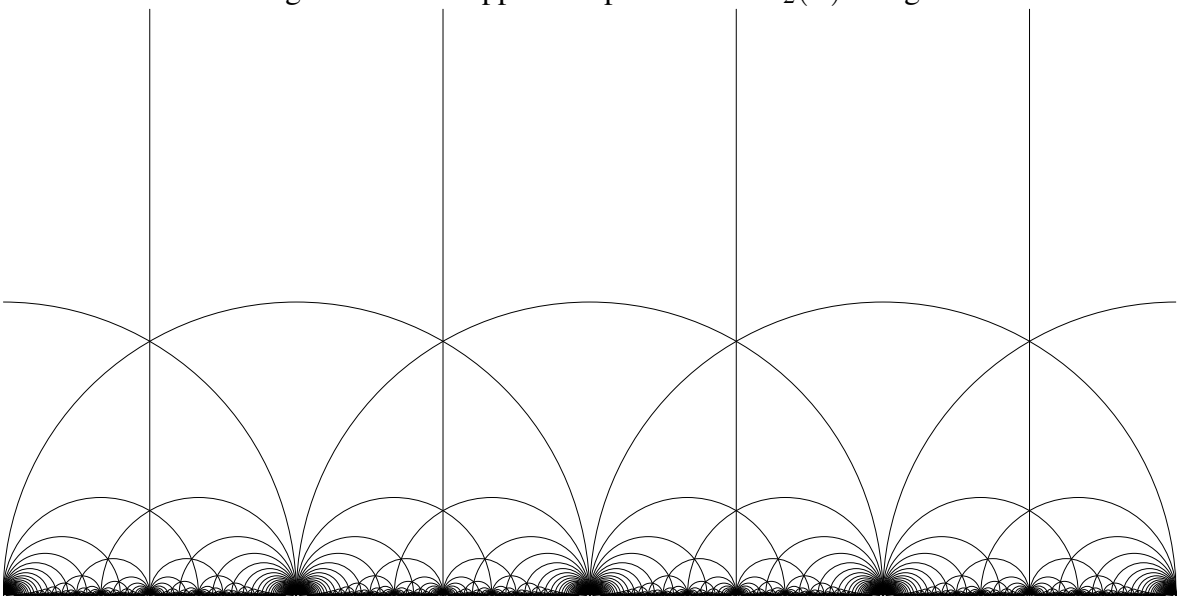
Consider the complex upper half plane $\mathfrak{H} := \{z \in \mathbb{C} : \Im z > 0\}$. On it we have an action of $\mathrm{SL}_2(\mathbb{Z})$ by

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

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 $\Gamma\backslash\mathfrak{H}^*$, 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 $\mathrm{SL}_2(\mathbb{Z})\backslash\mathfrak{H}$ 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.*

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

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:*

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

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].