Modular Forms

Lecturers: Peter Bruin and Sander Dahmen

Spring 2018

Contents

Introduction

Modular forms are a family of mathematical objects that are usually first encountered as holomorphic functions on the upper half-plane satisfying a certain transformation property. However, the study of these functions quickly reveals interesting connections to various other fields of mathematics, such as analysis, elliptic curves, number theory and representation theory.

The importance of modular forms is illustrated by the following quotation, attributed to Martin Eichler (1912–1992): “There are five fundamental operations in mathematics: addition, subtraction, multiplication, division, and modular forms.” Whether Eichler actually said this or not, it is indisputable that thanks to the remarkable properties of modular forms and their connections to other areas of mathematics, they have become an important object of study ever since the nineteenth century.

Further references

To conclude this introduction, we mention some useful references for the material treated in this course.

- A classical reference for modular forms for the full modular group $\mathrm{SL}_2(\mathbb{Z})$ is Serre's book [7, chapters VII and VIII].
- We recommend parts of Diamond and Shurman [4, chapters 1, 3, 4 and 5] for practically all the material covered in this course (and much more).
- Miyake [6, chapter 4] also treats most of the material, from a more analytic point of view than Diamond and Shurman.
- Another very comprehensive reference with an analytic flavour is the recent textbook of Cohen and Strömberg [3].
- For a broad perspective on classical modular forms, Hilbert modular forms, Siegel modular forms and applications of all of these, see the book by Bruinier, van der Geer, Harder and Zagier [1].
- For a more algebraic point of view, see Milne's course notes [5].
- Finally, for those interested in algorithmic aspects of modular forms, there is Stein's book [8].

One can experiment with modular forms using, for instance, the computer algebra packages Magma (`http://magma.maths.usyd.edu.au/`) and SageMath (`http://sagemath.org/`). In this course we will see in particular how to use SageMath for computations with modular forms.

Acknowledgements. These notes are based in part on notes from David Loeffler's course on modular forms taught at the University of Warwick in 2011.

Chapter 1

The modular group

1.1 Motivation: lattice functions

The word 'modular' refers (originally and in this course) to the so-called moduli space of complex elliptic curves. The latter can be described using the following basic concepts.

Definition. A *lattice* (of full rank) in the complex plane $\mathbb{C}$ is a subgroup $\Lambda \subset \mathbb{C}$ of the form

$$\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$$

where $\omega_1, \omega_2 \in \mathbb{C}$ are $\mathbb{R}$-linearly independent.

Two lattices Λ and Λ' are called *homothetic* if there exists a $\lambda \in \mathbb{C}^\times$ such that

$$\Lambda' = \lambda\Lambda := \{\lambda\omega \mid \omega \in \Lambda\}.$$

In this case we write $\Lambda \simeq \Lambda'$.

Let $\mathcal{L}$ denote the set of all lattices in $\mathbb{C}$. It turns out that any $\Lambda \in \mathcal{L}$ yields a complex elliptic curve, and conversely, any complex elliptic curve is isomorphic to $\mathbb{C}/\Lambda$ for some $\Lambda \in \mathcal{L}$. Furthermore, two complex elliptic curves $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ are isomorphic if and only if Λ and Λ' are homothetic. Therefore, in order to study isomorphism classes of complex elliptic curves, it suffices to study complex lattices modulo homothety; we denote the latter set by $\mathcal{L}/{\simeq}$. Furthermore, natural parametrizations of $\mathcal{L}/{\simeq}$ can be considered as natural parametrizations of the isomorphism classes of complex elliptic curves.

From the discussion above, it seems natural to consider functions $G\colon \mathcal{L}/{\simeq} \to \mathbb{C}$. (Actually, enlarging the codomain of G to the Riemann sphere $\mathbb{C}\cup\{\infty\}$ could be desirable, but we will ignore this for the time being.) Any such function corresponds naturally to a function $F\colon \mathcal{L} \to \mathbb{C}$ with the invariance property

$$F(\lambda\Lambda) = F(\Lambda) \quad \text{for all } \lambda \in \mathbb{C}^\times \text{ and } \Lambda \in \mathcal{L}.$$

It turns out to be too restrictive to only consider such function. Instead, we look at functions with a more general transformation property.

Definition. A function

$$\mathcal{F}\colon \mathcal{L} \to \mathbb{C}$$

is called *homogeneous of weight* $k \in \mathbb{Z}$ if it satisfies

$$\mathcal{F}(\lambda\Lambda) = \lambda^{-k}\mathcal{F}(\Lambda) \quad \text{for all } \lambda \in \mathbb{C}^\times \text{ and } \Lambda \in \mathcal{L}. \tag{1.1}$$

As a first example, for $k \in \mathbb{Z}$ with $k > 2$ consider the *Eisenstein seris*

$$\mathcal{G}_k\colon \mathcal{L} \to \mathbb{C}$$

defined by

$$\Lambda \to \sum_{\omega \in \Lambda - \{0\}} \frac{1}{\omega^k}$$

By e.g. comparing the sum to an integral, one can check that the series converges (this is where $k > 2$ is necessary). Furthermore, we immediately obtain the transformation property

$$\mathcal{G}_k(\lambda\Lambda) = \lambda^{-k}\mathcal{G}_k(\Lambda) \quad \text{for all } \lambda \in \mathbb{C}^\times \text{ and } \Lambda \in \mathcal{L}.$$

1.2 The upper half-plane and the modular group

Fundamental roles in the theory of modular forms are played by the *(complex) upper half-plane*

$$\begin{aligned} \mathbb{H} &:= \{z \in \mathbb{C} \mid \Im z > 0\} \\ &= \{x + iy \mid x, y \in \mathbb{R}, y > 0\}. \end{aligned}$$

and the *(full) modular group*

$$\mathrm{SL}_2(\mathbb{Z}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{Z}, ad - bc = 1 \right\}.$$

We will show how these objects, as well as a certain action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$, appear naturally in the study of homogeneous function on lattices described in the previous section. Analogously, one could consider the union of the complex upper and lower half plane $\mathbb{C} - \mathbb{R}$ (sometimes also denoted by $\mathbb{P}^1(\mathbb{C}) - \mathbb{P}^1(\mathbb{R})$) which is acted upon by

$$\mathrm{GL}_2(\mathbb{Z}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{Z}, ad - bc = \pm 1 \right\}$$

as we will describe below.

For $z \in \mathbb{C} - \mathbb{R}$ consider the lattice

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

Note that any lattice in $\mathbb{C}$ can be written as

$$\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \omega_2\Lambda_z \qquad \text{with } z := \omega_1/\omega_2 \in \mathbb{C} - \mathbb{R}.$$

By swapping ω_1 and ω_2 if necessary, we may assume that $\omega_1/\omega_2 \in \mathbb{H}$. We conclude that any homogeneous function $\mathcal{F}\colon \mathcal{L} \to \mathbb{C}$ is completely determined by its values on Λ_z for $z \in \mathbb{H}$. To any $\mathcal{F}$ as above we associate a function

$$f\colon \mathbb{H} \to \mathbb{C} \quad \text{by } z \mapsto \mathcal{F}(\Lambda_z), \tag{1.2}$$

from which the function $\mathcal{F}$ can be recovered as we just noted. In order to study the transformation properties of f, we first introduce an action on $\mathbb{C} - \mathbb{R}$, which restricts to an action on $\mathbb{H}$. This is motivated by the following properties about changing bases for a lattice in $\mathbb{C}$.

Lemma 1.1. *Let $\omega_1, \omega_2, \omega_1', \omega_2' \in \mathbb{C}^\times$ with $\omega_1/\omega_2, \omega_1'/\omega_2' \notin \mathbb{R}$.*

(i) We have $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \mathbb{Z}\omega_1' + \mathbb{Z}\omega_2'$ if and only if

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \gamma \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \quad \textit{for some } \gamma \in \mathrm{GL}_2(\mathbb{Z}). \tag{1.3}$$

(ii) Suppose $\omega_1/\omega_2 \in \mathbb{H}$. Then we have $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2 = \mathbb{Z}\omega_1' + \mathbb{Z}\omega_2'$ and $\omega_1'/\omega_2' \in \mathbb{H}$ if and only if

$$\begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix} = \gamma \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} \quad \textit{for some } \gamma \in \mathrm{SL}_2(\mathbb{Z}).$$

Let $\omega_1, \omega_2, \omega_1', \omega_2' \in \mathbb{C}^\times$ with $z := \omega_1/\omega_2, z' := \omega_1'/\omega_2' \in \mathbb{C} - \mathbb{R}$ and $\gamma \in \mathrm{GL}_2(\mathbb{Z})$ satisfying (1.3), then

$$z' = \frac{a\omega_1 + b\omega_2}{c\omega_1 + d\omega_2} = \frac{az+b}{cd+d}.$$

Note that the formula above is still well defined if we generalize from $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ to γ in

$$\mathrm{GL}_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{R}, ad - bc \neq 0 \right\}.$$

Now for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$ and $z \in \mathbb{C} - \mathbb{R}$, we write

$$\gamma z := \frac{az+b}{cz+d}$$

and introduce the *factor of automorphy*

$$j(\gamma, z) := cz + d \in \mathbb{C}^\times.$$

Proposition 1.2. *Let $\gamma, \gamma' \in \mathrm{GL}_2(\mathbb{R})$ and $z \in \mathbb{C} - \mathbb{R}$. Then*

(i)
$$\Im(\gamma z) = \frac{\det(\gamma)\Im z}{|j(\gamma, z)|^2};$$

(ii)
$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} z = z;$$

(iii)
$$\gamma(\gamma' z) = (\gamma\gamma')z.$$

Proof. For (i) write $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})$. We calculate

$$\begin{aligned} \Im(\gamma z) &= \Im \frac{az+b}{cz+d} \\ &= \Im \frac{(az+b)(c\bar{z}+d)}{|cz+d|^2} \\ &= \frac{\Im(ac|z|^2 + bd + adz + bc\bar{z})}{|cz+d|^2} \\ &= \frac{(ad-bc)\Im z}{|cz+d|^2} \\ &= \frac{\det(\gamma)\Im z}{|j(\gamma,z)|^2}. \end{aligned}$$

Part (ii) is trivial. The proof of part (iii) is a straightforward calculation; see Exercise 1.2. □

We also consider

$$\mathrm{GL}_2^+(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{R}, ad - bc > 0 \right\}.$$

Corollary 1.3. *(i) The map*

$$\begin{aligned} \mathrm{GL}_2(\mathbb{R}) \times \mathbb{C} - \mathbb{R} &\longrightarrow \mathbb{C} - \mathbb{R} \\ (\gamma, z) &\longmapsto \gamma z, \end{aligned}$$

defines an action of the group $\mathrm{GL}_2(\mathbb{R})$ on the set $\mathbb{C} - \mathbb{R}$.

(ii) The map

$$\begin{aligned} \mathrm{GL}_2^+(\mathbb{R}) \times \mathbb{H} &\longrightarrow \mathbb{H} \\ (\gamma, z) &\longmapsto \gamma z, \end{aligned}$$

defines an action of the group $\mathrm{GL}_2^+(\mathbb{R})$ *on the set* $\mathbb{H}$.

We make the trivial, but important remarks that the actions described above induce an action of $\mathrm{GL}_2(\mathbb{Z})$ on $\mathbb{C}-\mathbb{R}$ and an action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$. The latter will be our primary focus (as well as its restriction to so-called congruence subgroups later on, which will be discussed in Chapter 3). One more subgroup of $\mathrm{GL}_2^+(\mathbb{R})$ of (some) interest to us (together with its induced action on $\mathbb{H}$) is

$$\mathrm{SL}_2(\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \,\middle|\, a, b, c, d \in \mathbb{R}, ad - bc = 1 \right\}.$$

Let us come back to the transformation properties of the function f defined in (1.2).

Proposition 1.4. *Let* $\mathcal{F}\colon \mathcal{L} \to \mathbb{C}$ *be a homogeneous function of weight* $k \in \mathbb{Z}$ *and define the function*

$$f\colon \mathbb{H} \to \mathbb{C} \quad \text{by } z \mapsto \mathcal{F}(\Lambda_z).$$

Then

$$f(\gamma z) = j(\gamma, z)^k f(z) \quad \text{for all } \gamma \in \mathrm{SL}_2(\mathbb{Z}) \text{ and } z \in \mathbb{H}. \tag{1.4}$$

Proof. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and $z \in \mathbb{H}$. By Lemma 1.1, we have

$$\mathbb{Z}(az+b) + \mathbb{Z}(cz+d) = \mathbb{Z}z + \mathbb{Z}.$$

This gives us

$$\Lambda_{\gamma z} = \mathbb{Z}\frac{az+b}{cz+d} + \mathbb{Z} = (cz+d)^{-1}(\mathbb{Z}(az+b) + \mathbb{Z}(cz+d)) = (cz+d)^{-1}(\mathbb{Z}z + \mathbb{Z}) = j(\gamma, z)^{-1}\Lambda_z.$$

So finally,

$$f(\gamma z) = \mathcal{F}(\Lambda_{\gamma z}) = \mathcal{F}(j(\gamma, z)^{-1}\Lambda_z) = j(\gamma, z)^k \mathcal{F}(\Lambda_z) = j(\gamma, z)^k f(z).$$

□

Warning. Many authors work with the *projective modular group*

$$\mathrm{PSL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z}) \Big/ \left\{ \pm \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

instead of $\mathrm{SL}_2(\mathbb{Z})$. In these notes, we will mostly phrase the results in terms of $\mathrm{SL}_2(\mathbb{Z})$, but we will sometimes also give the analogous results for $\mathrm{PSL}_2(\mathbb{Z})$.

Remark. We will see in Theorem 1.5 below that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices

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

These satisfy the relations

$$S^4 = 1, \quad (ST)^3 = S^2 \quad \text{in } \mathrm{SL}_2(\mathbb{Z}).$$

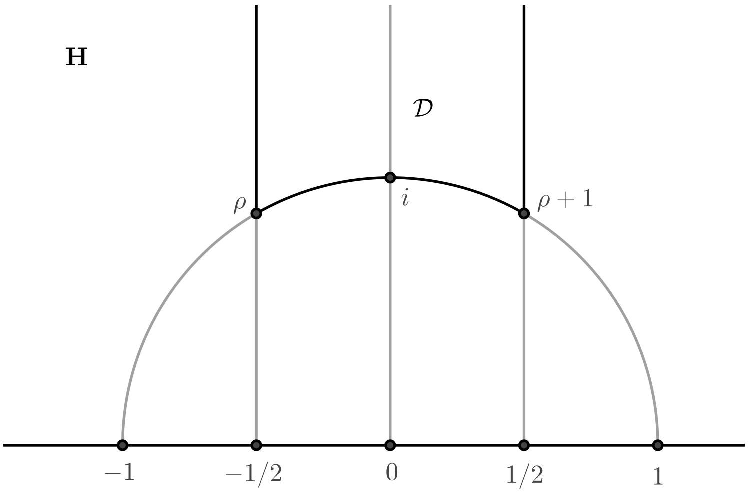
Moreover, one can show that these generate all relations, i.e. that $\langle S, T \mid S^4, S^2(ST)^3 \rangle$ is a presentation of the group $\mathrm{SL}_2(\mathbb{Z})$.

1.3 A fundamental domain

Let $\mathcal{D}$ be the closed subset of $\mathbb{H}$ given by

$$\mathcal{D} := \{z \in \mathbb{H} \mid -1/2 \le \Re z \le 1/2 \text{ and } |z| \ge 1\}.$$

It looks as follows:



Here we write ρ for the unique third root of unity in the upper half-plane, i.e.

$$\rho = \exp(2\pi i/3) = \frac{-1 + i\sqrt{3}}{2}.$$

Theorem 1.5. *Let $\mathcal{D}$ be the subset of $\mathbb{H}$ defined above.*

1. *Every point in $\mathbb{H}$ is equivalent, under the action of $\mathrm{SL}_2(\mathbb{Z})$, to a point of $\mathcal{D}$.*
2. *If $z, z' \in \mathcal{D}$ are two distinct points that are in the same $\mathrm{SL}_2(\mathbb{Z})$-orbit, then either $z' = z \pm 1$ (so z, z' are on the vertical parts of the boundary of $\mathcal{D}$) or $z' = -1/z$ (so z, z' are on the circular part of the boundary of $\mathcal{D}$).*
3. *Let z be in $\mathcal{D}$, and let H_z be the stabiliser of z in $\mathrm{SL}_2(\mathbb{Z})$. Then H_z is*

$$\begin{cases} \textit{cyclic of order } 6 \textit{ generated by } ST = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right) & \textit{if } z = \rho; \\ \textit{cyclic of order } 6 \textit{ generated by } TS = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right) & \textit{if } z = \rho + 1; \\ \textit{cyclic of order } 4 \textit{ generated by } S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) & \textit{if } z = i; \\ \textit{cyclic of order } 2 \textit{ generated by } \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) & \textit{otherwise.} \end{cases}$$

4. *The group $\mathrm{SL}_2(\mathbb{Z})$ is generated by S and T.*

Proof. Let z be any point in $\mathbb{H}$. We consider the imaginary part of γz for all $\gamma \in \langle S, T\rangle$. According to Proposition 1.2 part (i) this imaginary part is

$$\Im(\gamma z) = \frac{\Im z}{|cz + d|^2} \quad \text{if } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Given z, there are only finitely many $(c, d) \in \mathbb{Z}^2$, and in particular only finitely many $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \langle S, T\rangle$, such that $|cz + d| < 1$. This implies that there exists some $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ such that

$$|cz + d| \le |c'z + d'| \quad \text{for all } \gamma' = \begin{pmatrix} a' & b' \\ c' & c' \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),$$

or equivalently

$$\Im(\gamma z) \ge \Im(\gamma' z) \quad \text{for all } \gamma' \in \mathrm{SL}_2(\mathbb{Z}).$$

By multiplying γ from the left by a power of T, which has the effect of translating γz by an integer, we may in addition choose γ such that

$$-1/2 \le \Re(\gamma z) \le 1/2.$$

We claim that this γ satisfies

$$|\gamma z| \ge 1.$$

Namely, by the choice of γ, we have

$$\begin{aligned}\Im(\gamma z) &\ge \Im(S\gamma z)\\ &= \Im(-1/\gamma z)\\ &= \frac{\Im(\gamma z)}{|\gamma z|^2}.\end{aligned}$$

This implies $|\gamma z| \ge 1$, and hence $\gamma z \in \mathcal{D}$.

We conclude that for any $z \in \mathbb{H}$ there exists $\gamma \in \langle S, T\rangle$ such that $\gamma z \in \mathcal{D}$. In particular, this implies (1).

To prove (2), let $z, z' \in \mathcal{D}$ be distinct points in the same $\mathrm{SL}_2(\mathbb{Z})$-orbit. We may assume $\Im z' \ge \Im z$. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ be such that $z' = \gamma z$; in particular,

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

so $|cz+d| \le 1$. By the identity

$$|cz+d|^2 = |cx+d|^2 + |cy|^2 \quad (z = x+iy)$$

and the fact that $y > 1/2$ since $z \in D$, this is only possible if $|c| \le 1$.

If $c = 0$, then the condition $ad - bc = 1$ implies $a = d = \pm 1$, and hence $z' = z \pm b$. Because $\Re z$ and $\Re z'$ both lie in $[-1/2, 1/2]$, this implies $z = z' \pm 1$ and $\Re z = \pm 1/2$.

If $c = 1$, then we have

$$1 \ge |cz+d| = |z+d|;$$

this is only possible if $|z| = 1$ and $d = 0$, if $z = \rho$ and $d = 1$, or if $z = \rho + 1$ and $d = -1$. The case $d = 0$ implies $b = -1$ and $z' = \frac{az-1}{z+0} = a - 1/z$; this is only possible if $a = 0$, if $z = \rho$ and $a = -1$, or if $z = \rho + 1$ and $a = 1$. The case $d = 1$ implies $z = \rho$ and $a - b = 1$; this is only possible if $(a, b) = (1, 0)$ or $(a, b) = (0, -1)$.

The case $c = -1$ is completely analogous, since $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $-\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ act in the same way on $\mathbb{H}$.

Altogether, we obtain the following pairs (γ, z) where z and $z' = \gamma z$ are both in $\mathcal{D}$:

γ	z	$z' = \gamma z$	fixed points
$\pm\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$	all $z \in \mathcal{D}$	z	all $z \in \mathcal{D}$
$\pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$	$\Re z = -1/2$	$z+1$	none
$\pm\left(\begin{smallmatrix} 1 & -1 \\ 0 & 1 \end{smallmatrix}\right)$	$\Re z = 1/2$	$z-1$	none
$\pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$	$\|z\| = 1$	$-1/z$	i
$\pm\left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$	ρ	ρ	ρ
$\pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$	ρ	ρ	ρ
$\pm\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$	$\rho+1$	$\rho+1$	$\rho+1$
$\pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & -1 \end{smallmatrix}\right)$	$\rho+1$	$\rho+1$	$\rho+1$

Part (2) and (3) of the theorem can be read off from this table. It remains to show (4).

We choose any fixed z in the interior of $\mathcal{D}$. Let $\gamma \in \mathrm{SL}_2(\mathbb{Z})$; we have to show that γ is in $\langle S, T\rangle$. As we have seen in the first part of the proof, there exists $\gamma_0 \in \langle S, T\rangle$ such that $\gamma_0(\gamma z) \in \mathcal{D}$. This means that both z and $(\gamma_0\gamma)z$ lie in $\mathcal{D}$, and since z is not on the boundary of $\mathcal{D}$, part (3) implies $\gamma_0\gamma = \pm\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. We conclude that $\gamma = \pm\gamma_0$ is in $\langle S, T\rangle$. □

1.4 Exercises

Exercise 1.1. Prove Lemma 1.1. (For part (ii), you may use Proposition 1.2.)

Exercise 1.2. Prove part (iii) of Proposition 1.2.

Exercise 1.3.

(a) Show that the standard action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$ is transitive.

(b) Let $\gamma = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$ be an element of $\mathrm{SL}_2(\mathbb{R})$ with $\gamma \ne \pm\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. Prove that γ has exactly one fixed point in $\mathbb{H}$ if $|a+d| < 2$, and no fixed points in $\mathbb{H}$ otherwise.

Exercise 1.4.

(a) Let K be the stabiliser of $i \in \mathbb{H}$ under the standard action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$. Show that

$$K = \left\{ \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \,\middle|\, a, b \in \mathbb{R}, a^2 + b^2 = 1 \right\} \quad (= \mathrm{SO}_2(\mathbb{R})).$$

(b) Prove that there is a bijection

$$\begin{aligned} \mathrm{SL}_2(\mathbb{R})/K &\xrightarrow{\sim} \mathbb{H} \\ \gamma K &\longmapsto \gamma i. \end{aligned}$$

Exercise 1.5. Visit CoCalc on `https://cocalc.com/` and create an account.

Chapter 2

Modular forms for $\mathrm{SL}_2(\mathbb{Z})$

2.1 Definition of modular forms

Definition. Let f be a meromorphic function on $\mathbb{H}$. We say that f is *weakly modular* of weight $k \in \mathbb{Z}$ if it satisfies

$$f(\gamma z) = (cz+d)^k f(z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \text{ and } z \in \mathbb{H}.$$

Note that this is exactly the transformation from (1.4). This definition can be reformulated in several ways. To do this, we first introduce a right action of the group $\mathrm{SL}_2(\mathbb{R})$ on the set of meromorphic functions on $\mathbb{H}$. This action is called the *slash operator of weight k* and denoted by $(f,\gamma) \mapsto f|_k\gamma$. It is defined by

$$(f|_k\gamma)(z) := (cz+d)^{-k} f(\gamma z) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \text{ and } z \in \mathbb{H}. \tag{2.1}$$

For the proof that this is an action, see Exercise 2.1.

Saying that f is weakly modular is then equivalent to saying that f is invariant under the weight k action of $\mathrm{SL}_2(\mathbb{Z})$. Since $\mathrm{SL}_2(\mathbb{Z})$ is generated by the two matrices S and T, it suffices to check invariance under these two matrices. It is easy to check that invariance by T is equivalent to

$$f(z+1) = f(z) \quad \text{for all } z \in \mathbb{H},$$

and that invariance by S is equivalent to

$$f(-1/z) = z^k f(z) \quad \text{for all } z \in \mathbb{H}.$$

Remark. The property of weak modularity, applied to the matrix $\gamma = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, implies that

$$f(z) = (-1)^k f(z) \quad \text{for all } z \in \mathbb{H}.$$

So if k is odd, then the only meromorphic function on $\mathbb{H}$ that is weakly modular of weight k is the zero function.

We will make extensive use of the following notation:

$$\begin{aligned} q\colon \mathbb{H} &\to \mathbb{C} \\ z &\mapsto \exp(2\pi i z). \end{aligned}$$

Warning. Especially in older sources, $q(z)$ is defined to be $\exp(\pi i z)$ instead.

Let f be weakly modular of weight k. Applying the definition to the matrix $\gamma = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ shows that f is periodic with period 1:

$$f(z+1) = f(z).$$

This implies that f can be written in the form

$$f(z) = \tilde{f}(\exp(2\pi i z))$$

where $\tilde{f}$ is a meromorphic function on the punctured unit disc

$$\mathbb{D}^* := \{q \in \mathbb{C} \mid 0 < |q| < 1\}.$$

In other words, $\tilde{f}$ is defined by

$$\tilde{f}(q) := f\left(\frac{\log q}{2\pi i}\right).$$

The logarithm is multi-valued, but choosing a different value of the logarithm comes down to adding an integer multiple of $2\pi i$ to $\log q$, hence an integer to $\frac{\log q}{2\pi i}$. Since f is periodic with period 1, this formula for $\tilde{f}(q)$ does not depend on the chosen value of the logarithm.

Definition. Let f be a meromorphic function on $\mathbb{H}$ that is weakly modular of weight k. We say that f is *meromorphic at infinity* (or *at the cusp*) if $\tilde{f}$ can be continued to a meromorphic function on the open unit disc

$$\mathbb{D} = \{q \in \mathbb{C} \mid |q| < 1\}.$$

We say that f is *holomorphic at infinity* (or *at the cusp*) if this meromorphic continuation of $\tilde{f}$ is holomorphic at $q = 0$.

The condition that $\tilde{f}$ can be continued to a meromorphic on $\mathbb{D}$ is equivalent to the condition that $\tilde{f}$ can be written as a Laurent series

$$\tilde{f}(q) = \sum_{n=-\infty}^{\infty} a_n q^n \quad (a_n \in \mathbb{C},\ a_n = 0 \text{ for } n \text{ sufficiently negative})$$

that is convergent on $\{q \in \mathbb{C} \mid 0 < |q| < \epsilon\}$ for some $\epsilon > 0$. With this notation, f is holomorphic at infinity if and only if $a_n = 0$ for all $n < 0$. If f is holomorphic at infinity, we define the *value of f at infinity* as

$$f(\infty) := \tilde{f}(0) = a_0.$$

Definition. Let k be an integer. A *modular form* of weight k (for the group $\mathrm{SL}_2(\mathbb{Z})$) is a holomorphic function $f\colon \mathbb{H} \to \mathbb{C}$ that is weakly modular of weight k and holomorphic at infinity. A *cusp form* of weight k (for the group $\mathrm{SL}_2(\mathbb{Z})$) is a modular form f of weight k satisfying $f(\infty) = 0$.

2.2 Examples of modular forms: Eisenstein series

Let k be an even integer with $k \geq 4$. We define the *Eisenstein series* of weight k (for $\mathrm{SL}_2(\mathbb{Z})$) by

$$\begin{aligned} G_k\colon \mathbb{H} &\longrightarrow \mathbb{C} \\ z &\longmapsto \mathcal{G}_k(\Lambda_z) = \sum_{\substack{m,n\in\mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+n)^k}. \end{aligned}$$

Proposition 2.1. *The series above converges absolutely and uniformly on subsets of $\mathbb{H}$ of the form*

$$R_{r,s} = \{x + iy \mid |x| \leq r, y \geq s\}.$$

Proof. Let $z = x + iy \in R_{r,s}$ be given. We have the inequality

$$|mz + n|^2 = (mx + n)^2 + m^2y^2 \geq (mx + n)^2 + m^2s^2.$$

For fixed m and n, we distinguish the cases $|n| \leq 2r|m|$ and $|n| \geq 2r|m|$. In the first case, we have

$$|mz + n|^2 \geq m^2s^2 \geq \frac{s^2}{2}m^2 + \frac{s^2}{2(2r)^2}n^2 \geq \min\{s^2/2, s^2/(8r^2)\}(m^2 + n^2).$$

In the second case, the triangle inequality implies

$$|mz + n|^2 \geq (|mx| - |n|)^2 + m^2s^2 \geq (|n|/2)^2 + m^2s^2 \geq \min\{1/4, s^2\}(m^2 + n^2).$$

Combining both cases and putting

$$c = \min\{s^2/2, s^2/(8r^2), 1/4, s^2\},$$

we get the inequality

$$|mz + n| \geq c(m^2 + n^2)^{1/2} \quad \text{for all } m, n \in \mathbb{Z}, z \in R_{r,s}.$$

This implies that for any $z \in R_{r,s}$ we have

$$|G_k(z)| \leq \frac{1}{c^k} \sum_{(m,n)\neq(0,0)} \frac{1}{(m^2 + n^2)^{k/2}}.$$

We rearrange the sum by grouping together, for each fixed $j = 1, 2, 3, \ldots$, all pairs (m, n) with $\max\{|m|, |n|\} = j$. We note that for each j there are $8j$ such pairs (m, n), each of which satisfies

$$j^2 \leq m^2 + n^2 \quad (\leq 2j^2).$$

From this we obtain

$$\begin{aligned} |G_k(z)| &\leq \frac{1}{c^k} \sum_{j=1}^{\infty} \frac{8j}{j^k} \\ &= \frac{8}{c^k} \sum_{j=1}^{\infty} \frac{1}{j^{k-1}}, \end{aligned}$$

which is finite and independent of $z \in R_{r,s}$. □

The proposition above implies that the series defining $G_k(z)$ converges to a holomorphic function on $\mathbb{H}$.

Theorem 2.2. *For every even integer $k \geq 4$, the function*

$$G_k \colon \mathbb{H} \to \mathbb{C}$$

is a modular form of weight k.

Proof. As we have just seen, G_k is holomorphic on $\mathbb{H}$. That it has the correct transformation behaviour under the action of $\mathrm{SL}_2(\mathbb{Z})$ follows from Proposition 1.4.

It remains to check that $G_k(z)$ is holomorphic at infinity. We will do this in the next section by calculating the q-expansion of G_k. □

2.3 The q-expansions of Eisenstein series

We will need special values of the *Riemann zeta function.* This is a complex-analytic function defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \quad \text{for } s \in \mathbb{C} \text{ with } \Re s > 1. \tag{2.2}$$

We will only need the cases where s equals an even positive integer k.

We will also use the following notation for the sum of the t-th powers of the divisors of an integer n:

$$\sigma_t(n) = \sum_{\substack{d \mid n \\ d>0}} d^t.$$

We rewrite the infinite sum defining $G_k(z)$ as follows:

$$\begin{aligned} G_k(z) &= \sum_{\substack{m,n\in\mathbb{Z} \\ (m,n)\neq(0,0)}} \frac{1}{(mz+n)^k} \\ &= \sum_{n\neq 0} \frac{1}{n^k} + \sum_{m\neq 0} \sum_{n\in\mathbb{Z}} \frac{1}{(mz+n)^k}. \end{aligned}$$

Since k is even, we can further rewrite this (using the definition above of the Riemann zeta function) as

$$\begin{aligned} G_k(z) &= 2\sum_{n=1}^{\infty} \frac{1}{n^k} + 2\sum_{m=1}^{\infty} \sum_{n\in\mathbb{Z}} \frac{1}{(mz+n)^k} \\ &= 2\zeta(k) + 2\sum_{m=1}^{\infty} \sum_{n\in\mathbb{Z}} \frac{1}{(mz+n)^k}. \end{aligned} \tag{2.3}$$

Proposition 2.3. *Let $k \geq 2$ be an integer. Then we have*

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{d=1}^{\infty} d^{k-1} \exp(2\pi i d z) \quad \textit{for all } z \in \mathbb{H}.$$

Proof. We start with the classical formula (A.1) for the cotangent function:

$$\pi \frac{\cos(\pi z)}{\sin(\pi z)} = \frac{1}{z} + \sum_{n=1}^{\infty} \left(\frac{1}{z-n} + \frac{1}{z+n} \right) \quad \text{for all } z \in \mathbb{C} - \mathbb{Z}.$$

On the other hand, using the identity $\exp(\pm iz) = \cos z \pm i \sin z$ and the geometric series $1/(1-q) = \sum_{d=0}^{\infty} q^d$ for $|q| < 1$, we can rewrite the left-hand side for $z \in \mathbb{H}$ as

$$\begin{aligned} \pi \frac{\cos(\pi z)}{\sin(\pi z)} &= \pi i \frac{\exp(\pi i z) + \exp(-\pi i z)}{\exp(\pi i z) - \exp(-\pi i z)} \\ &= -\pi i - 2\pi i \frac{\exp(2\pi i z)}{1 - \exp(2\pi i z)} \\ &= -\pi i - 2\pi i \sum_{d=1}^{\infty} \exp(2\pi i d z). \end{aligned}$$

Combining the equations above, we obtain

$$\frac{1}{z} + \sum_{n=1}^{\infty} \left(\frac{1}{z-n} + \frac{1}{z+n} \right) = -\pi i - 2\pi i \sum_{d=1}^{\infty} \exp(2\pi i d z) \quad \text{for all } z \in \mathbb{H}. \tag{2.4}$$

Taking derivatives gives

$$\sum_{n\in\mathbb{Z}} \frac{1}{(z+n)^2} = (2\pi i)^2 \sum_{d=1}^{\infty} d\exp(2\pi i d z),$$

which is the desired equality in the case $k = 2$. The formula for general $k \geq 2$ is proved by induction. □

Applying the fact above to the last sum in (2.3), and using the identity $(-2\pi i)^k = (2\pi i)^k$ for k even, we deduce the following formula for all even $k \geq 4$:

$$\begin{aligned} G_k(z) &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{m=1}^{\infty}\sum_{d=1}^{\infty} d^{k-1}\exp(2\pi i d m z) \\ &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty}\sum_{d\mid n} d^{k-1}\exp(2\pi i n z) \\ &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n. \end{aligned} \tag{2.5}$$

(In replacing the sum over (d, m) by a sum over (d, n), we have taken $n = dm$.)

The *Bernoulli numbers* are the rational numbers B_k ($k \geq 0$) defined by the equation

$$\frac{t}{\exp(t)-1} = \sum_{k=0}^{\infty} \frac{B_k}{k!} t^k \in \mathbb{Q}[[t]].$$

We have

$$B_k \neq 0 \iff k = 1 \text{ or } k \text{ is even};$$

see Exercise 2.3. Furthermore, the first few non-zero Bernoulli numbers are

$$B_0 = 1,\quad B_1 = -\frac{1}{2},\quad B_2 = \frac{1}{6},\quad B_4 = -\frac{1}{30},\quad B_6 = \frac{1}{42},$$
$$B_8 = -\frac{1}{30},\quad B_{10} = \frac{5}{66},\quad B_{12} = -\frac{691}{2730},\quad B_{14} = \frac{7}{6}.$$

In Exercise 2.3, you will prove the formula

$$\zeta(k) = -\frac{(2\pi i)^k B_k}{2\cdot k!} \quad \text{for } k \geq 2 \text{ even.}$$

Substituting this into the formula (2.5) for $G_k(z)$, we obtain

$$G_k(z) = -\frac{(2\pi i)^k B_k}{k!} + 2\frac{(2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n.$$

It is useful to rescale the Eisenstein series G_k so that the coefficient of q becomes 1. This leads to the definition

$$E_k(z) = \frac{(k-1)!}{2(2\pi i)^k} G_k(z).$$

This immediately simplifies to

$$E_k(z) = -\frac{B_k}{2k} + \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n. \tag{2.6}$$

Note in particular that all coefficients in this q-expansion are rational numbers.

Remark. Another common normalisation of E_k is such that the constant coefficient (as opposed to the coefficient of q) becomes 1.

2.4 The Eisenstein series of weight 2

So far we have only defined Eisenstein series of weight k for $k \geq 4$. The construction does not generalise completely to the case $k = 2$, because the series

$$\sum_{\substack{(m,n)\in\mathbb{Z}^2\\(m,n)\neq(0,0)}} \frac{1}{(mz+n)^2}$$

fails to converge.

As it turns out, it is still useful to define a holomorphic function G_2 on $\mathbb{H}$ by the formula (2.3) for $k = 2$, and to define

$$E_2(z) = -\frac{1}{8\pi^2}G_2(z).$$

Then the formulae (2.5) and (2.6) are also valid for $k = 2$. One has to be careful, however, because the double series in (2.3) does not converge absolutely and the functions G_2 and E_2 are not modular forms.

Proposition 2.4. *The functions G_2 and E_2 satisfy the transformation formulae*

$$z^{-2}G_2(-1/z) = G_2(z) - \frac{2\pi i}{z}. \tag{2.7}$$

and

$$z^{-2}E_2(-1/z) = E_2(z) - \frac{1}{4\pi i z}. \tag{2.8}$$

The proof is based on following lemma, which gives an example of two double series that contain the same terms but sum to different values due to the order of summation being different.

Lemma 2.5. *For all $z \in \mathbb{H}$, we have*

$$\sum_{m\neq 0}\sum_{n\in\mathbb{Z}}\left(\frac{1}{mz+n} - \frac{1}{mz+n+1}\right) = 0 \tag{2.9}$$

and

$$\sum_{n\in\mathbb{Z}}\sum_{m\neq 0}\left(\frac{1}{mz+n} - \frac{1}{mz+n+1}\right) = -\frac{2\pi i}{z}. \tag{2.10}$$

Proof. We start with the telescoping sum

$$\sum_{-N\leq n<N}\left(\frac{1}{mz+n} - \frac{1}{mz+n+1}\right) = \frac{1}{mz-N} - \frac{1}{mz+N}.$$

Using this, we compute the inner sum of the first double series as

$$\begin{aligned}\sum_{n\in\mathbb{Z}}\left(\frac{1}{mz+n} - \frac{1}{mz+n+1}\right) &= \lim_{N\to\infty}\sum_{-N\leq n<N}\left(\frac{1}{mz+n} - \frac{1}{mz+n+1}\right)\\ &= \lim_{N\to\infty}\left(\frac{1}{mz-N} - \frac{1}{mz+N}\right)\\ &= 0,\end{aligned}$$

which implies the first identity.

On the other hand, again using the telescoping sum above, we can write the second double series as

$$\begin{aligned}\sum_{n\in\mathbb{Z}}\sum_{m\neq 0}\biggl(\frac{1}{mz+n}-\frac{1}{mz+n+1}\biggr) &= \lim_{N\to\infty}\sum_{-N\le n<N}\sum_{m\neq 0}\biggl(\frac{1}{mz+n}-\frac{1}{mz+n+1}\biggr)\\ &= \lim_{N\to\infty}\sum_{m\neq 0}\sum_{-N\le n<N}\biggl(\frac{1}{mz+n}-\frac{1}{mz+n+1}\biggr)\\ &= \lim_{N\to\infty}\sum_{m\neq 0}\biggl(\frac{1}{mz-N}-\frac{1}{mz+N}\biggr),\end{aligned}$$

and we cannot interchange the limit and the sum, because the series fails to converge uniformly when N varies in any interval of the form $[M,\infty)$. In fact, using (2.4) and the fact that $-N/z\in\mathbb{H}$, we can rewrite the sum over m as

$$\begin{aligned}\sum_{m\neq 0}\biggl(\frac{1}{mz-N}-\frac{1}{mz+N}\biggr) &= \sum_{m=1}^{\infty}\biggl(\frac{1}{mz-N}+\frac{1}{-mz-N}-\frac{1}{mz+N}-\frac{1}{-mz+N}\biggr)\\ &= \frac{2}{z}\sum_{m=1}^{\infty}\biggl(\frac{1}{-N/z-m}+\frac{1}{-N/z+m}\biggr)\\ &= \frac{2}{z}\biggl(\frac{z}{N}-\pi i-2\pi i\sum_{d=1}^{\infty}\exp(-2\pi i dN/z)\biggr)\end{aligned}$$

The series on the right-hand side converges uniformly for N in the interval $[1,\infty)$, because for all $N\ge 1$ the tail of the series for $d\ge D$ can be bounded using the triangle inequality as

$$\sum_{d=D}^{\infty}\bigl|\exp(-2\pi i dN/z)\bigr|\le\sum_{d=D}^{\infty}|q|^d\quad\text{with } q=\exp(-2\pi i/z);$$

the right-hand side is a geometric series that does not depend on N and tends to 0 as $D\to\infty$, since $|q|<1$. We can therefore interchange the limit and the sum, and we obtain

$$\begin{aligned}\sum_{n\in\mathbb{Z}}\sum_{m\neq 0}\biggl(\frac{1}{mz+n}-\frac{1}{mz+n+1}\biggr) &= \lim_{N\to\infty}\frac{2}{z}\biggl(\frac{z}{N}-\pi i-2\pi i\sum_{d=1}^{\infty}\exp(-2\pi i dN/z)\biggr)\\ &= -\frac{2\pi i}{z},\end{aligned}$$

which is what we had to prove. □

Proof of Proposition 2.4. We recall that

$$G_2(z)=2\zeta(2)+\sum_{m\neq 0}\sum_{n\in\mathbb{Z}}\frac{1}{(mz+n)^2}.$$

Subtracting the identity (2.9) and simplifying, we obtain the alternative expression

$$G_2(z)=2\zeta(2)+\sum_{m\neq 0}\sum_{n\in\mathbb{Z}}\frac{1}{(mz+n)^2(mz+n+1)}.$$

On the other hand, we have

$$\begin{aligned}z^{-2}G_2(-1/z) &= 2\zeta(2)z^{-2}+\sum_{m\neq 0}\sum_{n\in\mathbb{Z}}\frac{1}{(nz-m)^2}\\ &= 2\zeta(2)+\sum_{m\in\mathbb{Z}}\sum_{n\neq 0}\frac{1}{(nz-m)^2}\\ &= 2\zeta(2)+\sum_{n\in\mathbb{Z}}\sum_{m\neq 0}\frac{1}{(mz+n)^2};\end{aligned}$$

note that in the last step we just relabelled the variables, but did not change the summation order. Subtracting the identity (2.10) and simplifying, we obtain

$$z^{-2}G_2(-1/z) + \frac{2\pi i}{z} = 2\zeta(2) + \sum_{n\in\mathbb{Z}}\sum_{m\neq 0} \frac{1}{(mz+n)^2(mz+n+1)}.$$

By an argument similar to that used in the proof of Proposition 2.1, the double series on the right-hand side is absolutely convergent. We may therefore change the summation order. This shows that the right-hand side is equal to $G_2(z)$, which proves (2.7). Finally, (2.8) follows from (2.7) and the definition (2.4) of E_2. □

2.5 More examples: the modular form Δ and the modular function j

We define a function $\Delta\colon \mathbb{H} \to \mathbb{C}$ by

$$\Delta = \frac{(240E_4)^3 - (-504E_6)^2}{1728}. \tag{2.11}$$

Since E_4 and E_6 are modular forms of weight 4 and 6, respectively, Δ is a modular form of weight 12. Moreover, the specific linear combination of E_4^3 and E_6^2 is chosen such that the constant term of the q-expansion of Δ vanishes. This means that Δ is a cusp form of weight 12.

Using the known q-expansions of E_4 and E_6, one can compute the q-expansion of Δ as

$$\Delta = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 - 16744q^7 + \cdots$$

An infinite product expansion for Δ is given in the next section.

Furthermore, we define the *j-function* as

$$j(z) = \frac{(240E_4)^3}{\Delta}.$$

This is not a modular form (since Δ vanishes at infinity but E_4 does not, the j-function has a pole at infinity). However, the fact that the j-function is a quotient of two modular forms of the same weight (12 in this case) implies that it is a *modular function*, i.e. it satisfies $j(\gamma z) = j(z)$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and $z \in \mathbb{H}$) and is meromorphic on $\mathbb{H}$ and at infinity.

The j-function is extremely important in the theory of lattices and elliptic curves. For example, one can define the j-invariant $j(\Lambda)$ of a lattice $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, where $\omega_1/\omega_2 \in \mathbb{H}$, by $j(\omega_1/\omega_2)$ (we use the same j to denote the different functions); one can then show that the j-invariant gives a bijection

$$\begin{aligned} \{\text{lattices in } \mathbb{C}\}/(\text{homothety}) &\xrightarrow{\sim} \mathbb{C} \\ [\Lambda] &\longmapsto j(\Lambda). \end{aligned}$$

The q-expansion of j looks like

$$j(z) = q^{-1} + 744 + 196884q + 21493760q^2 + 864299970q^3 + \cdots$$

The coefficients of this series are famous for their role in the theory of *monstrous moonshine* (Conway, Norton, Borcherds et al.), which links these coefficients to the representation theory of the *monster group*.

2.6 The η-function

We define the *Dedekind eta function*, using $q_{24} := \exp(2\pi i z/24)$, by

$$\begin{aligned} \eta \colon \mathbb{H} &\longrightarrow \mathbb{C} \\ z &\longmapsto q_{24} \prod_{n=1}^{\infty} (1 - q^n) \end{aligned}$$

Since $\sum_{n=1}^{\infty} -q^n$ converges absolutely and uniformly on compact subsets of $\mathbb{H}$ (because $|q| < 1$), a standard result from complex analysis about infinite products (Theorem A.5) gives us that η converges to a holomorphic functions on $\mathbb{H}$ and that its zeroes coincide with the zeroes of the factors of the infinite product. Since these factors obviously do not have zeroes on $\mathbb{H}$, we arrive at the following result.

Proposition 2.6. *The Dedekind eta function $\eta \colon \mathbb{H} \to \mathbb{C}$ is holomorphic and non-vanishing.*

The transformation properties of η under the action of $\mathrm{SL}_2(\mathbb{Z})$ follow from the trivial observation that for all $z \in \mathbb{H}$ we have

$$\eta(z+1) = \exp(2\pi i/24)\eta(z)$$

and the fundamental transformation property below, which follows from the transformation property of E_2.

Proposition 2.7. *For all $z \in \mathbb{H}$ we have*

$$\eta(-1/z) = \sqrt{-iz}\eta(z)$$

where the branch of $\sqrt{-iz}$ is taken to have positive real part.

Proof. Let $z \in \mathbb{H}$. By invoking Theorem A.5 again, we may calculate the logarithmic derivative of η term by term. So we arrive at

$$\begin{aligned} \frac{d}{dz} \log(\eta(z)) &= \frac{2\pi i}{24} + \sum_{n=1}^{\infty} \frac{-2\pi i n q^n}{1 - q^n} = \frac{\pi i}{12} - 2\pi i \sum_{n=1}^{\infty} n \sum_{m=1}^{\infty} q^{nm} \\ &= \frac{\pi i}{12} - 2\pi i \sum_{m,n=1}^{\infty} n q^{nm} = \frac{\pi i}{12} - 2\pi i \sum_{l=1}^{\infty} \sigma(l) q^l \\ &= -2\pi i E_2(z). \end{aligned}$$

Together with the transformation property (2.8) of E_2, we arrive at

$$\begin{aligned} \frac{d}{dz} \log(\eta(-1/z)) &= -2\pi i z^{-2} E_2(-1/z) \\ &= -2\pi i E_2(z) + \frac{1}{2z} \\ &= \frac{d}{dz} \log(\sqrt{-iz}\eta(z)). \end{aligned}$$

This shows that there is a constant $c \in \mathbb{C}$ such that for all $z \in \mathbb{H}$ we have $\eta(-1/z) = c\sqrt{-iz}\eta(z)$. Specializing at $z = i$ shows that $c = 1$, which completes the proof of the proposition. □

The η function can be used to obtain an infinite product expansion for the modular form Δ introduced in the previous section. Define $f \colon \mathbb{H} \to \mathbb{C}$ by $f := \eta^{24}$. The holomorphicity and the transformation properties of η immediately imply that f is weakly modular of weight 12. Furthermore, $f = q + \mathcal{O}(q^2)$, so in fact f is a cusp form of weight 12. In Theorem 2.11, we will see that the $\mathbb{C}$-vector space of cusp forms of weight 12 is 1-dimensional. Since the Fourier coefficient of q of both Δ and η^{24} equals 1, we get

$$\Delta = \frac{(240E_4)^3 - (-504E_6)^2}{1728} = q\prod_{n=1}^{\infty}(1-q^n)^{24}.$$

The Fourier coefficients of this series are usually denoted by $\tau(n)$, so that (by definition)

$$\Delta = \sum_{n=1}^{\infty}\tau(n)q^n.$$

The function $n \mapsto \tau(n)$ is called *Ramanujan's τ-function.*

Remark. Ramanujan conjectured in 1916 some remarkable properties of τ, namely

- τ is multiplicative, i.e. $\tau(nm) = \tau(n)\tau(m)$ for all comprime $n, m \in \mathbb{Z}_{>0}$;
- $\tau(p^r) = \tau(p)\tau(p^{r-1}) - p^{11}\tau(p^{r-2})$ for all primes p and integers $r \geq 2$;
- $|\tau(p)| \leq 2p^{11/2}$ for all primes p.

The first two properties were already proven by Mordell in 1917 and the last by Deligne in 1974 as a consequence of his proof of the famous Weil conjectures. We will come back to the first two properties after we studied Hecke operators in Chapter 4.

2.7 The valence formula

We now come to a very important technical result about modular forms. To state and prove this result, we will use some definitions and results from complex analysis that are collected in §A.3.

Let f be meromorphic on $\mathbb{H}$ and weakly modular of weight k, let $z \in \mathbb{H}$, and let $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. It is not hard to check that the transformation formula $f|_k\gamma = f$ implies the equality

$$\mathrm{ord}_z\, f = \mathrm{ord}_{\gamma z}\, f,$$

so the order of f at z only depends on the $\mathrm{SL}_2(\mathbb{Z})$-orbit of z.

Recall that if f is meromorphic on $\mathbb{H}$, weakly modular of weight k and meromorphic at infinity, we constructed a meromorphic function $\tilde{f}$ on the open unit disc $\mathbb{D} = \{q \in \mathbb{C} \mid |q| < 1\}$. We define

$$\mathrm{ord}_{z=\infty}\, f = \mathrm{ord}_{q=0}\, \tilde{f}.$$

In particular, f is holomorphic at infinity (resp. vanishes at infinity) if and only if $\mathrm{ord}_\infty f \geq 0$ (resp. $\mathrm{ord}_\infty f > 0$).

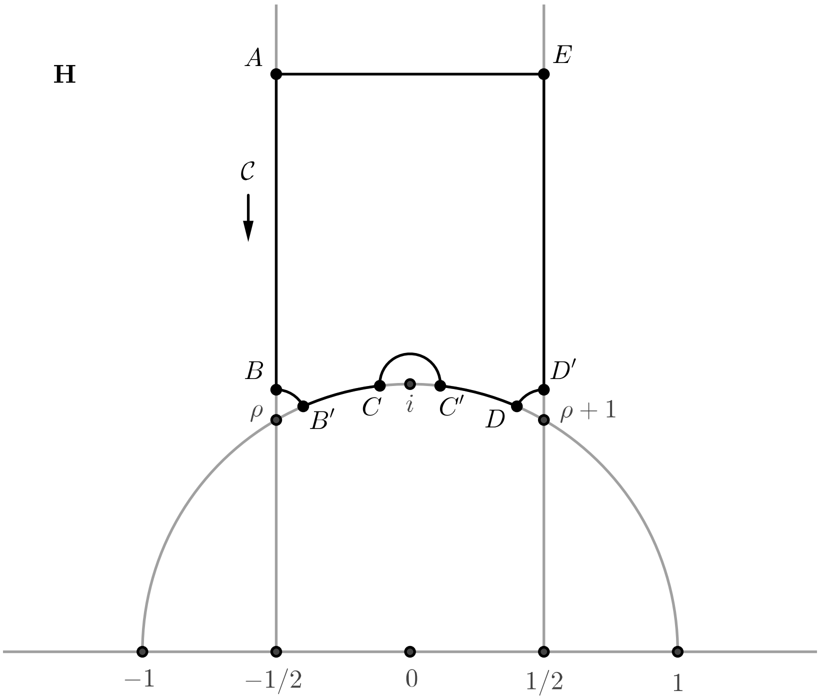
Theorem 2.8 (valence formula)**.** *Let f be a nonzero meromorphic function on $\mathbb{H}$ that is weakly modular of weight k (for the group $\mathrm{SL}_2(\mathbb{Z})$) and meromorphic at infinity. Then we have*

$$\mathrm{ord}_\infty\, f + \frac{1}{2}\,\mathrm{ord}_i\, f + \frac{1}{3}\,\mathrm{ord}_\rho\, f + \sum_{w\in W}\mathrm{ord}_w\, f = \frac{k}{12}.$$

Here W is the set $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ of $\mathrm{SL}_2(\mathbb{Z})$-orbits in $\mathbb{H}$, with the orbits of i and ρ omitted.

Proof. By the remark above, we may take all orbit representatives to lie in the fundamental domain $\mathcal{D}$. For simplicity of exposition, we assume that the boundary of $\mathcal{D}$ contains no zeroes or poles of f, except possibly at i, ρ and $\rho + 1$.

Let $\mathcal{C}$ be the contour in the following picture:



The small arcs around i, ρ, $\rho+1$ have radius r, and we will let r tend to 0. The segment AE has imaginary part R, and we will let R tend to ∞. In the case where the boundary of $\mathcal{D}$ does contain zeroes or poles of f, the contour $\mathcal{C}$ has to be modified using additional small arcs going around these zeroes or poles.

For R sufficiently large and r sufficiently small, the contour $\mathcal{C}$ contains all the zeroes and poles of f in $\mathcal{D}$ except those at i, ρ and $\rho+1$ (and infinity). Under this assumption, the argument principle (Theorem A.3) implies

$$\oint_{\mathcal{C}} \frac{f'(z)}{f(z)}dz = 2\pi i \sum_{w\in W} \operatorname{ord}_w f. \tag{2.12}$$

On the other hand, we can compute this integral by splitting up the contour $\mathcal{C}$ into eight parts, which we will consider separately.

First, we have

$$\begin{aligned}\int_{D'}^{E} \frac{f'}{f}(z)dz &= \int_B^A \frac{f'}{f}(z+1)dz \\ &= -\int_A^B \frac{f'}{f}(z)dz,\end{aligned}$$

so the integrals over the paths AB and $D'E$ cancel.

Second, from the equation

$$f(-1/z) = z^k f(z)$$

we obtain by differentiation

$$z^{-2}f'(-1/z) = kz^{k-1}f(z) + z^k f'(z)$$

and hence, dividing by the previous equation,

$$z^{-2}\frac{f'}{f}(-1/z) = \frac{k}{z} + \frac{f'}{f}(z).$$

We also note that

$$\frac{d}{dz}(-1/z) = z^{-2}.$$

Making the change of variables $z' = -1/z$, we therefore obtain

$$\begin{aligned}\int_{C'}^{D} \frac{f'}{f}(z)dz &= \int_{C}^{B'} \frac{f'}{f}(-1/z')(z')^{-2}dz' \\ &= \int_{C}^{B'} \left(\frac{k}{z'} + \frac{f'}{f}(z')\right)dz' \\ &= k\int_{C}^{B'} \frac{1}{z}dz - \int_{B'}^{C} \frac{f'}{f}(z)dz.\end{aligned}$$

This implies

$$\int_{B'}^{C} \frac{f'}{f}(z)dz + \int_{C'}^{D} \frac{f'}{f}(z)dz \longrightarrow k\frac{\pi i}{6} \quad \text{as } r \to 0,$$

since the angle $\angle C0B'$ tends to $\pi/6$ as $r \to 0$.

Third, as $r \to 0$, we have

$$\begin{aligned}\int_{B}^{B'} \frac{f'}{f}(z)dz &\longrightarrow -\frac{\pi i}{3}\operatorname{ord}_\rho(f), \\ \int_{C}^{C'} \frac{f'}{f}(z)dz &\longrightarrow -\pi i \operatorname{ord}_i(f), \\ \int_{D}^{D'} \frac{f'}{f}(z)dz &\longrightarrow -\frac{\pi i}{3}\operatorname{ord}_{\rho+1}(f) = -\frac{\pi i}{3}\operatorname{ord}_\rho(f).\end{aligned}$$

Fourth, to evaluate the integral from E to A, we make the change of variables $q = \exp(2\pi i z)$. By definition we have

$$f(z) = \tilde{f}(\exp(2\pi i z)),$$

and it follows that

$$f'(z) = 2\pi i \exp(2\pi i z)\tilde{f}'(\exp(2\pi i z)).$$

This implies

$$\frac{f'}{f}(z) = 2\pi i \exp(2\pi i z)\frac{\tilde{f}'}{\tilde{f}}(\exp(2\pi i z)).$$

Furthermore,

$$\frac{d}{dz}\exp(2\pi i z) = 2\pi i \exp(2\pi i z).$$

From this we obtain

$$\begin{aligned}\int_{E}^{A} \frac{f'}{f}(z)dz &= -\oint_{|q|=\exp(-2\pi R)} \frac{\tilde{f}'}{\tilde{f}}(q)dq \\ &= -2\pi i \operatorname{ord}_{q=0}\tilde{f} \\ &= -2\pi i \operatorname{ord}_{z=\infty} f.\end{aligned}$$

Summing the contributions of all the eight paths, we therefore obtain

$$\oint_{\mathcal{C}} \frac{f'}{f}(z)dz = k\frac{\pi i}{6} - \pi i \operatorname{ord}_i(f) - \frac{2\pi i}{3}\operatorname{ord}_\rho(f) - 2\pi i \operatorname{ord}_\infty(f).$$

Combining this with (2.12), we obtain

$$2\pi i \sum_{w\in W} \operatorname{ord}_w(f) = k\frac{\pi i}{6} - \pi i \operatorname{ord}_i(f) - \frac{2\pi i}{3}\operatorname{ord}_\rho(f) - 2\pi i \operatorname{ord}_\infty(f).$$

Rearranging this and dividing by $2\pi i$ yields the claim. □

2.8 Applications of the valence formula

We will now use Theorem 2.8 to prove a fundamental property of modular forms.

Notation. We write M_k for the $\mathbb{C}$-vector space of modular forms of weight k. We write $\mathrm{S}_k \subset \mathrm{M}_k$ for the subspace of M_k consisting of cusp forms of weight k.

Theorem 2.9. *1. The Eisenstein series E_4 has a simple zero at $z = \rho$ and no other zeroes.*

2. The Eisenstein series E_6 has a simple zero at $z = i$ and no other zeroes.

3. The modular form Δ of weight 12 has a simple zero at $z = \infty$ and no other zeroes.

Proof. If f is a modular form, the numbers $\operatorname{ord}_z f$ occurring in Theorem 2.8 are non-negative because f is holomorphic on $\mathbb{H}$ and at infinity. In the case $f = \Delta$, the q-expansion shows moreover that $\operatorname{ord}_\infty \Delta = 1$. One checks easily that the only way to get equality in Theorem 2.8 is if the location of the zeroes is as claimed. □

Corollary 2.10. *Multiplication by Δ is an isomorphism*

$$\begin{aligned} \mathrm{M}_k &\xrightarrow{\sim} \mathrm{S}_{k+12} \\ f &\longmapsto \Delta \cdot f. \end{aligned}$$

In particular, for all $k \in \mathbb{Z}$, we have

$$\dim \mathrm{S}_{k+12} = \dim \mathrm{M}_k.$$

Theorem 2.11. *The spaces M_k and S_k are finite-dimensional for every k. Furthermore, we have $\mathrm{M}_k = \{0\}$ if $k < 0$ or k is odd, and the dimensions of M_k for $k \geq 0$ even are given by*

$$\dim \mathrm{M}_k = \begin{cases} \lfloor k/12 \rfloor & \text{if } k \equiv 2 \pmod{12}, \\ \lfloor k/12 \rfloor + 1 & \text{if } k \not\equiv 2 \pmod{12}. \end{cases}$$

In particular, the dimensions of M_k and S_k for the first few values of k are given by

k	$\dim \mathrm{M}_k$	$\dim \mathrm{S}_k$
0	1	0
2	0	0
4	1	0
6	1	0
8	1	0
10	1	0
12	2	1
14	1	0
16	2	1

Proof. The fact that $\mathrm{M}_k = \{0\}$ for $k < 0$ follows from Theorem 2.8. The valence formula also easily implies $\mathrm{M}_0 = \mathbb{C}$ and $\mathrm{M}_2 = \{0\}$.

If k is odd and $f \in \mathrm{M}_k$, then applying the transformation formula

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z)$$

to the matrix $\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ implies that $f = 0$.

It remains to prove the theorem for even $k \geq 4$. In this case every modular form of weight k is a unique linear combination of E_k and a cusp form; this follows from the fact that E_k does not vanish at infinity. This gives a direct sum decomposition

$$\mathrm{M}_k = \mathrm{S}_k \oplus \mathbb{C} \cdot E_k \quad \text{for all even } k \geq 4.$$

In particular, this implies

$$\begin{aligned}\dim \mathrm{M}_k &= \dim \mathrm{S}_k + 1\\ &= \dim \mathrm{M}_{k-12} + 1.\end{aligned}$$

for all even $k \geq 4$. The theorem now follows by induction, starting from the known values of $\dim \mathrm{M}_k$ for $k \leq 2$. □

The following theorem is a very useful concrete consequence of the fact that spaces of modular forms are finite-dimensional.

Theorem 2.12. *Let f be a modular form of weight k with q-expansion $\sum_{n=0}^{\infty} a_n q^n$. Suppose that*

$$a_j = 0 \quad \text{for } j = 0, 1, \ldots, \lfloor k/12 \rfloor.$$

Then $f = 0$.

Proof. Suppose f is non-zero. Then the hypothesis implies that

$$\mathrm{ord}_\infty f \geq \lfloor k/12 \rfloor + 1 > k/12.$$

Therefore the left-hand side of the valence formula (Theorem 2.8) is strictly greater than $k/12$, contradiction. Hence $f = 0$. □

Corollary 2.13. *Let f, g be a modular form of the same weight k, with q-expansions $\sum_{n=0}^{\infty} a_n q^n$ and $\sum_{n=0}^{\infty} b_n q^n$, respectively. Suppose that*

$$a_j = b_j \quad \text{for } j = 0, 1, \ldots, \lfloor k/12 \rfloor.$$

Then $f = g$.

Theorem 2.12 is a very powerful tool. It allows us to conclude that two modular forms are identical if we only know a priori that their q-expansions agree to a certain finite precision. An example of a formula that can be proved using this principle is

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{j=1}^{n-1} \sigma_3(j)\sigma_3(n-j) \quad \text{for all } n \geq 1; \tag{2.13}$$

see Exercise 2.8. This identity is very hard to prove (or even conjecture) without using modular forms.

2.9 Exercises

Exercise 2.1. Prove that the formula (2.1) indeed defines a right action of $\mathrm{SL}_2(\mathbb{R})$ on the set of meromorphic functions on $\mathbb{H}$.

Exercise 2.2. We recall the notation

$$\sigma_t(n) = \sum_{d \mid n} d^t \quad \text{for all integers } t \geq 0 \text{ and } n \geq 1,$$

where d runs over the set of positive divisors of n.

(a) Let m, n and t be positive integers such that m and n are coprime. Show that

$$\sigma_t(mn) = \sigma_t(m)\sigma_t(n).$$

(b) Let n and t be positive integers, and let

$$n = \prod_{p \text{ prime}} p^{e_p} \quad (e_p \geq 0;\ e_p = 0 \text{ for all but finitely many } p)$$

be the prime factorisation of n. Show that

$$\sigma_t(n) = \prod_{p \text{ prime}} \frac{p^{(e_p+1)t} - 1}{p^t - 1}.$$

Exercise 2.3.

(a) Using the definition of the Bernoulli numbers B_k, prove the identity

$$\pi z \frac{\cos \pi z}{\sin \pi z} = \sum_{k \geq 0 \text{ even}} (2\pi i)^k \frac{B_k}{k!} z^k \quad \text{for all } |z| < 1.$$

(b) Using the formula (A.1), prove the identity

$$\pi z \frac{\cos \pi z}{\sin \pi z} = 1 - 2 \sum_{k \geq 2 \text{ even}} \zeta(k) z^k \quad \text{for all } |z| < 1.$$

(c) Deduce that the values of the Riemann zeta function at even integers $k \geq 2$ are given by

$$\zeta(k) = -\frac{(2\pi i)^k B_k}{2 \cdot k!}.$$

(d) Prove that B_k is non-zero if and only if $k = 1$ or k is even.

Exercise 2.4.

(a) Show that $G_4(\exp(2\pi i/3)) = 0$. (*Hint:* $G_4(-1/z) = z^4 G_4(z)$.)

(b) Show that $G_6(i) = 0$.

Exercise 2.5. Using the fact that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, prove that the transformation behaviour of the function E_2 under any element $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ is given by

$$(cz+d)^{-2} E_2\left(\frac{az+b}{cz+d}\right) = E_2(z) - \frac{1}{4\pi i} \frac{c}{cz+d}.$$

Exercise 2.6. Define $f\colon \mathbb{H} \to \mathbb{C}$ by

$$f(z) := G_2(z) - \frac{\pi}{\Im z}.$$

(a) Show that

$$f(\gamma z) = j(\gamma, z)^2 f(z) \quad \text{for all } \gamma \in \mathrm{SL}_2(\mathbb{Z}) \text{ and } z \in \mathbb{H}.$$

(b) Is f a modular form?

Exercise 2.7. Show that the q-expansion coefficients of the modular form Δ defined by (2.11) are integers, despite the division by 1728 occurring in the definition.

Exercise 2.8. Using the fact that the space M_8 is one-dimensional, prove the formula (2.13).

Exercise 2.9. We continue along the lines of the previous exercise.

(a) Prove the formula

$$\sigma_9(n) = \frac{21}{11}\sigma_5(n) - \frac{10}{11}\sigma_3(n) + \frac{5040}{11}\sum_{j=1}^{n-1}\sigma_3(j)\sigma_5(n-j) \quad \text{for all } n \in \mathbb{Z}_{>0}.$$

(b) Find similar expressions for σ_{13} in terms of σ_3 and σ_9, and in terms of σ_5 and σ_7.

(c) Express σ_{13} in terms of σ_3 and σ_5.

Exercise 2.10.

(a) Show that there exists $C \in \mathbb{R}_{>0}$ such that any element in $\mathbb{H}$ is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to some $z \in \mathbb{H}$ with $\Im(z) \geq C$. (You can take e.g. $C = \sqrt{3}/2$.)

(b) Deduce that if $f\colon \mathbb{H} \to \mathbb{C}$ is a modular form of weight 0, then $|f|$ attains a maximum on $\mathbb{H}$.

(c) Conclude that the space of modular forms of weight zero consists exactly of the constant functions $\mathbb{H} \to \mathbb{C}$. (*Hint:* use the maximum modulus principle.)

Exercise 2.11.

(a) Find rational numbers λ and μ such that

$$\Delta = \lambda E_4^3 + \mu E_{12}.$$

(b) Let $\tau(n)$ be the n-th coefficient in the q-expansion of Δ, so that

$$\Delta = \sum_{n=1}^{\infty} \tau(n) q^n.$$

Prove *Ramanujan's congruence*:

$$\tau(n) \equiv \sigma_{11}(n) \pmod{691}.$$

Exercise 2.12. Show that the ring $\mathbb{C}[E_2, E_4, E_6]$ is closed under differentiation.

Exercise 2.13.

(a) Show that the modular functions (for $\mathrm{SL}_2(\mathbb{Z})$) form a field F (with addition and multiplication defined pointwise).

(b) Prove that $F = \mathbb{C}(j)$ and that j is transcendental over $\mathbb{C}$.

Exercise 2.14. Consider the modular function $j\colon \mathbb{H} \to \mathbb{C}$.

(a) Show that $j(i) = 1728$ and $j(\rho) = 0$ (where $\rho = \exp(2\pi i/3)$).

(b) Let $z \in \mathcal{D}$ (the standard fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$). Prove:

$$(z \text{ lies on the boundary of } \mathcal{D} \text{ or } \Re z = 0) \;\Rightarrow\; j(z) \in \mathbb{R}.$$

(c) Show that $\bar{j}\colon \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H} \to \mathbb{C}$ given by $\bar{j}([z]) := j(z)$ is well-defined and prove that $\bar{j}$ is bijective.
(Here $[z]$ denotes the orbit of z under the action of $\mathrm{SL}_2(\mathbb{Z})$.)

(d) Prove the converse to part (b).

Exercise 2.15.

(a) Show that M_k is spanned by all $E_4^a E_6^b$ with $a, b \in \mathbb{Z}_{\geq 0}$ and $4a + 6b = k$.

(b) Show that E_4 and E_6 are algebraically independent over $\mathbb{C}$.

The above exercise shows that the *ring of modular forms* (for $\mathrm{SL}_2(\mathbb{Z})$) $M := \bigoplus_{k\in\mathbb{Z}} M_k$ is isomorphic to the two-variable polynomial ring $\mathbb{C}[x, y]$ via the isomorphism $\mathbb{C}[x, y] \xrightarrow{\sim} M$ given by $(x, y) \mapsto (E_4, E_6)$. (If we *grade* the rings by assigning grade k to a modular form of weight k and grades 4 and 6 to x and y respectively, we get an isomorphism of graded rings.)

Chapter 3

Modular forms for congruence subgroups

3.1 Congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$

So far, we have considered functions satisfying a suitable transformation property with respect to the action of the full group $\mathrm{SL}_2(\mathbb{Z})$. It turns out to be very useful to also consider functions having this transformation behaviour only with respect to certain subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Definition. Let N be a positive integer. The *principal congruence subgroup of level* N is the group

$$\Gamma(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; \gamma \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \pmod N \right\}.$$

In other words, $\Gamma(N)$ is the kernel of the reduction map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. The reduction map is surjective; see Exercise 3.1. We therefore get an isomorphism

$$\mathrm{SL}_2(\mathbb{Z})/\Gamma(N) \xrightarrow{\sim} \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

In particular, this implies that $\Gamma(N)$ is a normal subgroup of finite index in $\mathrm{SL}_2(\mathbb{Z})$, namely

$$(\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)) = \#\,\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

Definition. A *congruence subgroup* (of $\mathrm{SL}_2(\mathbb{Z})$) is a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ containing $\Gamma(N)$ for some $N \ge 1$. The least such N is called the *level* of Γ.

We note that every congruence subgroup has finite index in $\mathrm{SL}_2(\mathbb{Z})$. The converse is false; there exist subgroups of finite index in $\mathrm{SL}_2(\mathbb{Z})$ that do not contain $\Gamma(N)$ for any N.

Examples. The most important examples of congruence subgroups are the groups

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; \begin{array}{r} a \equiv d \equiv 1 \pmod N, \\ c \equiv 0 \pmod N \end{array} \right\}$$

and

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; c \equiv 0 \pmod N \right\}.$$

We have a chain of inclusions

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z}).$$

These inclusions are in general strict; however, all of them are equalities for $N = 1$, and we have $\Gamma_0(2) = \Gamma_1(2)$.

Proposition 3.1. *The congruence subgroup $\Gamma_1(N)$ is normal in $\Gamma_0(N)$, and there is an isomorphism*

$$\begin{aligned}\Gamma_0(N)/\Gamma_1(N) &\xrightarrow{\sim} (\mathbb{Z}/N\mathbb{Z})^\times\\ \left[\begin{pmatrix} a & b\\ c & d\end{pmatrix}\right] &\longmapsto d \bmod N.\end{aligned}$$

For the proof, see Exercise 3.2.

The groups introduced above are the most important examples of congruence subgroups (although they are certainly not the only ones). It turns out that $\Gamma_0(N)$ and $\Gamma_1(N)$ have a useful "moduli interpretation".

To show how this works for the group $\Gamma_0(N)$, we consider pairs (L, G) with $L \subset \mathbb{C}$ a lattice and G a cyclic subgroup of order N of the quotient $\mathbb{C}/L$. To these data we attach another lattice L', namely the inverse image of G in $\mathbb{C}$ with respect to the quotient map $\mathbb{C} \to \mathbb{C}/L$. Then we can choose a basis (ω_1, ω_2) for L with the property that L' equals $\mathbb{Z}\omega_1 + \frac{1}{N}\mathbb{Z}\omega_2$. For any $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, the basis $(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$ of L again has the property above if and only if c is divisible by N, i.e. if and only if $\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)$ is in $\Gamma_0(N)$. Restricting to bases (ω_1, ω_2) with $\omega_1/\omega_2 \in \mathbb{H}$ and taking the quotient by the action of the subgroup $\Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$, we obtain a bijection between the set of homothety classes of pairs (L, G) as above and the quotient set $\Gamma_0(N)\backslash\mathbb{H}$.

An analogous argument shows that there is a bijection between the set of homothety classes of pairs (L, P), where $L \subset \mathbb{C}$ is a lattice and P is a point of order N in the group $\mathbb{C}/L$, and the set $\Gamma_1(N)\backslash\mathbb{H}$. We refer to Exercise 3.7 for details.

Definition. Let f be a meromorphic function on $\mathbb{H}$, let k be an integer, and let Γ be a congruence subgroup. We say that f is *weakly modular of weight k for the group* Γ (or *of level* Γ) if it satisfies the transformation formula

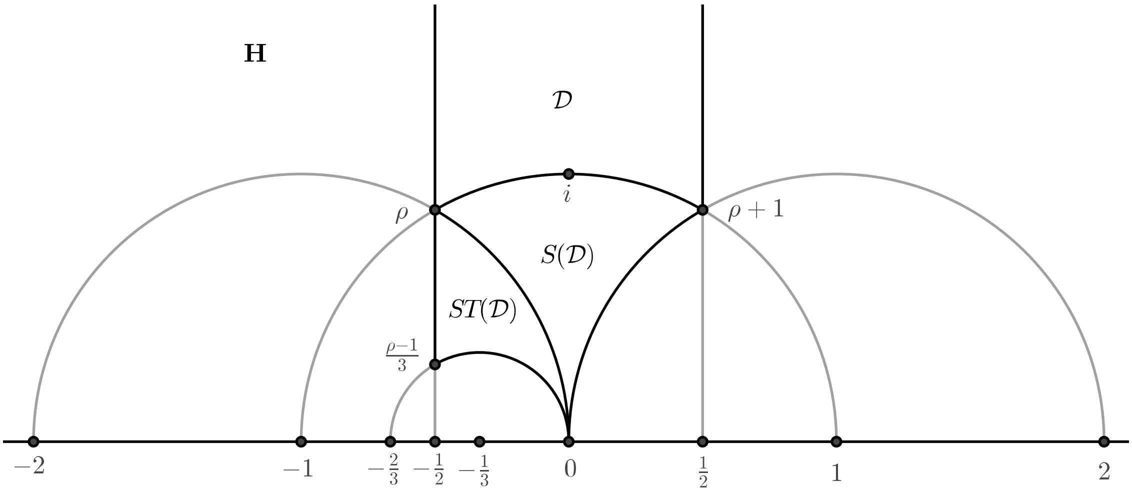
$$f|_k\gamma = f \quad \text{for all } \gamma \in \Gamma.$$

To generalise the definition of modular forms to this setting, we will have to answer the question how to generalise the notion of being holomorphic at infinity.

Example. Take $\Gamma = \Gamma_0(2) = \Gamma_1(2)$. A system of coset representatives for the quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$ is

$$\left\{\begin{pmatrix} 1 & 0\\ 0 & 1\end{pmatrix}, \begin{pmatrix} 0 & -1\\ 1 & 0\end{pmatrix}, \begin{pmatrix} 0 & -1\\ 1 & 1\end{pmatrix}\right\} = \{1, S, ST\}.$$

(It is important to take this quotient instead of $\mathrm{SL}_2(\mathbb{Z})/\Gamma$.) Using this, one can draw the following picture of a fundamental domain for Γ:



There are now two points "at infinity" that are in the closure of $\mathcal{D}$ in the Riemann sphere, but not in $\mathbb{H}$, namely ∞ and 0.

3.2 Fundamental domains and cusps

Proposition 3.2. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and let R be a set of coset representatives for the quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$. Then the set*

$$\mathcal{D}_\Gamma = \bigcup_{\gamma\in R} \gamma\mathcal{D}$$

has the property that for any $z \in \mathbb{H}$ there exists $\gamma \in \Gamma$ such that $\gamma z \in \mathcal{D}_\Gamma$. Furthermore, this γ is unique up to multiplication by an element of $\Gamma \cap \{\pm 1\}$, except possibly if γz lies on the boundary of $\mathcal{D}_\Gamma$.

Proof. Let $z \in \mathbb{H}$. By Theorem 1.5, there exist $z_0 \in \mathcal{D}$ and $\gamma_0 \in \mathrm{SL}_2(\mathbb{Z})$ such that $z = \gamma_0 z_0$. Since R is a set of coset representatives, we can express γ_0 uniquely as $\gamma_0 = \gamma^{-1}\gamma'$ with $\gamma \in \Gamma$ and $\gamma' \in R$. We now have

$$\gamma z = \gamma\gamma_0 z_0 = \gamma' z_0 \in \mathcal{D}_\Gamma.$$

For the statement about uniqueness of γ, see Exercise 3.4. □

Definition. The *projective line over* $\mathbb{Q}$ is the set

$$\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}.$$

The group $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathbb{P}^1(\mathbb{Q})$ by the same formula giving the action on $\mathbb{H}$:

$$\gamma t = \frac{at+b}{ct+d} \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}), t \in \mathbb{P}^1(\mathbb{Q}).$$

Here the right-hand side is to be interpreted as a/c if $t = \infty$, and as ∞ if $ct + d = 0$.

Lemma 3.3. *The action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{P}^1(\mathbb{Q})$ is transitive.*

Proof. It suffices to show that for every $t \in \mathbb{Q}$, there exists $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma\infty = t$. We write $t = a/c$ with a, c coprime integers. Then there exist integers r, s such that $ar + cs = 1$; the matrix $\gamma = \left(\begin{smallmatrix} a & -s \\ c & r \end{smallmatrix}\right)$ has the required property. □

One easily checks that the stabiliser of ∞ in $\mathrm{SL}_2(\mathbb{Z})$ is

$$\mathrm{SL}_2(\mathbb{Z})_\infty = \left\{ \pm\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \,\middle|\, b \in \mathbb{Z} \right\}.$$

This shows that we have a bijection

$$\begin{aligned} \mathrm{SL}_2(\mathbb{Z})/\mathrm{SL}_2(\mathbb{Z})_\infty &\xrightarrow{\sim} \mathbb{P}^1(\mathbb{Q}) \\ \gamma\,\mathrm{SL}_2(\mathbb{Z})_\infty &\longmapsto \gamma\infty. \end{aligned}$$

Definition. Let Γ be a congruence subgroup. The set of *cusps* of Γ is the set of Γ-orbits in $\mathbb{P}^1(\mathbb{Q})$, i.e. the quotient

$$\mathrm{Cusps}(\Gamma) = \Gamma\backslash\mathbb{P}^1(\mathbb{Q}).$$

Note that by what we have just seen, an equivalent definition is

$$\mathrm{Cusps}(\Gamma) = \Gamma\backslash\mathrm{SL}_2(\mathbb{Z})/\mathrm{SL}_2(\mathbb{Z})_\infty.$$

In particular, we have a surjective map

$$\Gamma\backslash\mathrm{SL}_2(\mathbb{Z}) \twoheadrightarrow \mathrm{Cusps}(\Gamma).$$

Let $\mathfrak{c}$ be a cusp of Γ, and let t be an element of the corresponding Γ-orbit in $\mathbb{P}^1(\mathbb{Q})$. We denote by Γ_t the stabiliser of t in Γ, i.e.

$$\Gamma_t = \{\gamma \in \Gamma \mid \gamma t = t\}.$$

By the lemma, we can choose a matrix $\gamma_t \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma_t\infty = t$. For every $\gamma \in \Gamma$, we now have the equivalences

$$\begin{aligned} \gamma \in \Gamma_t &\iff \gamma t = t \\ &\iff \gamma\gamma_t\infty = \gamma_t\infty \\ &\iff \gamma_t^{-1}\gamma\gamma_t\infty = \infty \\ &\iff \gamma_t^{-1}\gamma\gamma_t \in \mathrm{SL}_2(\mathbb{Z})_\infty. \end{aligned}$$

This shows that

$$\Gamma_t = \Gamma \cap \gamma_t \,\mathrm{SL}_2(\mathbb{Z})_\infty \gamma_t^{-1}.$$

In particular, we have an injective map

$$\Gamma_t\backslash(\gamma_t \,\mathrm{SL}_2(\mathbb{Z})_\infty\gamma_t^{-1}) \rightarrowtail \Gamma\backslash\mathrm{SL}_2(\mathbb{Z}).$$

This implies that Γ_t is of finite index in $\gamma_t \,\mathrm{SL}_2(\mathbb{Z})\gamma_t^{-1}$. It is useful to conjugate by γ_t and define

$$H_\mathfrak{c} = \gamma_t^{-1}\Gamma\gamma_t \cap \mathrm{SL}_2(\mathbb{Z})_\infty. \tag{3.1}$$

Hence $H_\mathfrak{c}$ is a subgroup of finite index in $\mathrm{SL}_2(\mathbb{Z})_\infty$. It is independent of the choice of t and γ_t; see Exercise 3.5.

Lemma 3.4. *Let H be a subgroup of finite index in $\mathrm{SL}_2(\mathbb{Z})_\infty$. Then H is one of the following:*

1. *infinite cyclic generated by $\left(\begin{smallmatrix}1 & h\\ 0 & 1\end{smallmatrix}\right)$ with $h \ge 1$;*
2. *infinite cyclic generated by $\left(\begin{smallmatrix}-1 & h\\ 0 & -1\end{smallmatrix}\right)$ with $h \ge 1$;*
3. *isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}$, generated by $\left(\begin{smallmatrix}-1 & 0\\ 0 & -1\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1 & h\\ 0 & 1\end{smallmatrix}\right)$ with $h \ge 1$.*

Furthermore, h is the index of $\{\pm 1\}H$ in $\mathrm{SL}_2(\mathbb{Z})_\infty$.

We refer to Exercise 3.6 for the proof.

Definition. Let $\mathfrak{c} \in \mathrm{Cusps}(\Gamma)$, and let t be an element of the corresponding Γ-orbit in $\mathbb{P}^1(\mathbb{Q})$. The *width* of $\mathfrak{c}$, denoted by $h_\Gamma(\mathfrak{c})$, is the integer h defined as in Lemma 3.4 (with $H = H_\mathfrak{c}$), i.e. the index of $\{\pm 1\}H_\mathfrak{c}$ in $\mathrm{SL}_2(\mathbb{Z})_\infty$. Furthermore, the cusp $\mathfrak{c}$ is called *irregular* if $H_\mathfrak{c}$ is of the form (2) in Lemma 3.4, *regular* otherwise.

Remark. Suppose Γ is a *normal* congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. By definition, this means that $\gamma^{-1}\Gamma\gamma = \Gamma$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. From (3.1) it then follows that all the groups $H_\mathfrak{c}$ for $\mathfrak{c} \in \mathrm{Cusps}(\Gamma)$ are equal. In particular, all cusps of Γ have the same width, and either all are regular or all are irregular.

Before continuing, we state and prove a group-theoretic lemma.

Lemma 3.5. *Let G be a group acting transitively on a set X, and let H be a subgroup of finite index in G.*

1. *For any $x \in X$, the stabiliser $\mathrm{Stab}_H\, x$ has finite index in $\mathrm{Stab}_G\, x$, and we have an injection*

$$(\mathrm{Stab}_H\, x)\backslash(\mathrm{Stab}_G\, x) \rightarrowtail H\backslash G$$

with image $H\backslash H\,\mathrm{Stab}_G\, x$.

2. *Let $x_0 \in X$. There is a surjective map*

$$\begin{aligned} H\backslash G &\twoheadrightarrow H\backslash X \\ Hg &\mapsto Hgx_0, \end{aligned}$$

and for every $x \in X$, the cardinality of the fibre of this map over Hx equals $(\operatorname{Stab}_G x : \operatorname{Stab}_H x)$*.*

3. *If R is a set of orbit representatives for the quotient $H\backslash X$, we have*

$$\sum_{x \in R} (\operatorname{Stab}_G x : \operatorname{Stab}_H x) = (G : H).$$

Proof. Part (1) is standard and just recalled here.

As for part (2), the transitivity of the G-action on X implies that for every $x \in X$ we can choose an element $g_x \in G$ such that $g_x x_0 = x$. This implies the surjectivity of the map $H\backslash G \to H\backslash X$. Let T_{Hx} denote the fibre of this map over Hx, so that by definition

$$T_{Hx} = \bigl\{Hg \in H\backslash G \mid Hgx_0 = Hx\bigr\}.$$

Replacing Hg by $Hg'g_x$, we obtain a bijection

$$\begin{aligned} T_{Hx} &\cong \bigl\{Hg' \in H\backslash G \mid Hg'g_x x_0 = Hx\bigr\} \\ &= \bigl\{Hg' \in H\backslash G \mid Hg'x = Hx\bigr\} \\ &= H\backslash(H \operatorname{Stab}_G x) \\ &\cong (\operatorname{Stab}_H x)\backslash(\operatorname{Stab}_G x), \end{aligned}$$

where in the last step we have used part (1). Taking cardinalities, we obtain the claim.

Finally, summing over a system of representatives R for the quotient $H\backslash X$, we obtain

$$\begin{aligned} (G : H) &= \#(H\backslash G) \\ &= \sum_{x \in R} \#T_{Hx} \\ &= \sum_{x \in R} (\operatorname{Stab}_G x : \operatorname{Stab}_H x). \end{aligned}$$

This proves part (3). □

Corollary 3.6. *Let Γ be a congruence subgroup, and let $\bar\Gamma$ be the image of Γ in $\mathrm{PSL}_2(\mathbb{Z})$. Then we have*

$$\begin{aligned} \sum_{\mathfrak{c} \in \mathrm{Cusps}(\Gamma)} h_\Gamma(\mathfrak{c}) &= (\mathrm{PSL}_2(\mathbb{Z}) : \bar\Gamma) \\ &= (\mathrm{SL}_2(\mathbb{Z}) : \{\pm 1\}\Gamma). \end{aligned}$$

Proof. Apply part (3) of the lemma to $G = \mathrm{PSL}_2(\mathbb{Z})$, $H = \bar\Gamma$ and $X = \mathbb{P}^1(\mathbb{Q})$. □

Example. Let p be a prime number. We consider the group $\Gamma = \Gamma_0(p)$. We note that $\Gamma_0(p)$ contains the principal congruence subgroup $\Gamma(p)$, and there is an isomorphism

$$\Gamma_0(p)\backslash\mathrm{SL}_2(\mathbb{Z}) \xrightarrow{\sim} K_p\backslash\mathrm{SL}_2(\mathbb{F}_p)$$

where

$$\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$$

and

$$\begin{aligned} K_p &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{F}_p) \;\middle|\; c = 0 \right\} \\ &= \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \;\middle|\; a \in \mathbb{F}_p^\times, b \in \mathbb{F}_p \right\}. \end{aligned}$$

It is known that

$$\#\,\mathrm{SL}_2(\mathbb{F}_p) = p(p-1)(p+1).$$

Furthermore, the description above of K_p implies

$$\#K_p = p(p-1).$$

We therefore obtain

$$\begin{aligned} (\mathrm{SL}_2(\mathbb{Z}) : \Gamma) &= (\mathrm{SL}_2(\mathbb{F}_p) : K_p) \\ &= \frac{\#\,\mathrm{SL}_2(\mathbb{F}_p)}{\#K_p} \\ &= \frac{p(p-1)(p+1)}{p(p-1)} \\ &= p+1. \end{aligned}$$

(Another way of computing this is to find a transitive action of $\mathrm{SL}_2(\mathbb{F}_p)$ on $\mathbb{P}^1(\mathbb{F}_p)$ such that some point of $\mathbb{P}^1(\mathbb{F}_p)$ has stabiliser K_p.)

To compute the set of cusps of Γ, we determine the Γ-orbits in $\mathbb{P}^1(\mathbb{Q})$. The orbit of $\infty \in \mathbb{P}^1(\mathbb{Q})$ is

$$\begin{aligned} \Gamma \cdot \infty &= \left\{ \begin{pmatrix} a & b \\ cp & d \end{pmatrix} \infty \;\middle|\; a, b, c, d \in \mathbb{Z}, ad - bcp = 1 \right\} \\ &= \left\{ \frac{a}{cp} \;\middle|\; a, c \in \mathbb{Z}, \gcd(a, cp) = 1 \right\} \\ &= \left\{ \frac{r}{s} \;\middle|\; r, s \in \mathbb{Z}, \gcd(r, s) = 1, p \mid s \right\}. \end{aligned}$$

(Here a fraction with denominator 0 is interpreted as ∞.) Likewise, the orbit of $0 \in \mathbb{P}^1(\mathbb{Q})$ is

$$\begin{aligned} \Gamma \cdot 0 &= \left\{ \begin{pmatrix} a & b \\ cp & d \end{pmatrix} 0 \;\middle|\; a, b, c, d \in \mathbb{Z}, ad - bcp = 1 \right\} \\ &= \left\{ \frac{b}{d} \;\middle|\; b, d \in \mathbb{Z}, \gcd(b, d) = 1, p \nmid d \right\}. \end{aligned}$$

From this description of the two orbits it is clear that every element of $\mathbb{P}^1(\mathbb{Q})$ is in exactly one of them. In particular, $\Gamma_0(p)$ has two cusps, namely the two elements $[\infty]$ and $[0]$ of $\Gamma_0(p)\backslash\mathbb{P}^1(\mathbb{Q})$.

Next, we determine the widths of these two cusps. For the cusp $\mathfrak{c} = [\infty]$, we take $t = \infty$ and $\gamma_t = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. This gives $H_\mathfrak{c} = \mathrm{SL}_2(\mathbb{Z})_\infty$ and $h_\Gamma(\mathfrak{c}) = 1$. For the cusp $\mathfrak{c} = [0]$, we take $t = 0$ and $\gamma_t = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. We have

$$\Gamma_t = \left\{ \pm\begin{pmatrix} 1 & 0 \\ cp & 1 \end{pmatrix} \;\middle|\; c \in \mathbb{Z} \right\}.$$

An easy calculation implies

$$H_\mathfrak{c} = \left\{ \pm\begin{pmatrix} 1 & cp \\ 0 & 1 \end{pmatrix} \;\middle|\; c \in \mathbb{Z} \right\}.$$

In particular, $h_\Gamma(\mathfrak{c}) = p$.

3.3 Modular forms for congruence subgroups

Let Γ be a congruence subgroup, let k be an integer, and let f be a meromorphic function on $\mathbb{H}$ that is weakly modular of weight k for the group Γ. Let $\mathfrak{c}$ be a cusp of Γ, and let $t \in \mathbb{P}^1(\mathbb{Q})$ be an element of the corresponding Γ-orbit in $\mathbb{P}^1(\mathbb{Q})$. We choose $\gamma_t \in \mathrm{SL}_2(\mathbb{Z})$ such that $\gamma_t\infty = t \in \mathbb{P}^1(\mathbb{Q})$. Then the meromorphic function $f|_k\gamma_t$ is invariant under the weight k action of the group $H_\mathfrak{c}$. By the definition of the width $h_\Gamma(\mathfrak{c})$ and of (ir)regularity of the cusp $\mathfrak{c}$, the group $H_\mathfrak{c}$ contains the element $\left(\begin{smallmatrix}1 & \tilde h_\Gamma(\mathfrak{c})\\ 0 & 1\end{smallmatrix}\right)$, where

$$\tilde h_\Gamma(\mathfrak{c}) = \begin{cases} h_\Gamma(\mathfrak{c}) & \text{if the cusp } \mathfrak{c} \text{ is regular,}\\ 2h_\Gamma(\mathfrak{c}) & \text{if the cusp } \mathfrak{c} \text{ is irregular.}\end{cases}$$

This means that the function $f|_k\gamma_t$ satisfies

$$(f|_k\gamma_t)(z + \tilde h_\Gamma(\mathfrak{c})) = (f|_k\gamma_t)(z).$$

On the punctured unit disc $\mathbb{D}^*$, we can therefore express $f|_k\gamma_t$ as a function of the variable

$$q_\mathfrak{c} = \exp(2\pi i z/\tilde h_\Gamma(\mathfrak{c})).$$

In other words, there exists a function $\tilde f_\mathfrak{c}\colon \mathbb{D}^* \to \mathbb{C}\cup\{\infty\}$ such that

$$(f|_k\gamma_t)(z) = \tilde f_\mathfrak{c}(\exp(2\pi i z/\tilde h_\Gamma(\mathfrak{c}))).$$

We say that f is *meromorphic at the cusp* $\mathfrak{c}$ if $\tilde f_\mathfrak{c}$ can be continued to a meromorphic function on $\mathbb{D}$. In this case, we can write $\tilde f_\mathfrak{c}$ as a Laurent series

$$\tilde f_\mathfrak{c}(q_\mathfrak{c}) = \sum_{n\in\mathbb{Z}} a_{\mathfrak{c},n} q_\mathfrak{c}^n,$$

where $a_{\mathfrak{c},n} = 0$ for $n \ll 0$. Furthermore, we say that f is *holomorphic at* $\mathfrak{c}$ if in addition $\tilde f_\mathfrak{c}$ is holomorphic at $q_\mathfrak{c} = 0$, and that f *vanishes at* $\mathfrak{c}$ if $\tilde f_\mathfrak{c}$ vanishes at $q_\mathfrak{c} = 0$. Finally, if f is not identically zero and is meromorphic at $\mathfrak{c}$, we define the *order* of f at $\mathfrak{c}$ as the least n such that $a_{\mathfrak{c},n} \ne 0$. The notation for this order is $\mathrm{ord}_{\Gamma,\mathfrak{c}}(f)$.

Definition. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and let k be an integer. A *modular form* of weight k for the group Γ is a holomorphic function $f\colon \mathbb{H} \to \mathbb{C}$ that is weakly modular of weight k for Γ and holomorphic at all cusps of Γ. Such an f is called a *cusp form* (of weight k for the group Γ) if it vanishes at all cusps of Γ.

As in the case of modular forms for $\mathrm{SL}_2(\mathbb{Z})$, it is straightforward to check that the set of modular forms of weight k for Γ is a $\mathbb{C}$-vector space.

Notation. We write $\mathrm{M}_k(\Gamma)$ for the $\mathbb{C}$-vector space of modular forms of weight k for the group Γ, and $\mathrm{S}_k(\Gamma)$ for the subspace of cusp forms.

For proving that a holomorphic function that is weakly modular is actually modular, checking directly the condition that it is holomorphic at all cusps might be a bit complicated in practice. The theorem below can be used to translate this into checking that it is holomorphic at infinity and that the Fourier coefficients do not grow too quickly. The converse also holds.

Theorem 3.7. *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and let k be an integer. Let $f\colon \mathbb{H} \to \mathbb{C}$ be a holomorphic function which is weakly modular of weight k for Γ. Then the following two properties are equivalent:*

(i) f is holomorphic at all cusps;

(ii) f is holomorphic at infinity and there exist $C, d \in \mathbb{R}_{>0}$ such that for the Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} a_n q_\infty^n$$

we have

$$|a_n| \le Cn^d \quad \textit{for all } n \in \mathbb{Z}_{>0}.$$

Proof. '(ii) ⇒ (i)': See Exercise 3.19.
'(i) ⇒ (ii)': This will be discussed in Chapter 6. (We note that this implication will never be used in these notes.) □

3.4 Example: the θ-function

Definition. The *Jacobi theta function* is the holomorphic function $\theta\colon \mathbb{H} \to \mathbb{C}$ defined by

$$\theta(z) = \sum_{n\in\mathbb{Z}} q^{n^2} = 1 + 2\sum_{n=1}^{\infty} q^{n^2} \quad (q = \exp(2\pi i z)).$$

Note that uniform convergence of the series on compact sets follows immediately by comparing it with the geometric series, from which the holomorphicity follows. Obviously, θ satisfies

$$\theta(z+1) = \theta(z) \quad \text{for all } z \in \mathbb{H}. \tag{3.2}$$

There is yet another type of transformation satisfied by θ.

Theorem 3.8. *The function θ satisfies the transformation formula*

$$\theta\left(\frac{-1}{4z}\right) = \sqrt{-2iz}\,\theta(z) \qquad \textit{for all } z \in \mathbb{H} \tag{3.3}$$

where the branch of $\sqrt{-2iz}$ is taken to have positive real part.

Proof. Since both sides are holomorphic functions on $\mathbb{H}$, it suffices to prove the identity for z on the imaginary axis. (Namely, the difference between the left-hand side and the right-hand side will then be zero on a subset of $\mathbb{H}$ possessing a limit point in $\mathbb{H}$, which implies that it is identically zero.)

Let us write $z = ia/2$ with $a > 0$. From Theorem A.6 and Corollary A.8, we obtain

$$\sum_{m\in\mathbb{Z}} \exp(-\pi a m^2) = \frac{1}{\sqrt{a}} \sum_{n\in\mathbb{Z}} \exp(-\pi n^2/a).$$

Substituting $a = -2iz$ gives

$$\sum_{m\in\mathbb{Z}} \exp(2\pi i m^2 z) = \frac{1}{\sqrt{-2iz}} \sum_{n\in\mathbb{Z}} \exp(-2\pi i n^2/(4z)).$$

This implies the claim. □

Corollary 3.9. *The function θ satisfies the transformation formula*

$$\theta\left(\frac{z}{4z+1}\right) = \sqrt{4z+1}\,\theta(z) \qquad \textit{for all } z \in \mathbb{H} \tag{3.4}$$

where the branch of $\sqrt{4z+1}$ is taken to have positive real part.

Proof. Let $z' := -1/(4z) - 1 \in \mathbb{H}$ and note that

$$\frac{z}{4z+1} = -\frac{1}{4z'}.$$

Now apply (3.3) with z' instead of z, followed by (3.2), and finally apply (3.3) (again). □

Theorem 3.10. *Let k be an even positive integer. Then the function*

$$\theta^k \colon z \mapsto \theta(z)^k$$

is a modular form of weight $k/2$ for the group $\Gamma_1(4)$.

Proof. First note that it suffices to prove that $f := \theta^2 \in \mathrm{M}_1(\Gamma_1(4))$. Let $T := \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ as usual and let $A := \left(\begin{smallmatrix} 1 & 0 \\ 4 & 1 \end{smallmatrix}\right)$. From (3.2) and (3.4) we get respectively

$$f|_1 T = f \quad \text{and} \quad f|_1 A = f.$$

According to Exercise 3.13, the group generated by A and T equals $\Gamma_1(4)$. We arrive at the fact that f is holomorphic and weakly modular of weight 1 for the group $\Gamma_1(4)$. By construction, f is holomorphic at infinity. By Theorem 3.7 it remains to show that the absolute values of the Fourier coefficients of f are bounded by a polynomial in the index. This is left as an (easy) exercise. □

3.5 Eisenstein series of weight 2

The space of modular forms of weight 2 is trivial, and the “Eisenstein series” E_2 is not a modular form. However, we can use E_2 to define modular forms of weight 2 for congruence subgroups of higher level as follows. For every positive integer e, we define a holomorphic function $E_2^{(e)} \colon \mathbb{H} \to \mathbb{C}$ by

$$E_2^{(e)}(z) = E_2(z) - eE_2(ez).$$

By Exercise 2.5, for any element $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(e)$ we have

$$\begin{aligned}
(cz+d)^{-2}E_2^{(e)}\left(\frac{az+b}{cz+d}\right) &= (cz+d)^{-2}E_2\left(\frac{az+b}{cz+d}\right) - e(cz+d)^{-2}E_2\left(e\frac{az+b}{cz+d}\right) \\
&= (cz+d)^{-2}E_2\left(\frac{az+b}{cz+d}\right) - e((c/e)(ez)+d)^{-2}E_2\left(\frac{a(ez)+be}{(c/e)(ez)+d}\right) \\
&= E_2(z) - \frac{1}{4\pi i}\frac{c}{cz+d} - e\left(E_2(ez) - \frac{1}{4\pi i}\frac{c/e}{(c/e)(ez)+d}\right) \\
&= E_2(z) - eE_2(Ez) \\
&= E_2^{(e)}(z).
\end{aligned}$$

This shows that the function $E_2^{(e)}$ is weakly modular of weight 2 for $\Gamma_0(e)$. It then follows from Theorem 3.7 that $E_2^{(e)}$ is holomorphic at the cusps and hence is a modular form for $\Gamma_0(e)$.

3.6 The valence formula for congruence subgroups

We now generalise Theorem 2.8 to arbitrary congruence subgroups.

Notation. For any congruence subgroup Γ, we will write $\bar{\Gamma}$ for the image of Γ under the natural quotient map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{PSL}_2(\mathbb{Z})$. We will also write

$$\Gamma_z = \mathrm{Stab}_\Gamma z \quad \text{and} \quad \bar{\Gamma}_z = \mathrm{Stab}_{\bar{\Gamma}} z \quad \text{for all } z \in \mathbb{H}.$$

Theorem 3.11 (valence formula for congruence subgroups)**.** *Let Γ be a congruence subgroup, and let k be an integer. Let f be a non-zero meromorphic function on $\mathbb{H}$ that is weakly modular of weight k for the group Γ and meromorphic at all cusps of Γ. Let*

$$\epsilon_{\Gamma,\mathfrak{c}} = \begin{cases} 1 & \textit{if } -1 \notin \Gamma \textit{ and } \mathfrak{c} \textit{ is regular,} \\ 2 & \textit{if } -1 \in \Gamma \textit{ or } \mathfrak{c} \textit{ is irregular,} \end{cases}$$

and

$$\bar\epsilon_{\Gamma,\mathfrak{c}} = \begin{cases} 1 & \textit{if } \mathfrak{c} \textit{ is regular,} \\ 2 & \textit{if } \mathfrak{c} \textit{ is irregular.} \end{cases}$$

Then we have

$$\sum_{z\in\Gamma\backslash\mathbb{H}} \frac{\operatorname{ord}_z(f)}{\#\Gamma_z} + \sum_{\mathfrak{c}\in\mathrm{Cusps}(\Gamma)} \frac{\operatorname{ord}_{\Gamma,\mathfrak{c}}(f)}{\epsilon_{\Gamma,\mathfrak{c}}} = \frac{k}{24}(\mathrm{SL}_2(\mathbb{Z}) : \Gamma).$$

and

$$\sum_{z\in\Gamma\backslash\mathbb{H}} \frac{\operatorname{ord}_z(f)}{\#\bar\Gamma_z} + \sum_{\mathfrak{c}\in\mathrm{Cusps}(\Gamma)} \frac{\operatorname{ord}_{\Gamma,\mathfrak{c}}(f)}{\bar\epsilon_{\Gamma,\mathfrak{c}}} = \frac{k}{12}(\mathrm{PSL}_2(\mathbb{Z}) : \bar\Gamma).$$

Proof. The proof is based on Theorem 2.8 and Lemma 3.5. Let us write

$$d = (\mathrm{SL}_2(\mathbb{Z}) : \Gamma).$$

Let R be a system of coset representatives for the quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$; then we have $\#R = d$. We now define

$$F(z) = \prod_{\gamma\in R} (f|_k\gamma)(z).$$

This function is weakly modular of weight dk for the full modular group $\mathrm{SL}_2(\mathbb{Z})$ and meromorphic at ∞. By the valence formula for $\mathrm{SL}_2(\mathbb{Z})$ (Theorem 2.8), we therefore have

$$\operatorname{ord}_\infty F + \frac{1}{2}\operatorname{ord}_i F + \frac{1}{3}\operatorname{ord}_\rho F + \sum_{w\in W} \operatorname{ord}_w F = \frac{dk}{12}.$$

were W is the set $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ of $\mathrm{SL}_2(\mathbb{Z})$-orbits in $\mathbb{H}$, with the orbits of i and ρ omitted. We note that this can be rewritten as

$$\frac{1}{2}\operatorname{ord}_\infty F + \sum_{z\in\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \frac{\operatorname{ord}_z F}{\#\,\mathrm{SL}_2(\mathbb{Z})_z} = \frac{dk}{24}.$$

(In this formula and in the rest of the proof, we will implicitly choose orbit and coset representatives where necessary.)

Let $z \in \mathbb{H}$. We apply Lemma 3.5 to the groups $G = \mathrm{SL}_2(\mathbb{Z})$ and $H = \Gamma$, with X taken to be the $\mathrm{SL}_2(\mathbb{Z})$-orbit of z. We rewrite $\operatorname{ord}_z F$ as follows:

$$\begin{aligned} \operatorname{ord}_z F &= \sum_{\gamma\in\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})} \operatorname{ord}_z(f|_k\gamma) \\ &= \sum_{\gamma\in\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})} \operatorname{ord}_{\gamma z} f \\ &= \sum_{w\in\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})z} (\mathrm{SL}_2(\mathbb{Z})_w : \Gamma_w)\operatorname{ord}_w f. \end{aligned}$$

In the last sum, we have used the fact that $\operatorname{ord}_{\gamma z} f$ depends only on γz and not on γ, and we have applied Lemma 3.5.

Since $\mathrm{SL}_2(\mathbb{Z})_w$ is finite and independent of $w \in \Gamma\backslash\mathrm{SL}_2(\mathbb{Z})z$, we may write

$$(\mathrm{SL}_2(\mathbb{Z})_w : \Gamma_w) = \frac{\#\,\mathrm{SL}_2(\mathbb{Z})_z}{\#\Gamma_w}$$

and divide the identity above by $\#\,\mathrm{SL}_2(\mathbb{Z})_z$; this gives

$$\frac{\operatorname{ord}_z F}{\#\,\mathrm{SL}_2(\mathbb{Z})_z} = \sum_{w\in\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})z} \frac{\operatorname{ord}_w f}{\#\Gamma_w}.$$

Summing over (a system of orbit representatives for) the quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, we obtain

$$\begin{aligned}\sum_{z\in\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \frac{\operatorname{ord}_z F}{\#\,\mathrm{SL}_2(\mathbb{Z})_z} &= \sum_{z\in\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}}\ \sum_{w\in\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})z} \frac{\operatorname{ord}_w f}{\#\Gamma_w} \\ &= \sum_{w\in\Gamma\backslash\mathbb{H}} \frac{\operatorname{ord}_w f}{\#\Gamma_w}.\end{aligned}$$

In Exercises 3.14 and 3.15, it is shown that the orders of f and F at the cusps satisfy

$$\frac{1}{2}\operatorname{ord}_\infty F = \sum_{\mathfrak{c}\in\mathrm{Cusps}(\Gamma)} \frac{\operatorname{ord}_{\Gamma,\mathfrak{c}}(f)}{\epsilon_{\Gamma,\mathfrak{c}}}. \tag{3.5}$$

We conclude that

$$\begin{aligned}\sum_{w\in\Gamma\backslash\mathbb{H}} \frac{\operatorname{ord}_w f}{\#\Gamma_w} + \sum_{\mathfrak{c}\in\mathrm{Cusps}(\Gamma)} \frac{\operatorname{ord}_{\Gamma,\mathfrak{c}}(f)}{\epsilon_{\Gamma,\mathfrak{c}}} &= \sum_{z\in\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}} \frac{\operatorname{ord}_z F}{\#\,\mathrm{SL}_2(\mathbb{Z})_z} + \frac{1}{2}\operatorname{ord}_\infty(F) \\ &= \frac{k}{24}(\mathrm{SL}_2(\mathbb{Z}) : \Gamma),\end{aligned}$$

which proves the first formula from the theorem. For the second formula, we first note the identities

$$\#(\Gamma\cap\{\pm1\})(\mathrm{SL}_2(\mathbb{Z}) : \Gamma) = 2(\mathrm{PSL}_2(\mathbb{Z}) : \bar\Gamma)$$

and

$$\epsilon_{\Gamma,\mathfrak{c}} = \#(\Gamma\cap\{\pm1\})\bar\epsilon_{\Gamma,\mathfrak{c}}.$$

The second identity can be checked by distinguishing the three possible cases: $-1\in\Gamma$ and $\mathfrak{c}$ regular; $-1\notin\Gamma$ and $\mathfrak{c}$ regular; $-1\notin\Gamma$ and $\mathfrak{c}$ irregular. The second formula now follows from the first by multiplying by $\#(\Gamma\cap\{\pm1\})$ and rewriting. □

Corollary 3.12. *Let $f\in\mathrm{M}_k(\Gamma)$ be a modular form with q-expansion $\sum_{n=0}^\infty a_n q^n$ at some cusp $\mathfrak{c}$ of Γ. Suppose we have*

$$a_j = 0 \quad \textit{for } j = 0, 1, \ldots, \left\lfloor \frac{k}{24}\epsilon_{\Gamma,\mathfrak{c}}(\mathrm{SL}_2(\mathbb{Z}) : \Gamma)\right\rfloor.$$

Then $f = 0$. Similarly, two forms in $\mathrm{M}_k(\Gamma)$ are equal whenever their q-expansions at $\mathfrak{c}$ agree to this precision.

Corollary 3.13. *The space of modular forms of weight k for a congruence subgroup Γ has dimension at most $1 + \left\lfloor \frac{k}{12}(\mathrm{SL}_2(\mathbb{Z}) : \Gamma)\right\rfloor$.*

There also exist formulae giving the dimensions of $\mathrm{M}_k(\Gamma)$ and $\mathrm{S}_k(\Gamma)$; these are rather complicated and will not be given here. In the book of Diamond and Shurman, a whole chapter is devoted to dimension formulae [4, Chapter 3].

3.7 Dirichlet characters

To continue developing the theory of modular forms for congruence subgroups (and in particular $\Gamma_0(N)$ and $\Gamma_1(N)$), it is essential to study Dirichlet characters first.

Definition. Let N be a positive integer. A *Dirichlet character* modulo N is a function

$$\chi\colon \mathbb{Z} \to \mathbb{C}$$

with the property that there exists a group homomorphism $\chi'\colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ such that

$$\chi(d) = \begin{cases} \chi'(d \bmod N) & \text{if } \gcd(d, N) = 1, \\ 0 & \text{if } \gcd(d, N) \ne 1. \end{cases}$$

Alternatively, a Dirichlet character modulo N is a function $\chi\colon \mathbb{Z} \to \mathbb{C}$ such that $\chi(m) = 0$ if and only if $\gcd(m, N) > 1$, and $\chi(mm') = \chi(m)\chi(m')$ for all $m \in \mathbb{Z}$.

The terminology "Dirichlet character" is often also used for the group homomorphism χ' itself. Since $(\mathbb{Z}/N\mathbb{Z})^\times$ is finite, the image of any group homomorphism $\chi'\colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ is contained in the the torsion subgroup of $\mathbb{C}^\times$, i.e. the group of roots of unity.

For fixed N, the set of Dirichlet characters modulo N is a group under pointwise multiplication. This group can be identified with $\mathrm{Hom}((\mathbb{Z}/N\mathbb{Z})^\times, \mathbb{C}^\times)$. By decomposing $(\mathbb{Z}/N\mathbb{Z})^\times$ as a product of cyclic groups, one sees that $\mathrm{Hom}((\mathbb{Z}/N\mathbb{Z})^\times, \mathbb{C}^\times)$ is *non-canonically* isomorphic to $(\mathbb{Z}/N\mathbb{Z})^\times$. In particular, its order is $\phi(N)$, where ϕ is Euler's ϕ-function.

Let M, N be positive integers with $M \mid N$, and let χ be a Dirichlet character modulo M. Then χ can be *lifted* to a Dirichlet character $\chi^{(N)}$ modulo N by putting

$$\chi^{(N)}(m) = \begin{cases} \chi(m) & \text{if } \gcd(m, N) = 1, \\ 0 & \text{if } \gcd(m, N) > 1. \end{cases}$$

The *conductor* of a Dirichlet character χ modulo N is the smallest divisor M of N such that there exists a Dirichlet character χ_M modulo M satisfying $\chi = \chi_M^{(N)}$. A Dirichlet character χ modulo N is called *primitive* if its conductor equals N.

Example. Modulo 1, we have the trivial character $\mathbf{1}\colon (\mathbb{Z}/1\mathbb{Z})^\times = \{0\} \to \mathbb{C}$. The corresponding Dirichlet character $\mathbf{1}\colon \mathbb{Z} \to \mathbb{C}$ is the constant function 1. For any N, lifting $\mathbf{1}$ to a Dirichlet character modulo N gives the function

$$\begin{aligned} \mathbf{1}^{(N)}\colon \mathbb{Z} &\to \mathbb{C} \\ m &\mapsto \begin{cases} 1 & \text{if } \gcd(m, N) = 1, \\ 0 & \text{if } \gcd(m, N) = 1. \end{cases} \end{aligned}$$

Example. Let $N = 4$. The group $(\mathbb{Z}/4\mathbb{Z})^\times$ has order 2. There exists a unique non-trivial Dirichlet character χ modulo 4, given by

$$\chi(m) = \begin{cases} 1 & \text{if } m \equiv 1 \pmod{4}, \\ -1 & \text{if } m \equiv 3 \pmod{4}, \\ 0 & \text{if } m \equiv 0, 2 \pmod{4}. \end{cases} \tag{3.6}$$

Example. We consider the case where N is a prime number p. We put

$$\left(\frac{a}{p}\right) = \begin{cases} 0 & \text{if } p \mid a, \\ 1 & \text{if } p \nmid a \text{ and } a \text{ is a square modulo } p, \\ -1 & \text{if } a \text{ is not a square modulo } p. \end{cases}$$

Then the map $a \mapsto \left(\frac{a}{p}\right)$ is a Dirichlet character modulo p. It is of conductor p if $p > 2$, and of conductor 1 if $p = 2$.

Consider two Dirichlet characters χ_1, χ_2 modulo N_1 and N_2, respectively. Let N be any common multiple of N_1 and N_2. Then χ_1 and χ_2 can be multiplied to give a Dirichlet character modulo N by putting

$$\chi = \chi_1\chi_2 = \chi_1^{(N)}\chi_2^{(N)}.$$

3.8 Application of modular forms to sums of squares

As an interesting application of modular forms, we will now study how they can be used to answer the following classical question.

Question. Given positive integers n and k, in how many ways can n be written as a sum of k squares of integers?

To make this question more precise, let us write

$$r_k(n) = \#\{(x_1, \ldots, x_k) \in \mathbb{Z}^k \mid x_1^2 + \cdots + x_k^2 = n\}.$$

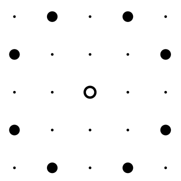
Then the question is how to (efficiently) compute $r_k(n)$. Note in particular that by this definition, changing the signs or the order of the x_i in some representation of n as a sum of k squares is regarded as giving a different representation. For example, we have

$$r_2(5) = 8,$$

the eight representations being

$$\begin{aligned} 5 &= 1^2 + 2^2 = (-1)^2 + 2^2 = 1^2 + (-2)^2 = (-1)^2 + (-2)^2 \\ &= 2^2 + 1^2 = (-2)^2 + 1^2 = 2^2 + (-1)^2 = (-2)^2 + (-1)^2. \end{aligned}$$

This can also be viewed geometrically as saying that in the square lattice $\mathbb{Z}^2 \subset \mathbb{R}^2$, there are 8 points whose distance from the origin equals $\sqrt{5}$:



Two of the most famous theorems concerning the question above were proved by Pierre de Fermat (1601–1665) and Joseph-Louis Lagrange (1736–1813).

Theorem 3.14 (Fermat). *Let n be an odd positive integer. Then n is a sum of two squares if and only if every prime number $p \mid n$ with $p \equiv 3 \pmod 4$ occurs an even number of times in the prime factorisation of n.*

Corollary 3.15. *Let p be a prime number. Then p is a sum of two squares if and only $p = 2$ or $p \equiv 1 \pmod 4$.*

Theorem 3.16 (Lagrange). *Every non-negative integer is a sum of four squares.*

These theorems can be proved without using modular forms; indeed, Fermat and Lagrange had no modular forms available to them. However, making use of modular forms leads to new insights and to generalisations of the results above.

In the theorems below, χ is the Dirichlet character modulo 4 given by (3.6).

The following formulae were found by C. G. J. Jacobi (1804–1851).

Theorem 3.17 (Jacobi, 1829)**.** *The functions $r_2(n)$ and $r_4(n)$ are given by the following formulae: for all $n \geq 1$, we have*

$$r_2(n) = 4\sum_{d|n} \chi(d),$$

$$r_4(n) = 8\sum_{\substack{d|n\\ 4\nmid d}} d.$$

In the first sum, d runs over the set of all positive divisors of n. In the second sum, d runs over the subset of such divisors that are not divisible by 4.

The following formulae for the cases $k = 6$ and $k = 8$ follow from work of Jacobi, F. G. M. Eisenstein (1823–1852) and H. J. S. Smith (1826–1883).

Theorem 3.18 (Jacobi, Eisenstein, Smith)**.** *The functions $r_6(n)$ and $r_8(n)$ are given by the following formulae: for all $n \geq 1$, we have*

$$r_6(n) = \sum_{d|n}\bigl(16\chi(n/d) - 4\chi(d)\bigr)d^2,$$

$$r_8(n) = 16\sum_{d|n}(-1)^{n-d}d^3.$$

Joseph Liouville (1809–1882) conjectured formulae for $r_k(n)$ in the cases $k = 10$ and $k = 12$. These formulae were later proved by J. W. L. Glaisher (1848–1928).

Theorem 3.19 (Glaisher (1907), conjectured by Liouville (1864/65))**.** *The functions $r_{10}(n)$ and $r_{12}(n)$ are given (partially) by the following formulae: for all $n \geq 1$, we have*

$$r_{10}(n) = \frac{4}{5}\sum_{d|n}(\chi(d) + 16\chi(n/d))d^2 + \frac{8}{5}\sum_{\substack{z\in\mathbb{Z}[i]\\ |z|^2=n}} z^4,$$

and for all even *$n \geq 2$, we have*

$$r_{12}(n) = 8\sum_{d|n} d^5 - 512\sum_{d|n/4} d^5.$$

It turns out that the function θ^k introduced in §3.4 is closely related to the counting function $r_k(n)$, as we will see in Exercise (3.16). We will use this relation to prove the formulae given above for $r_k(n)$ in the cases $k = 4$ and $k = 8$. The formulae for $k = 2$ and $k = 6$ can be proved using Eisenstein series with character; for these we refer to the exercises.

By Theorem 3.10, the function θ^k is in $\mathrm{M}_{k/2}(\Gamma_1(4))$. The group $\Gamma_1(4)$ has index 12 in $\mathrm{SL}_2(\mathbb{Z})$. The valence formula for congruence groups (Theorem 3.11) therefore implies that the dimension of the space $\mathrm{M}_{k/2}(\Gamma_1(4))$ satisfies the bound

$$\dim \mathrm{M}_{k/2}(\Gamma_1(4)) \leq 1 + \lfloor k/4 \rfloor.$$

Furthermore, the group $\Gamma_1(4)$ has three cusps, two of which are regular; see Exercise 3.8. Another consequence of Theorem 3.11 is therefore that for $k \in \{2, 4, 6, 8\}$, the space $\mathrm{S}_{k/2}(\Gamma_1(4))$ is trivial.

The easiest cases (given what we have seen so far) are $k = 4$ and $k = 8$. For $k = 4$, the relevant space $\mathrm{M}_2(\Gamma_1(4))$ has dimension 2. We already know two linearly independent elements of this space, namely $E_2^{(2)}(z) = E_2(z) - 2E_2(2z)$ and $E_2^{(4)}(z) = E_2(z) - 4E_2(4z)$. These elements therefore form a basis for $\mathrm{M}_2(\Gamma_1(4))$. From the q-expansion

$$E_2(z) = -\frac{1}{24} + \sum_{n=1}^{\infty}\sigma_1(n)q^n,$$

we compute

$$\begin{aligned} E_2(z) &= -\frac{1}{24} + q + 3q^2 + O(q^3), \\ E_2(2z) &= -\frac{1}{24} + q^2 + O(q^3), \\ E_2(4z) &= -\frac{1}{24} + O(q^3). \end{aligned}$$

On the other hand, we have

$$\theta(z)^4 = 1 + 8q + 24q^2 + O(q^3).$$

This implies that we have

$$\theta(z)^4 = c_1 E_2(z) + c_2 E_2(2z) + c_4 E_2(4z),$$

where the coefficients c_1, c_2, c_4 are obtained by solving the linear system

$$\begin{pmatrix} -1/24 & -1/24 & -1/24 \\ 1 & 0 & 0 \\ 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 24 \end{pmatrix}.$$

The unique solution of this system is

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \begin{pmatrix} 8 \\ 0 \\ -32 \end{pmatrix}.$$

This means that

$$\begin{aligned} \theta(z)^4 &= 8(E_2(z) - 4E_2(4z)) \\ &= 8E_2^{(4)}(z). \end{aligned}$$

Taking coefficients, we obtain

$$\begin{aligned} r_4(n) &= 8\Biggl(\sum_{d\mid n} d - 4 \sum_{d\mid(n/4)} d\Biggr) \\ &= 8 \sum_{\substack{d\mid n \\ 4\nmid d}} d, \end{aligned}$$

where the sum over the divisors of $n/4$ is only included if n is divisible by 4.

For $k = 8$, the relevant space $\mathrm{M}_4(\Gamma_1(4))$ has dimension 3. We already know three linearly independent elements of this space, namely $E_4(z)$, $E_4(2z)$ and $E_4(4z)$; these elements therefore form a basis for $\mathrm{M}_4(\Gamma_1(4))$. We have

$$E_4(z) = \frac{1}{240} + \sum_{n=1}^{\infty} \sigma_3(n) q^n.$$

Doing a similar calculation as above gives

$$\theta(z)^8 = 16E_4(z) - 32E_4(2z) + 256E_4(4z).$$

This implies

$$r_8(n) = 16\Biggl(\sum_{d\mid n} d^3 - 2 \sum_{d\mid(n/2)} d^3 + 16 \sum_{d\mid(n/4)} d^3\Biggr). \tag{3.7}$$

3.9 Exercises

Exercise 3.1. Show that the reduction map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective by completing the steps below.

- Let $\gamma \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ and choose a lift $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in M_2(\mathbb{Z})$. Show that $\gcd(a, b, N) = 1$.
- Show that there exist $k, l \in \mathbb{Z}$ such that $\gcd(a + kN, b + lN) = 1$. (*Hint:* $\gcd(a, b, N) = \gcd(\gcd(a, b), N)$.)
- Show that $\left(\begin{smallmatrix} a+kN & b+lN \\ c+mN & d+nN \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ for certain $m, n \in \mathbb{Z}$.

Exercise 3.2. Prove Proposition 3.1

Exercise 3.3. Recall that if N is a positive integer, $\Gamma(N)$ denotes the principal congruence subgroup of level N.

Let D and N be positive integers, and let β be a 2×2 matrix with integral entries and determinant D.

(a) Prove that $\beta\Gamma(DN)\beta^{-1}$ is contained in $\Gamma(N)$.

(b) Deduce that $\Gamma(N) \cap \beta^{-1}\Gamma(N)\beta$ contains $\Gamma(DN)$.

(c) Now let Γ be any congruence subgroup, and let α be in $\mathrm{GL}_2^+(\mathbb{Q})$. Prove that the group $\Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha$ is again a congruence subgroup.

Exercise 3.4. Prove that the element $\gamma \in \Gamma$ from Proposition 3.2 is unique up to multiplication by an element of $\Gamma \cap \{\pm 1\}$, except possibly if γz lies on the boundary of $\mathcal{D}_\Gamma$.

Exercise 3.5. Show that the subgroup $H_{\mathfrak{c}} \subset \mathrm{SL}_2(\mathbb{Z})_\infty$ defined by (3.1) does not depend on the choice of t and γ_t.

Exercise 3.6. Prove Lemma 3.4.

Exercise 3.7. Let $\mathcal{L}_1(N)$ be the set of pairs (Λ, P) where Λ is a lattice in $\mathbb{C}$ and P is a point of order N in the group $\mathbb{C}/\Lambda$.

(a) Show that on $\mathcal{L}_1(N)$ there is an equivalence relation $\sim$ with the property that $(\Lambda, P) \sim (\Lambda', P')$ if and only if there exists $\alpha \in \mathbb{C}^\times$ such that for any $\omega \in \mathbb{C}$ with $\omega + \Lambda = P$ in $\mathbb{C}/\Lambda$ we have $\alpha\Lambda = \Lambda'$ and $\alpha\omega + \Lambda' = P'$ in $\mathbb{C}/\Lambda'$.

(b) Recall that $\Gamma_1(N)$ is the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ consisting of matrices of the form $\left(\begin{smallmatrix} a & b \\ Nc & d \end{smallmatrix}\right)$ with $a, b, c, d \in \mathbb{Z}$, $a \equiv d \equiv 1 \pmod{N}$ and $ad - Nbc = 1$. Prove that there is a bijection

$$\mathcal{L}_1(N)/{\sim} \cong \Gamma_1(N)\backslash\mathbb{H}.$$

(*Hint:* consider lattices together with a suitable $\mathbb{Z}$-basis (ω_1, ω_2), and use a similar argument as for the case of $\Gamma_0(N)$ treated in §3.1.)

Exercise 3.8. Show that the cusps of $\Gamma_1(4)$, viewed as $\Gamma_1(4)$-orbits in $\mathbb{P}^1(\mathbb{Q})$, are represented by the elements 0, 1/2 and ∞ of $\mathbb{P}^1(\mathbb{Q})$. For each of these cusps $\mathfrak{c}$, determine whether $\mathfrak{c}$ is regular or irregular, and compute its width $h_\Gamma(\mathfrak{c})$.

Exercise 3.9. Let p be an odd prime number. Determine a set of representatives for the $\Gamma_1(p)$-orbits in $\mathbb{P}^1(\mathbb{Q})$. For each of the corresponding cusps $\mathfrak{c}$ of $\Gamma_1(p)$, compute its width $h_\Gamma(\mathfrak{c})$.

Exercise 3.10. Let N be a positive integer, and let H be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^\times$. Show that the set

$$\Gamma_H = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; a, d \bmod N \text{ are in } H \text{ and } c \equiv 0 \pmod{N} \right\}$$

is a congruence subgroup, and determine its level.

Exercise 3.11. Let Γ and Γ' be two congruence subgroups such that $\Gamma' \subset \Gamma$. Let f be a meromorphic function on $\mathbb{H}$ that is weakly modular of weight k for Γ, and hence also for Γ'. Let $\mathfrak{c}' \in \mathrm{Cusps}(\Gamma')$, and let $\mathfrak{c}$ be its image under the natural map $\mathrm{Cusps}(\Gamma') \to \mathrm{Cusps}(\Gamma)$.

(a) Prove that $h_\Gamma(\mathfrak{c})$ divides $h_{\Gamma'}(\mathfrak{c}')$ and that $\tilde{h}_\Gamma(\mathfrak{c})$ divides $\tilde{h}_{\Gamma'}(\mathfrak{c}')$.

(b) Prove the identity

$$\frac{\mathrm{ord}_{\Gamma',\mathfrak{c}'}(f)}{\tilde{h}_{\Gamma'}(\mathfrak{c}')} = \frac{\mathrm{ord}_{\Gamma,\mathfrak{c}}(f)}{\tilde{h}_\Gamma(\mathfrak{c})}.$$

Exercise 3.12. Let $\Gamma' \subset \Gamma$ be two congruence subgroups, let $k \in \mathbb{Z}$, and let f be a meromorphic function on $\mathbb{H}$ that is weakly modular of weight k for Γ.

(a) Show that there is a canonical surjective map $\mathrm{Cusps}(\Gamma') \to \mathrm{Cusps}(\Gamma)$.

(b) Let $\mathfrak{c}'$ be in $\mathrm{Cusps}(\Gamma')$, and let $\mathfrak{c}$ be its image in $\mathrm{Cusps}(\Gamma)$. Show that f is holomorphic at $\mathfrak{c}$ if and only if f (viewed as a weakly modular function of weight k for Γ') is holomorphic at $\mathfrak{c}'$. Show also that f vanishes at $\mathfrak{c}$ if and only if f (viewed as a weakly modular function of weight k for Γ') vanishes at $\mathfrak{c}'$.

(c) Deduce that if f is a modular form (resp. a cusp form) of weight k for Γ, then f is a modular form (resp. a cusp form) of weight k for Γ'. (This shows that we have inclusions $\mathrm{M}_k(\Gamma) \subset \mathrm{M}_k(\Gamma')$ and $\mathrm{S}_k(\Gamma) \subset \mathrm{S}_k(\Gamma')$; this fact has been used implicitly in the lectures.)

Exercise 3.13. (basically taken from [4, Exercise 1.2.4]) Let $A := \left(\begin{smallmatrix} 1 & 0 \\ 4 & 1 \end{smallmatrix}\right)$ and $T := \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Show that $\langle A, T\rangle = \Gamma_1(4)$ as follows.

Denote $\Gamma := \langle A, T\rangle$. Let $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(4)$. Use the identity

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b' \\ c & nc+d \end{pmatrix}$$

to show that unless $c = 0$, some $\alpha\gamma$ with $\gamma \in \Gamma$ has bottom row (c', d') with $|d'| < |c'|/2$. Use the identity

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 4n & 1 \end{pmatrix} = \begin{pmatrix} a' & b \\ c+4nd & d \end{pmatrix}$$

to show that unless $d = 0$, some $\alpha\gamma$ with $\gamma \in \Gamma$ has bottom row (c', d') with $|c'| < 2|d'|$. Each multiplication reduces the positive integer quantity $\min\{|c|, 2|d|\}$, so the process must stop. Show that this means that $\alpha\gamma \in \Gamma$ for some $\gamma \in \Gamma$, hence $\alpha \in \Gamma$.

The goal of the next two exercises is to prove the formula (3.5). We use the notation from (the proof of) Theorem 3.11.

Exercise 3.14. Consider the set

$$Z = \mathrm{SL}_2(\mathbb{Z}) \Big/ \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \,\middle|\, b \in \mathbb{Z} \right\}$$

equipped with the natural left action of $\mathrm{SL}_2(\mathbb{Z})$. Let u denote the class of the unit matrix in Z.

(a) Show that there exists a unique map $Z \to \mathbb{P}^1(\mathbb{Q})$ that is compatible with the $\mathrm{SL}_2(\mathbb{Z})$-action and sends u to ∞.

(b) Consider the map

$$\begin{aligned} \Gamma\backslash Z &\to \mathrm{Cusps}(\Gamma) \\ x &\mapsto \bar{x} \end{aligned}$$

obtained by taking the quotient by Γ on both sides of the map from (a). Show that for each $\mathfrak{c} \in \mathrm{Cusps}(\Gamma)$, the fibre of this map over $\mathfrak{c}$ has cardinality $2/\epsilon_{\Gamma,\mathfrak{c}}$.

(c) Show that for each $x \in \Gamma\backslash Z$, the fibre of the natural map $\Gamma\backslash\mathrm{SL}_2(\mathbb{Z}) \to \Gamma\backslash Z$ over x has cardinality $\tilde{h}_\Gamma(\bar{x})$. (*Hint:* use Lemma 3.5(2).)

Exercise 3.15. Choose a congruence subgroup Γ' contained in Γ such that Γ' is normal in $\mathrm{SL}_2(\mathbb{Z})$. Let $\tilde{h}_{\Gamma'}$ be the common value of $\tilde{h}_{\Gamma'}(\mathfrak{c})$ for all cusps $\mathfrak{c}$ of Γ' (note that these are indeed equal because Γ' is normal in $\mathrm{SL}_2(\mathbb{Z})$).

(a) Show that all fibres of the natural map $\Gamma'\backslash\mathrm{SL}_2(\mathbb{Z}) \to \Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$ have cardinality $(\Gamma : \Gamma')$.

(b) Prove the identity

$$\sum_{\gamma\in\Gamma'\backslash\mathrm{SL}_2(\mathbb{Z})} \mathrm{ord}_{\Gamma',\overline{\gamma u}}(f) = (\Gamma : \Gamma')\tilde{h}_{\Gamma'}\,\mathrm{ord}_\infty(F).$$

(*Hint:* use part (a) to show that $F^{(\Gamma:\Gamma')} = \prod_{\gamma'\in\Gamma'\backslash\mathrm{SL}_2(\mathbb{Z})} f|_k\gamma'$, and use this identity to rewrite $\mathrm{ord}_\infty(F)$.)

(c) Prove the identity

$$\sum_{\gamma\in\Gamma'\backslash\mathrm{SL}_2(\mathbb{Z})} \mathrm{ord}_{\Gamma',\overline{\gamma u}}(f) = (\Gamma : \Gamma')\tilde{h}_{\Gamma'} \sum_{x\in\Gamma\backslash Z} \mathrm{ord}_{\Gamma,\bar{x}}(f).$$

(*Hint:* apply Exercise 3.11 to the left-hand side, and then use part (a) and Exercise 3.14(c) to rewrite the resulting sum.)

(d) Deduce the formula (3.5).

Exercise 3.16. Show that for any $k \geq 1$, we have

$$\theta^k = \sum_{n=0}^{\infty} r_k(n)q^n \quad \text{for all } k \geq 0$$

(where $q = \exp(2\pi i z)$ as usual), or equivalently

$$r_k(n) = a_n(\theta^k).$$

Exercise 3.17. Deduce from the formula (3.7) that

$$r_8(n) = 16\sum_{d|n}(-1)^{n-d}d^3.$$

Exercise 3.18. Let N be a positive integer. We consider the set

$$C_N = \left\{\begin{pmatrix} x \\ y \end{pmatrix} \in (\mathbb{Z}/N\mathbb{Z})^2 \mid \langle x, y\rangle = \mathbb{Z}/N\mathbb{Z}\right\}\Big/\{\pm 1\},$$

where $\langle x, y\rangle$ denotes the (additive) subgroup of $\mathbb{Z}/N\mathbb{Z}$ generated by x and y, and where the group $\{\pm 1\}$ acts from the right on C_N by $\binom{x}{y}\epsilon = \binom{\epsilon x}{\epsilon y}$. Note that the set C_N has a natural left $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$-action.

(a) Prove that there is a natural bijection

$$\mathrm{Cusps}(\Gamma(N)) \cong C_N.$$

(b) Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a congruence subgroup of level N. Let H be the image of Γ under the map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Show that there is a natural bijection

$$\mathrm{Cusps}(\Gamma) \cong H\backslash C_N.$$

(c) Describe how the widths of the cusps of a given congruence subgroup of level N can be determined using computations "in characteristic N", i.e. involving $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ and C_N instead of $\mathrm{SL}_2(\mathbb{Z})$ and $\mathbb{P}^1(\mathbb{Q})$.

(d) Use parts (b) and (c) to solve Exercise 3.9: given an odd prime number p, describe the set $\mathrm{Cusps}(\Gamma_1(p))$, and for each $\mathfrak{c} \in \mathrm{Cusps}(\Gamma_1(p))$, compute $h_\Gamma(\mathfrak{c})$.

Exercise 3.19. The goal of this exercise is to prove the implication '(ii) ⇒ (i)' of Theorem 3.7. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and let k be an integer. Let $f\colon \mathbb{H} \to \mathbb{C}$ be a holomorphic function that is weakly modular of weight k for Γ and holomorphic at the cusp ∞. Suppose that there exist positive real numbers C, d such that the coefficients a_n in the Fourier expansion

$$f(z) = \sum_{n=0}^{\infty} a_n q_\infty^n$$

satisfy

$$|a_n| \le Cn^d \quad \text{for all } n \in \mathbb{Z}_{>0}.$$

(a) Prove that there exist positive real numbers C_1 and C_2 such that for all $z \in \mathbb{H}$ we have

$$|f(z)| \le C_1 + C_2(\Im z)^{-d-1}.$$

(*Hint:* bound $|f(z)|$ by comparing $\sum_{n=1}^{\infty} |a_n q_\infty^n|$ to an integral of the form $\int_0^\infty t^d \exp(-at)dt$.)

(b) Prove that for any $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, the function $z \mapsto (f|_k\alpha)(z)$ grows at most polynomially when $\Im z \to \infty$, i.e. that there exist positive real numbers C_3 and e such that

$$\bigl|(f|_k\alpha)(z)\bigr| \le C_3(\Im z)^e \quad \text{for all } z \in \mathbb{H} \text{ with } \Im z \ge 1.$$

(c) Deduce that f is a modular form of weight k for Γ.

(*Hint:* A version of this exercise with more intermediate steps is [4, Exercise 1.2.6].)

In the exercises below, N denotes a positive integer.

Exercise 3.20.

(a) Let χ be a Dirichlet character modulo N. Prove that

$$\sum_{j=0}^{N-1} \chi(j) = \begin{cases} \phi(N) & \text{if } \chi = \mathbf{1}_N, \\ 0 & \text{otherwise.} \end{cases}$$

(b) Let j be an integer. Prove that

$$\sum_{\chi \in D_N} \chi(j) = \begin{cases} \phi(N) & \text{if } j \in N\mathbb{Z}, \\ 0 & \text{otherwise.} \end{cases}$$

where D_N is the group of all Dirichlet characters modulo N.

Exercise 3.21. For integers $k > 0$ and $n \ge 0$, write

$$r_k(n) = \#\{(x_1, \ldots, x_k) \in \mathbb{Z}^k \mid x_1^2 + \cdots + x_k^2 = n\}.$$

Furthermore, let χ be the unique non-trivial Dirichlet character modulo 4. In this exercise you may assume without proof that there exist modular forms $E_1^{\mathbf{1},\chi} \in \mathrm{M}_1(\Gamma_1(4))$ and $E_3^{\mathbf{1},\chi}, E_3^{\chi,\mathbf{1}} \in$

$\mathrm{M}_3(\Gamma_1(4))$ with q-expansions

$$\begin{aligned} E_1^{\mathbf{1},\chi} &= \frac{1}{4} + \sum_{n=1}^{\infty}\Bigl(\sum_{d\mid n}\chi(d)\Bigr)q^n, \\ E_3^{\mathbf{1},\chi} &= -\frac{1}{4} + \sum_{n=1}^{\infty}\Bigl(\sum_{d\mid n}\chi(d)d^2\Bigr)q^n, \\ E_3^{\chi,\mathbf{1}} &= \sum_{n=1}^{\infty}\Bigl(\sum_{d\mid n}\chi(n/d)d^2\Bigr)q^n. \end{aligned}$$

(These are examples of Eisenstein series for $\Gamma_1(4)$. For a construction of the last two forms, see Exercise 3.26. Eisenstein series of weight 1 will not be constructed in this course.)

(a) Prove the formula

$$r_2(n) = 4\sum_{d\mid n}\chi(d) \quad \text{for all } n \ge 1.$$

(*Note:* If you know about arithmetic in the ring $\mathbb{Z}[i]$ of Gaussian integers, you can also prove this formula by counting ideals of norm n in $\mathbb{Z}[i]$.)

(b) Prove the formula

$$r_6(n) = \sum_{d\mid n}(16\chi(n/d) - 4\chi(d))d^2 \quad \text{for all } n \ge 1.$$

Exercise 3.22. Let $\chi\colon \mathbb{Z} \to \mathbb{C}$ be a Dirichlet character modulo N. The *L-function* of χ is the holomorphic function $L(\chi, s)$ (of the variable s) defined by

$$L(\chi, s) = \sum_{n=1}^{\infty}\chi(n)n^{-s}.$$

(a) Prove that the sum converges absolutely and uniformly on every right half-plane of the form $\{s \in \mathbb{C} \mid \Re s \ge \sigma\}$ with $\sigma > 1$.

(b) Prove the identity

$$L(\chi, s) = \prod_{p \text{ prime}}\frac{1}{1-\chi(p)p^{-s}} \quad \text{for } \Re s > 1.$$

(*Hint:* expand $\dfrac{1}{1-\chi(p)t}$ in a power series in t.)

Note: The functions $L(\chi, s)$ were introduced by P. G. Lejeune-Dirichlet in the proof of his famous theorem on primes in arithmetic progressions:

Theorem 3.20 (Dirichlet, 1837)**.** *Let N and a be coprime positive integers. Then there exist infinitely many prime numbers p with $p \equiv a \pmod{N}$.*

Exercise 3.23. Let χ be a Dirichlet character modulo N. We consider the function $\mathbb{Z} \to \mathbb{C}$ sending an integer m to the complex number

$$\tau(\chi, m) = \sum_{n=0}^{N-1}\chi(n)\exp(2\pi i m n/N).$$

(This can be viewed as a discrete Fourier transform of χ.) The case $m = 1$ deserves special mention: the complex number

$$\tau(\chi) = \tau(\chi, 1) = \sum_{n=0}^{N-1}\chi(n)\exp(2\pi i n/N)$$

is called the *Gauss sum* attached to χ.

(a) Compute $\tau(\chi)$ for all non-trivial Dirichlet characters χ modulo 4 and modulo 5, respectively.

(b) Suppose that χ is primitive. Prove that for all $m \in \mathbb{Z}$ we have

$$\tau(\chi, m) = \bar{\chi}(m)\tau(\chi).$$

(*Hint:* writing $d = \gcd(m, N)$, distinguish the cases $d = 1$ and $d > 1$.)

(c) Deduce that if χ is primitive, we have

$$\tau(\chi)\tau(\bar{\chi}) = \chi(-1)N$$

and

$$\tau(\chi)\overline{\tau(\chi)} = N.$$

Exercises 3.24, 3.25 and 3.26 are optional. The goal is to construct Eisenstein series with character. In each exercise you may use the results of all preceding exercises.

Exercise 3.24. Let χ be a primitive Dirichlet character modulo N. The *generalised Bernoulli numbers* attached to χ are the complex numbers $B_k(\chi)$ for $k \geq 0$ defined by the identity

$$\sum_{k=0}^{\infty} \frac{B_k(\chi)}{k!} t^k = \frac{t}{\exp(Nt) - 1} \sum_{j=1}^{N} \chi(j) \exp(jt)$$

in the ring $\mathbb{C}[[t]]$ of formal power series in t.

(a) Let ζ be a primitive N-th root of unity in $\mathbb{C}$. Prove that if χ is non-trivial (i.e. $N > 1$), then we have

$$\sum_{j=0}^{N-1} \chi(j) \frac{x + \zeta^j}{x - \zeta^j} = \frac{2N}{\tau(\bar{\chi})(x^N - 1)} \sum_{m=0}^{N-1} \bar{\chi}(m) x^m$$

in the field $\mathbb{C}(x)$ of rational functions in the variable x. (*Hint:* compute residues.)

(b) Prove that for every integer $k \geq 2$ such that $(-1)^k = \chi(-1)$, the special value of the Dirichlet L-function of χ at k is

$$L(\chi, k) = -\frac{(2\pi i)^k B_k(\bar{\chi})}{2\tau(\bar{\chi}) N^{k-1} k!}.$$

Exercise 3.25. Let $k \geq 3$, and let α and β be Dirichlet characters modulo M and N, respectively. For all $k \geq 3$, we define a function $G_k^{\alpha,\beta} \colon \mathbb{H} \to \mathbb{C}$ by

$$G_k^{\alpha,\beta}(z) = \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n) \neq (0,0)}} \frac{\alpha(m)\bar{\beta}(n)}{(mz + n)^k}.$$

(a) Prove that the function $G_k^{\alpha,\beta}$ is weakly modular of weight k for the congruence subgroup

$$\Gamma_1(M, N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; \begin{array}{l} a \equiv d \equiv 1 \pmod{\mathrm{lcm}(M, N)}, \\ c \equiv 0 \pmod{M}, \quad b \equiv 0 \pmod{N} \end{array} \right\}.$$

(b) Show that $G_k^{\alpha,\beta}$ is the zero function unless $\alpha(-1)\beta(-1) = (-1)^k$.

(c) Prove the identity

$$G_k^{\alpha,\beta}(z) = 2\alpha(0) \sum_{n>0} \frac{\bar{\beta}(n)}{n^k} + 2 \sum_{m>0} \alpha(m) \sum_{n \in \mathbb{Z}} \frac{\bar{\beta}(n)}{(mz + n)^k}.$$

Exercise 3.26. Keeping the notation of Exercise 3.25, assume in addition that $\alpha(-1)\beta(-1) = (-1)^k$ and that the character β is primitive.

(a) Prove that for all $w \in \mathbb{H}$ we have

$$\sum_{n\in\mathbb{Z}} \frac{\bar\beta(n)}{(w+n)^k} = \frac{(-2\pi i)^k \tau(\bar\beta)}{N^k(k-1)!} \sum_{d=1}^{\infty} \beta(d) d^{k-1} \exp(2\pi i d w/N).$$

(b) Deduce the formula

$$\begin{aligned} G_k^{\alpha,\beta}(z) = {} & -\alpha(0)\frac{(2\pi i)^k B_k(\beta)}{\tau(\beta) N^{k-1} k!} \\ & + \frac{2(-2\pi i)^k \tau(\bar\beta)}{N^k(k-1)!} \sum_{d=1}^{\infty} \Bigl(\sum_{d|n} \alpha(n/d)\beta(d) d^{k-1}\Bigr) \exp(2\pi i n z/N). \end{aligned}$$

(c) Let $E_k^{\alpha,\beta}(z)$ be the unique scalar multiple of $G_k^{\alpha,\beta}(Nz)$ such that the coefficient of q in the q-expansion of $E_k^{\alpha,\beta}$ equals 1. Prove the identity

$$E_k^{\alpha,\beta}(z) = -\alpha(0)\frac{B_k(\beta)}{2k} + \sum_{n=1}^{\infty} \Bigl(\sum_{d|n} \alpha(n/d)\beta(d) d^{k-1}\Bigr) q^n.$$

(d) Prove that $E_k^{\alpha,\beta}(z)$ is a modular form of weight k for $\Gamma_1(MN)$.

Chapter 4

Hecke operators and eigenforms

So far, we have viewed M_k as a complex vector space. It turns out that this vector space has a very interesting additional structure: it is a module over a commutative ring called the *Hecke algebra*.

There are various ways of introducing Hecke operators. We will start with a group-theoretic construction, and then show how this construction can be interpreted in terms of lattices.

4.1 The operators T_α

We start by extending the action of $\mathrm{SL}_2(\mathbb{Z})$ on the set of meromorphic function $\mathbb{H}$ to an action of the group

$$\mathrm{GL}_2^+(\mathbb{Q}) = \{\gamma \in \mathrm{GL}_2(\mathbb{Q}) \mid \det\gamma > 0\}.$$

This is done (and one easily checks that this defines indeed an action), for any $k \in \mathbb{Z}$, by putting

$$(f|_k\gamma)(z) := \frac{(\det\gamma)^k}{(cz+d)^k} f\left(\frac{az+b}{cz+d}\right) \quad \text{for all } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{Q}) \text{ and } z \in \mathbb{H}.$$

Let Γ be a congruence subgroup, and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. We define

$$\Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha.$$

Let f be a modular form of weight k for Γ. Then $f|_k\alpha$ is invariant under the right action of $\alpha^{-1}\Gamma\alpha$, as a consequence (the details are left to the reader) it is a modular form for Γ'. We define

$$T_\alpha f = \sum_{[\gamma]\in\Gamma'\backslash\Gamma} f|_k\alpha\gamma,$$

where we have chosen representatives $\gamma \in \Gamma$ for the cosets $[\gamma] \in \Gamma'\backslash\Gamma$. One readily checks that $f|_k\alpha\gamma$ does not depend on the choice of γ for $[\gamma]$ and furthermore that $[\Gamma : \Gamma'] < \infty$ so that the sum above is finite. It follows that $T_\alpha f$ is well defined.

Proposition 4.1. *Let Γ be a congruence subgroup, let k be an integer, and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Then for any $f \in \mathrm{M}_k(\Gamma)$, the function $T_\alpha f$ is again in $\mathrm{M}_k(\Gamma)$. Furthermore, if f is in $\mathrm{S}_k(\Gamma)$, then so is $T_\alpha f$.*

For the proof, see Exercise 4.1.

By the proposition, the map $f \mapsto T_\alpha f$ defines an endomorphism of the $\mathbb{C}$-vector space $\mathrm{M}_k(\Gamma)$, and this operator preserves the subspace $\mathrm{S}_k(\Gamma)$.

Remark. Note that we have an isomorphism

$$\Gamma'\backslash\Gamma \xrightarrow{\sim} (\alpha^{-1}\Gamma\alpha)\backslash(\alpha^{-1}\Gamma\alpha\Gamma).$$

Left multiplication by α identifies the right-hand side with $\Gamma\backslash\Gamma\alpha\Gamma$. Alternatively, noting that $\alpha\Gamma'\alpha^{-1}$ equals $\alpha\Gamma\alpha^{-1}\cap\Gamma$, we see that left multiplication by α is an isomorphism

$$\begin{aligned}\Gamma'\backslash\Gamma &\xrightarrow{\sim} (\alpha\Gamma\alpha^{-1}\cap\Gamma)\backslash\alpha\Gamma\\ &\xrightarrow{\sim} \Gamma\backslash\Gamma\alpha\Gamma.\end{aligned}$$

Either way, we have a composed isomorphism

$$\begin{aligned}\Gamma'\backslash\Gamma &\xrightarrow{\sim} \Gamma\backslash\Gamma\alpha\Gamma\\ \Gamma'\gamma &\longmapsto \Gamma\alpha\gamma.\end{aligned}$$

This shows that $T_\alpha f$ can also be expressed as

$$T_\alpha f = \sum_{[\gamma]\in\Gamma\backslash\Gamma\alpha\Gamma} f|_k\gamma.$$

An important special case is where α normalises Γ. In this case we have $\Gamma' = \Gamma$ and $T_\alpha f = f|_k\alpha$. If moreover α is in Γ, then $T_\alpha f = f$.

4.2 Hecke operators for $\Gamma_1(N)$

We now choose a positive integer N and take $\Gamma = \Gamma_1(N)$. We recall that we have a group isomorphism

$$\begin{aligned}\Gamma_1(N)\backslash\Gamma_0(N) &\xrightarrow{\sim} (\mathbb{Z}/N\mathbb{Z})^\times\\ \Gamma_1(N)\begin{pmatrix} a & b\\ c & d\end{pmatrix} &\longmapsto d \bmod N.\end{aligned}$$

Furthermore, given an integer d coprime to N, we can find a matrix $\alpha = \left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(N)$ with lower right entry d. For any such α, we put

$$\langle d\rangle f = T_\alpha f \quad \text{for all } f\in \mathrm{M}_k(\Gamma_1(N)).$$

More concretely, this means

$$(\langle d\rangle f)(z) = (cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right).$$

As the notation suggests, $T_\alpha f$ only depends on the class of α in $\Gamma_1(N)\backslash\Gamma_0(N)$, that is to say, on $d \bmod N$. This gives an action of the group $(\mathbb{Z}/N\mathbb{Z})^\times$ on $\mathrm{M}_k(\Gamma_1(N))$.

Proposition 4.2. *The subspace of* $(\mathbb{Z}/N\mathbb{Z})^\times$*-invariants in* $\mathrm{M}_k(\Gamma_1(N))$ *is equal to* $\mathrm{M}_k(\Gamma_0(N))$.

For the proof, see Exercise 4.3.

Next, we take $\alpha = \left(\begin{smallmatrix} 1 & 0\\ 0 & p\end{smallmatrix}\right)$; note that this is the first matrix we consider that is not in $\mathrm{SL}_2(\mathbb{Z})$.

Definition. Let N be a positive integer, and let p be a prime number. The *Hecke operator* T_p is the $\mathbb{C}$-linear endomorphism of $\mathrm{M}_k(\Gamma_1(N))$ defined by

$$T_p f = \frac{1}{p} T_{\left(\begin{smallmatrix} 1 & 0\\ 0 & p\end{smallmatrix}\right)} f \quad \text{for all } f\in \mathrm{M}_k(\Gamma_1(N)).$$

Remark. The factor $\frac{1}{p}$ is included to give nicer formulae.

We will need to work out the definition above of the operator T_p more concretely.

Lemma 4.3. *Let N be a positive integer, let p be a prime number, and let*

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \quad \Gamma = \Gamma_1(N), \quad \Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha.$$

Then we have

$$\Gamma' = \left\{ \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; \gamma \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \pmod N \text{ and } p \mid b \right\}.$$

A system of coset representatives for the quotient $\Gamma'\backslash\Gamma$ is given by

$$\begin{cases} \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \;\middle|\; 0 \le b \le p-1 \right\} & \text{if } p \mid N, \\ \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \;\middle|\; 0 \le b \le p-1 \right\} \cup \left\{ \begin{pmatrix} ap & 1 \\ cN & 1 \end{pmatrix} \right\} & \text{if } p \nmid N, \end{cases}$$

where $a, c \in \mathbb{Z}$ are such that $ap - cN = 1$.

Proof. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. We compute

$$\begin{aligned} \alpha^{-1}\gamma\alpha &= \begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & p^{-1} \end{pmatrix} \begin{pmatrix} a & bp \\ c & dp \end{pmatrix} \\ &= \begin{pmatrix} a & bp \\ c/p & d \end{pmatrix}. \end{aligned}$$

Hence Γ' consists of those matrices that are in Γ and whose upper right coefficient is divisible by p; this implies the first claim of the lemma.

To find systems of coset representatives, we consider the map

$$\Gamma = \Gamma_1(N) \to \mathrm{SL}_2(\mathbb{F}_p)$$

in the cases $p \mid N$ and $p \nmid N$, respectively.

In the case $p \mid N$, the image of the map above consists of all matrices of the form $\left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$, and the inverse image of $\left\{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)\right\} \subset \mathrm{SL}_2(\mathbb{F}_p)$ equals Γ'. We therefore have

$$\Gamma'\backslash\Gamma \cong \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \Bigg\backslash \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \;\middle|\; b \in \mathbb{F}_p \right\}.$$

This description shows that the matrices $\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$ for $0 \le b \le p-1$ form a system of coset representatives for the quotient $\Gamma'\backslash\Gamma$.

In the case $p \nmid N$, the reduction map $\Gamma \to \mathrm{SL}_2(\mathbb{F}_p)$ is surjective and the inverse image of the group of lower triangular matrices in $\mathrm{SL}_2(\mathbb{F}_p)$ under this map is Γ'. This implies

$$\Gamma'\backslash\Gamma \cong \left\{ \begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix} \;\middle|\; a \in \mathbb{F}_p^\times, c \in \mathbb{F}_p \right\} \Bigg\backslash \mathrm{SL}_2(\mathbb{F}_p).$$

It is left to the reader to check (Exercise 4.4) that a system of coset representatives for this quotient is

$$\left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \;\middle|\; b \in \mathbb{F}_p \right\} \cup \left\{ \begin{pmatrix} 0 & 1 \\ cN & 1 \end{pmatrix} \bmod p \right\},$$

where c is any integer with $cN \equiv -1 \pmod p$. Given c, there exists a unique $a \in \mathbb{Z}$ such that $ap - cN = 1$. This implies that the set of matrices given in the lemma is a system of coset representatives for $\Gamma'\backslash\Gamma$. □

We now apply this lemma to the definition of the operator $T_{\left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)}$. For $0 \le b \le p-1$, we have

$$\begin{aligned}\left(f|_k\left(\begin{pmatrix}1&0\\0&p\end{pmatrix}\begin{pmatrix}1&b\\0&1\end{pmatrix}\right)\right)(z) &= \left(f|_k\begin{pmatrix}1&b\\0&p\end{pmatrix}\right)(z)\\ &= \frac{p^k}{p^k}f\left(\frac{z+b}{p}\right)\\ &= f\left(\frac{z+b}{p}\right).\end{aligned}$$

In the case $p \nmid N$, choosing a matrix $\left(\begin{smallmatrix}ap&1\\cN&1\end{smallmatrix}\right)$ as above, we note that

$$\begin{pmatrix}1&0\\0&p\end{pmatrix}\begin{pmatrix}ap&1\\cN&1\end{pmatrix} = \begin{pmatrix}ap&1\\cNp&p\end{pmatrix} = \begin{pmatrix}a&1\\cN&p\end{pmatrix}\begin{pmatrix}p&0\\0&1\end{pmatrix}.$$

This implies

$$\begin{aligned}\left(f|_k\left(\begin{pmatrix}1&0\\0&p\end{pmatrix}\begin{pmatrix}ap&1\\cN&1\end{pmatrix}\right)\right)(z) &= \left(f|_k\left(\begin{pmatrix}a&1\\cN&p\end{pmatrix}\begin{pmatrix}p&0\\0&1\end{pmatrix}\right)\right)(z)\\ &= \left((\langle p\rangle f)|_k\begin{pmatrix}p&0\\0&1\end{pmatrix}\right)(z)\\ &= p^k(\langle p\rangle f)(pz).\end{aligned}$$

The action of $T_{\left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)}$ on f is therefore given by the formula

$$(T_{\left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)}f)(z) = \sum_{b=0}^{p-1} f\left(\frac{z+b}{p}\right) + p^k(\langle p\rangle f)(pz),$$

where the last term is only included if $p \nmid N$.

Notation. From now on, if f is a modular form (of some weight) for a group of the form $\Gamma_1(N)$, we will write $a_n(f)$ for the n-th coefficient of the q-expansion of f at the cusp ∞ of Γ.

Theorem 4.4. *Let N be a positive integer, and let p be a prime number. The Hecke operator T_p on $\mathrm{M}_k(\Gamma_1(N))$ is given by*

$$(T_pf)(z) = \frac{1}{p}\sum_{b=0}^{p-1} f\left(\frac{z+b}{p}\right) + p^{k-1}(\langle p\rangle f)(pz),$$

and its effect on q-expansions at the cusp ∞ of Γ is given by

$$a_n(T_pf) = a_{pn}(f) + p^{k-1}a_{n/p}(\langle p\rangle f) \quad \textit{for all } n \ge 0,$$

or equivalently

$$T_pf = \sum_{n=0}^{\infty}\bigl(a_{pn}(f) + p^{k-1}a_{n/p}(\langle p\rangle f)\bigr)q^n \quad \textit{with } q = \exp(2\pi iz).$$

Here the expression $a_{n/p}(\langle p\rangle f)$ is only included if $p \nmid N$ and $p \mid n$.

Proof. The first formula follows from the definition of T_p and the expression above for $T_{\left(\begin{smallmatrix}1&0\\0&p\end{smallmatrix}\right)}$. It remains to prove the q-expansion formula. We note that by definition of the q-expansion we have

$$\begin{aligned}
(T_p f)(z) &= \frac{1}{p}\sum_{b=0}^{p-1} f\left(\frac{z+b}{p}\right) + p^{k-1}(\langle p\rangle f)(pz) \\
&= \frac{1}{p}\sum_{b=0}^{p-1}\sum_{n=0}^{\infty} a_n(f)\exp\left(2\pi i\frac{n(z+b)}{p}\right) + p^{k-1}\sum_{n=0}^{\infty} a_n(\langle p\rangle f)\exp(2\pi i pnz) \\
&= \frac{1}{p}\sum_{n=0}^{\infty}\left(\sum_{b=0}^{p-1}\exp\left(\frac{2\pi i nb}{p}\right)\right) a_n(f)\exp\left(\frac{2\pi i nz}{p}\right) + p^{k-1}\sum_{n=0}^{\infty} a_n(\langle p\rangle f)\exp(2\pi i pnz).
\end{aligned}$$

Next we note that

$$\sum_{b=0}^{p-1}\exp\left(\frac{2\pi i nb}{p}\right) = \begin{cases} p & \text{if } p \mid n, \\ 0 & \text{if } p \nmid n. \end{cases}$$

This implies

$$\begin{aligned}
(T_p f)(z) &= \sum_{\substack{n\ge 0\\ p\mid n}} a_n(f)\exp\left(\frac{2\pi i nz}{p}\right) + p^{k-1}\sum_{n=0}^{\infty} a_n(\langle p\rangle f)\exp(2\pi i pnz) \\
&= \sum_{n=0}^{\infty} a_{pn}(f)\exp(2\pi i nz) + p^{k-1}\sum_{n=0}^{\infty} a_{n/p}(\langle p\rangle f)\exp(2\pi i nz),
\end{aligned}$$

as claimed. □

4.3 Lattice interpretation of Hecke operators

We now give a more conceptual interpretation of Hecke operators in the case $N = 1$. (A similar explanation exists for other subgroups, but it is a bit more involved.)

Let $f \in \mathrm{M}_k = \mathrm{M}_k(\mathrm{SL}_2(\mathbb{Z}))$. As in §1.1, we write $\mathcal{L}$ for the set of all lattices in $\mathbb{C}$ and Λ_z for the lattice $\mathbb{Z}z + \mathbb{Z}$. We recall that there is a unique function

$$\mathcal{F}\colon \mathcal{L} \to \mathbb{C}$$

that is homogeneous of weight k and satisfies

$$\mathcal{F}(\Lambda_z) = f(z) \quad \text{for all } z \in \mathbb{H}.$$

Similarly, if p is a prime number, we write $T_p\mathcal{F}$ for the homogeneous function of weight k corresponding to $T_p f$.

Proposition 4.5. *Let $k \in \mathbb{Z}$, let $f \in \mathrm{M}_k(\mathrm{SL}_2(\mathbb{Z}))$, and let $\mathcal{F}$ be the homogeneous function associated to f. Then for every prime number p, we have*

$$(T_p\mathcal{F})(\Lambda) = \frac{1}{p}\sum_{\substack{\Lambda'\supset\Lambda\\ (\Lambda':\Lambda)=p}} \mathcal{F}(\Lambda') \quad \textit{for all } \Lambda \in \mathcal{L}.$$

Proof. By the homogeneity of $\mathcal{F}$, it suffices to consider the case where $\Lambda = \Lambda_z$ with $z \in \mathbb{H}$. Using

Theorem 4.4 and the fact that the group of diamond operators is trivial for $N = 1$, we compute

$$\begin{aligned}(T_p\mathcal{F})(\Lambda_z) &= (T_p f)(z)\\ &= \frac{1}{p}\sum_{b=0}^{p-1} f\Bigl(\frac{z+b}{p}\Bigr) + p^{k-1}f(pz)\\ &= \frac{1}{p}\sum_{b=0}^{p-1} \mathcal{F}\Bigl(\mathbb{Z}\frac{z+b}{p}+\mathbb{Z}\Bigr) + p^{k-1}\mathcal{F}(\mathbb{Z}pz+\mathbb{Z})\\ &= \frac{1}{p}\Biggl(\sum_{b=0}^{p-1} \mathcal{F}\Bigl(\mathbb{Z}\frac{z+b}{p}+\mathbb{Z}\Bigr) + \mathcal{F}\Bigl(\mathbb{Z}z+\mathbb{Z}\frac{1}{p}\Bigr)\Biggr).\end{aligned}$$

It is not hard to check that the lattices $\mathbb{Z}\frac{z+b}{p}+\mathbb{Z}$ $(0 \le b \le p-1)$ and $\mathbb{Z}z+\mathbb{Z}\frac{1}{p}$ are precisely the lattices containing Λ_z with index p. This proves the formula. □

Proposition 4.5 shows that the Hecke operator T_p "averages" the values $\mathcal{F}(\Lambda')$ over all lattices Λ' containing Λ with index p. It is not a real average, however, because the sum is divided by p instead of the number of such lattices, which is $p+1$.

4.4 The Hecke algebra

Let $N \ge 1$ and $k \in \mathbb{Z}$. By Proposition 4.1, the operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and T_p for p prime preserve the $\mathbb{C}$-vector spaces $\mathrm{M}_k(\Gamma_1(N))$ and $\mathrm{S}_k(\Gamma_1(N))$ of modular forms and cusp forms, respectively.

Definition. Let $N \ge 1$ and $k \in \mathbb{Z}$. The *Hecke algebra* acting on $\mathrm{M}_k(\Gamma_1(N))$ is the $\mathbb{C}$-subalgebra of $\mathrm{End}_{\mathbb{C}}\,\mathrm{M}_k(\Gamma_1(N))$ generated by

- the $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$;
- the T_p for p prime.

This algebra is denoted by $\mathbb{T}(\mathrm{M}_k(\Gamma_1(N)))$.

We similarly define the Hecke algebra $\mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$ as the $\mathbb{C}$-subalgebra of $\mathrm{End}_{\mathbb{C}}\,\mathrm{S}_k(\Gamma_1(N))$ generated by the same operators. Then we have a surjective ring homomorphism

$$\mathbb{T}(\mathrm{M}_k(\Gamma_1(N))) \twoheadrightarrow \mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$$

defined by sending each operator on $\mathrm{M}_k(\Gamma_1(N))$ to its restriction to $\mathrm{S}_k(\Gamma_1(N))$.

Proposition 4.6. *For every $N \ge 1$, the Hecke algebra $\mathbb{T}(\mathrm{M}_k(\Gamma_1(N)))$ is commutative.*

Proof. We first note that all diamond operators commute, because the group $(\mathbb{Z}/N\mathbb{Z})^\times$ is commutative; more precisely, for all $d, e \in (\mathbb{Z}/N\mathbb{Z})^\times$, we have

$$\langle d\rangle\langle e\rangle = \langle de\rangle = \langle ed\rangle = \langle e\rangle\langle d\rangle.$$

Next, we show that T_p and $\langle d\rangle$ commute for all prime numbers p and all $d \in (\mathbb{Z}/N\mathbb{Z})^\times$. We first lift d to an integer coprime to Np and then to a matrix

$$\gamma_d = \begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \Gamma_0(Np).$$

As before, we put

$$\alpha = \begin{pmatrix} 1 & 0\\ 0 & p\end{pmatrix}, \quad \Gamma = \Gamma_1(N), \quad \Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha.$$

We also define

$$\beta_d = \alpha^{-1}\gamma_d\alpha = \begin{pmatrix} a & pb \\ c/p & d \end{pmatrix} \in \Gamma_0(N).$$

We now compute

$$\begin{aligned} T_p(\langle d\rangle f) &= T_p(f|_k\gamma_d) \\ &= \frac{1}{p}\sum_{\gamma\in\Gamma'\backslash\Gamma}(f|_k\gamma_d)|_k\alpha\gamma \\ &= \frac{1}{p}\sum_{\gamma\in\Gamma'\backslash\Gamma} f|_k\alpha\beta_d\gamma. \end{aligned}$$

The matrix $\gamma_d \in \Gamma_0(N)$ normalises Γ, so β_d normalises $\alpha^{-1}\Gamma\alpha$. Furthermore, $\beta_d \in \Gamma_0(N)$ also normalises Γ, and hence Γ'. Conjugation by β_d therefore induces a permutation of the set $\Gamma'\backslash\Gamma$, so we can replace $\gamma \in \Gamma'\backslash\Gamma$ by $\delta = \beta_d\gamma\beta_d^{-1} \in \Gamma'\backslash\Gamma$. This gives

$$\begin{aligned} T_p(\langle d\rangle f) &= \frac{1}{p}\sum_{\delta\in\Gamma'\backslash\Gamma} f|_k\alpha\delta\beta_d \\ &= \frac{1}{p}\left(\sum_{\delta\in\Gamma'\backslash\Gamma} f|_k\alpha\delta\right)\Big|_k\beta_d \\ &= \langle d\rangle(T_pf). \end{aligned}$$

This shows that T_p and $\langle d\rangle$ commute, as claimed.

Next, we have to prove that T_p and $T_{p'}$ commute for all prime numbers p, p'. There are various ways to show this. An intrinsic proof would be based on group theory; however, we will give a more computational proof using q-expansions. Let $f \in \mathrm{M}_k(\Gamma_1(N))$. For all $n \geq 0$, we have by Theorem 4.4

$$a_n(T_{p'}f) = a_{p'n}(f) + (p')^{k-1}a_{n/p'}(\langle p'\rangle f)$$

and (applying the theorem to $T_{p'}$)

$$a_n(T_pT_{p'}f) = a_{pn}(T_{p'}f) + p^{k-1}a_{n/p}(\langle p\rangle T_{p'}f)$$

Using the fact that $\langle p\rangle$ and $T_{p'}$ commute and using the first formula, we can rewrite this as

$$\begin{aligned} a_n(T_pT_{p'}f) &= a_{pn}(T_{p'}f) + p^{k-1}a_{n/p}(T_{p'}(\langle p\rangle f)) \\ &= a_{p'pn}(f) + (p')^{k-1}a_{pn/p'}(\langle p'\rangle f) \\ &\quad + p^{k-1}a_{p'n/p}(\langle p\rangle f) + p^{k-1}(p')^{k-1}a_{n/(pp')}(\langle p\rangle\langle p'\rangle f). \end{aligned}$$

(As before, we put $a_m(f) = 0$ if m is not an integer, and $\langle p\rangle = 0$ if $p \mid N$.) The right-hand side is symmetric in p and p' for all n, which shows that $T_pT_{p'}f = T_{p'}T_pf$. Since this holds for all $f \in \mathrm{M}_k(\Gamma_1(N))$, we conclude that T_p and $T_{p'}$ commute in $\mathbb{T}(\mathrm{M}_k(\Gamma_1(N)))$. □

Definition. Let $N \geq 1$ and $k \in \mathbb{Z}$. The *Hecke operators* $T_n \in \mathbb{T}(\mathrm{M}_k(\Gamma_1(N)))$ for $n \geq 1$ are defined as follows, starting from the T_p for p prime defined before:

$$T_1 = \mathrm{id}, \quad T_p \text{ as before for } p \text{ prime},$$
$$T_{p^r} = T_pT_{p^{r-1}} - p^{k-1}\langle p\rangle T_{p^{r-2}} \quad \text{for } p \text{ prime and } r = 2, 3, \ldots,$$
$$T_n = \prod_{p \text{ prime}} T_{p^{e_p}} \quad \text{for } n = \prod_{p \text{ prime}} p^{e_p}.$$

Note that the ordering of the factors in the products does not matter because the Hecke algebra is commutative. Furthermore, the definition implies

$$T_{mn} = T_mT_n \quad \text{if } m \text{ and } n \text{ are coprime.}$$

Also, if $p|N$, then $\langle p\rangle = 0$, and the recursion for T_{p^r} simply yields $T_{p^r} = T_p^r$.

4.5 The effect of Hecke operators on q-expansions

Proposition 4.7. *Let $f \in \mathrm{M}_k(\Gamma_1(N))$, and let $m \geq 1$. Then the q-expansion coefficients of $T_m f$ at the cusp ∞ of $\Gamma_1(N)$ are given by*

$$a_n(T_m f) = \sum_{d \mid \gcd(m,n)} d^{k-1} a_{mn/d^2}(\langle d\rangle f) \quad \textit{for all } n \geq 0.$$

In particular, we have

$$a_0(T_m f) = \sum_{d\mid m} d^{k-1} a_0(\langle d\rangle f), \quad a_1(T_m f) = a_m(f).$$

Proof. Since $T_1 = \mathrm{id}$, the claim is true for $m = 1$; by Theorem 4.4 it also holds when m is a prime number.

Let p be a prime number; we prove by induction that the claim holds when m is a power of p. The known base cases are $m = 1$ and $m = p$. For the induction step, let $r \geq 2$ and suppose the formula holds for $m = p^0, \ldots, p^{r-1}$. We have to prove

$$a_n(T_{p^r} f) = \sum_{\substack{0\leq j\leq r\\ p^j \mid n}} p^{j(k-1)} a_{p^{r-2j}n}(\langle p\rangle^j f) \quad \text{for all } n \geq 0.$$

By the definition of T_{p^r}, we have

$$T_{p^r} f = T_p T_{p^{r-1}} f - p^{k-1}\langle p\rangle T_{p^{r-2}} f,$$

so that

$$\begin{aligned}
a_n(T_{p^r} f) &= a_n(T_p T_{p^{r-1}} f) - p^{k-1} a_n(\langle p\rangle T_{p^{r-2}} f)\\
&= a_{pn}(T_{p^{r-1}} f) + p^{k-1} a_{n/p}(\langle p\rangle T_{p^{r-1}} f) - p^{k-1} a_n(\langle p\rangle T_{p^{r-2}} f)\\
&= a_{pn}(T_{p^{r-1}} f) + p^{k-1} a_{n/p}(T_{p^{r-1}}\langle p\rangle f) - p^{k-1} a_n(T_{p^{r-2}}\langle p\rangle f)\\
&= \sum_{\substack{0\leq j\leq r-1\\ p^j\mid pn}} p^{j(k-1)} a_{p^{r-1-2j}pn}(\langle p\rangle^j f) + p^{k-1} \sum_{\substack{0\leq j\leq r-1\\ p^{j+1}\mid n}} p^{j(k-1)} a_{p^{r-1-2j}n/p}(\langle p\rangle^{j+1} f)\\
&\quad - p^{k-1} \sum_{\substack{0\leq j\leq r-2\\ p^j\mid n}} p^{j(k-1)} a_{p^{r-2-2j}n}(\langle p\rangle^{j+1} f)\\
&= \sum_{\substack{0\leq j\leq r-1\\ p^j\mid pn}} p^{j(k-1)} a_{p^{r-2j}n}(\langle p\rangle^j f) + \sum_{\substack{0\leq j\leq r-1\\ p^{j+1}\mid n}} p^{(j+1)(k-1)} a_{p^{r-2-2j}n}(\langle p\rangle^{j+1} f)\\
&\quad - \sum_{\substack{0\leq j\leq r-2\\ p^j\mid n}} p^{(j+1)(k-1)} a_{p^{r-2-2j}n}(\langle p\rangle^{j+1} f)\\
&= \sum_{\substack{0\leq j\leq r-1\\ p^j\mid pn}} p^{j(k-1)} a_{p^{r-2j}n}(\langle p\rangle^j f) + \sum_{\substack{1\leq j\leq r\\ p^j\mid n}} p^{j(k-1)} a_{p^{r-2j}n}(\langle p\rangle^j f)\\
&\quad - \sum_{\substack{1\leq j\leq r-1\\ p^j\mid pn}} p^{j(k-1)} a_{p^{r-2j}n}(\langle p\rangle^j f).
\end{aligned}$$

The first and third summation cancel except for the term corresponding to $j = 0$, which is $a_{p^r n}(f)$. This gives

$$a_n(T_{p^r f}) = a_{p^r n}(f) + \sum_{\substack{1\leq j\leq r\\ p^j\mid n}} p^{j(k-1)} a_{p^{r-2j}n}(\langle p\rangle^j f).$$

This proves the claim for $m = p^r$. By induction, we conclude that the claim holds for all prime powers m and for all $n \geq 0$.

Next we prove that if m and m' are coprime and the claim holds for m and m', then the claim also holds for mm'. We compute

$$\begin{aligned} a_n(T_{mm'}f) &= a_n(T_m(T_{m'}f)) \\ &= \sum_{d\mid\gcd(m,n)} d^{k-1} a_{mn/d^2}(\langle d\rangle T_{m'}f) \\ &= \sum_{d\mid\gcd(m,n)} d^{k-1} a_{mn/d^2}(T_{m'}\langle d\rangle f) \\ &= \sum_{d\mid\gcd(m,n)} d^{k-1} \sum_{d'\mid\gcd(m',mn/d^2)} (d')^{k-1} a_{m'(mn/d^2)/(d')^2}(\langle d'\rangle(\langle d\rangle f)) \\ &= \sum_{d\mid\gcd(m,n)} \sum_{d'\mid\gcd(m',mn/d^2)} (dd')^{k-1} a_{mm'n/(dd')^2}(\langle dd'\rangle f). \end{aligned}$$

Since m and m' are coprime, we have

$$\gcd(m', mn/d^2) = \gcd(m', n) \quad \text{for all } d \mid m.$$

Furthermore, as d and d' range over the divisors of $\gcd(m, n)$ and $\gcd(m', n)$, respectively, $e = dd'$ ranges over the divisors of $\gcd(mm', n)$. This implies

$$a_n(T_{mm'}f) = \sum_{e\mid\gcd(mm',n)} e^{k-1} a_{mm'n/e^2}(\langle e\rangle f),$$

which proves the claim for mm'.

The claim for any $m \geq 1$ is now proved by induction on the number of prime factors of m. □

4.6 Hecke eigenforms

Definition. A *(Hecke) eigenform* is a non-zero modular form $f \in \mathrm{M}_k(\Gamma_1(N))$ that is an eigenvector for all Hecke operators T_p for p prime and all diamond operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$. A *normalised eigenform* is an eigenform f satisfying $a_1(f) = 1$.

Let $f \in \mathrm{M}_k(\Gamma_1(N))$ be an eigenform. If λ_n is the eigenvalue of T_n on f, then taking a_1 in the equation $T_n f = \lambda_n f$ and applying Proposition 4.7, we obtain

$$a_n(f) = \lambda_n a_1(f) \quad \text{for all } n \geq 1. \tag{4.1}$$

If f is an eigenform with $a_1(f) = 0$, then (4.1) shows that $a_n(f) = 0$ for all $n \geq 1$. Since $f \neq 0$ by assumption, this is only possible if $k = 0$, in which case f is a constant function. This means that when $k \geq 1$, any eigenform can be scaled to a *normalised* eigenform.

Theorem 4.8. *Let $f \in \mathrm{M}_k(\Gamma_1(N))$ be a normalised eigenform. Then the eigenvalues of the Hecke operators on f are equal to the q-expansion coefficients of f at the cusp ∞ of $\Gamma_1(N)$, i.e.*

$$T_n f = a_n(f) \cdot f \quad \textit{for all } n \geq 1.$$

Proof. This follows from (4.1) and the fact that $a_1(f) = 1$. □

We now consider the eigenvalues of the diamond operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$. Let us write

$$\chi\colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}$$

for the map that sends every $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ to the eigenvalue of $\langle d \rangle$ on f. Then for all $d, e \in (\mathbb{Z}/N\mathbb{Z})^\times$ we have

$$\chi(de)f = \langle de \rangle f = \langle d \rangle(\langle e \rangle f) = \langle d \rangle(\chi(e)f) = \chi(d)\chi(e)f,$$

so χ is a Dirichlet character modulo N (see §3.7).

Definition. Let $N \geq 1$, let $k \in \mathbb{Z}$, and let $\chi\colon (\mathbb{Z}/N\mathbb{Z})^\times$ be a group homomorphism. The space of *modular forms of weight k and character χ for $\Gamma_1(N)$* is the $\mathbb{C}$-linear subspace of $\mathrm{M}_k(\Gamma_1(N))$ consisting of the forms f satisfying

$$\langle d \rangle f = \chi(d) f \quad \text{for all } d \in (\mathbb{Z}/N\mathbb{Z})^\times.$$

It is denoted by $\mathrm{M}_k(\Gamma_1(N), \chi)$.

Remark. Some alternative notations for this space are $\mathrm{M}_k(N, \chi)$ and $\mathrm{M}_k(\Gamma_0(N), \chi)$. However, note that $\mathrm{M}_k(\Gamma_0(N), \chi)$ is in general not a subspace of $\mathrm{M}_k(\Gamma_0(N))$.

Similarly, we define the space of *cusp forms of weight k and character χ for $\Gamma_1(N)$* as the intersection

$$\mathrm{S}_k(\Gamma_1(N), \chi) = \mathrm{M}_k(\Gamma_1(N), \chi) \cap \mathrm{S}_k(\Gamma_1(N))$$

in $\mathrm{M}_k(\Gamma_1(N))$.

Proposition 4.9. *The $\mathbb{C}$-vector space $\mathrm{M}_k(\Gamma_1(N))$ has a decomposition*

$$\mathrm{M}_k(\Gamma_1(N)) = \bigoplus_{\chi} \mathrm{M}_k(\Gamma_1(N), \chi)$$

as the direct sum of its subspaces $\mathrm{M}_k(\Gamma_1(N), \chi)$ with χ running through all Dirichlet characters modulo N. The analogous statement holds for the space $\mathrm{S}_k(\Gamma_1(N))$.

Proof. Because the operators $\langle d \rangle$ have finite order, they are diagonalisable. Since they commute with each other, the space $\mathrm{M}_k(\Gamma_1(N))$ can be decomposed as a direct sum of simultaneous eigenspaces for all the $\langle d \rangle$ (see also Theorem A.9). Let E be such an eigenspace, and let $\chi(d)$ denote the eigenvalue of $\langle d \rangle$ on E. Then for all $d, e \in (\mathbb{Z}/N\mathbb{Z})^\times$, the fact that $\langle de \rangle = \langle d \rangle \langle e \rangle$ implies $\chi(de) = \chi(d)\chi(e)$, so χ is a Dirichlet character. Therefore the simultaneous eigenspaces E are precisely the spaces $\mathrm{M}_k(\Gamma_1(N), \chi)$ with χ running through the Dirichlet characters modulo N. □

Theorem 4.10. *Let $f \in \mathrm{M}_k(\Gamma_1(N), \chi)$ be a normalised eigenform, and let $\sum_{n=0}^{\infty} a_n q^n$ be the q-expansion of f at ∞. Then the a_n for $n \geq 1$ can be expressed recursively in terms of the a_p for p prime by*

$$\begin{aligned} a_1 &= 1, \\ a_{p^r} &= a_p a_{p^{r-1}} - p^{k-1}\chi(p) a_{p^{r-2}} \quad \textit{for } p \textit{ prime and } r = 2, 3, \ldots, \\ a_n &= \prod_{p \text{ prime}} a_{p^{e_p}} \quad \textit{for } n = \prod_{p \text{ prime}} p^{e_p}. \end{aligned}$$

Proof. This follows from Theorem 4.8 and the definition of the Hecke operators. □

Corollary 4.11. *With the notation of the theorem, we have*

$$a_{mn} = a_m a_n \quad \textit{if } \gcd(m, n) = 1.$$

Example. We take $N = 1$ and $k = 12$. We recall that there is a cusp form $\Delta \in \mathrm{S}_{12}(\mathrm{SL}_2(\mathbb{Z}))$, which is unique up to scaling since this space of cusp forms is one-dimensional. Since the Hecke algebra preserves the space of cusp forms, Δ is automatically an eigenform, and it has trivial character because $N = 1$.

We also recall that Ramanujan's τ-function is defined by the q-expansion of Δ as follows:

$$\Delta = \sum_{n=1}^{\infty} \tau(n) q^n.$$

Applying the results above to the eigenform Δ, we obtain the recurrence relations

$$\tau(p^r) = \tau(p)\tau(p^{r-1}) - p^{11}\tau(p^{r-2}) \quad \text{for } p \text{ prime and } r = 2,\, 3,\, \ldots$$

and

$$\tau(mn) = \tau(m)\tau(n) \quad \text{if } \gcd(m,n) = 1,$$

which were conjectured by Ramanujan.

4.7 Exercises

Exercise 4.1.

(a) Let Γ be a congruence subgroup, $k \in \mathbb{Z}$, $f \in M_k(\Gamma)$, $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, and denote $\Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha$. Prove that $f|_k\alpha \in M_k(\Gamma')$ (provide all details).

(b) Prove Proposition 4.1.

Exercise 4.2. Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$, and let $k \in \mathbb{Z}$.

(a) Let Γ' be a congruence subgroup contained in Γ, and let g be in $\mathrm{M}_k(\Gamma')$. Show that g is in $\mathrm{M}_k(\Gamma)$ if and only if g is invariant under the weight k action of Γ. (*Hint:* use Exercise 3.12.)

(b) Let f be in $\mathrm{M}_k(\Gamma)$, and let α be in $\mathrm{GL}_2^+(\mathbb{Q})$. Show that there exists a congruence subgroup Γ' contained in $\Gamma \cap \alpha^{-1}\Gamma\alpha$ such that for all $\gamma \in \Gamma$, the function $f|_k\alpha\gamma$ is in $\mathrm{M}_k(\Gamma')$.

(c) Show that the function

$$T_\alpha f = \sum_{\gamma \in \Gamma' \backslash \Gamma} f|_k\alpha\gamma$$

is in $\mathrm{M}_k(\Gamma)$.

Exercise 4.3. Prove Proposition 4.2.

Exercise 4.4. Read the proof of Lemma 4.3 and check that

$$\left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \,\middle|\, b \in \mathbb{F}_p \right\} \cup \left\{ \begin{pmatrix} 0 & 1 \\ cN & 1 \end{pmatrix} \bmod p \right\},$$

where c is any integer with $cN \equiv -1 \pmod{p}$, indeed forms a system of coset representatives for the quotient given at the end of the proof.

Exercise 4.5. Let N be a positive integer, let p be a prime number, and let

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \quad \Gamma = \Gamma_0(N) \text{ (instead of } \Gamma_1(N)), \quad \Gamma' = \Gamma \cap \alpha^{-1}\Gamma\alpha.$$

Determine a system of coset representatives for the quotient $\Gamma' \backslash \Gamma$.

Exercise 4.6. Prove that for any even integer $k \geq 4$ and prime p we have

$$T_p G_k = \sigma_{k-1}(p) G_k$$

for the Eisenstein series G_k and the Hecke operator T_p on $M_k(\mathrm{SL}_2(\mathbb{Z}))$.

Exercise 4.7. Let p be a prime and consider the lattice $\Lambda := \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ where $\omega_1, \omega_2 \in \mathbb{C}^*$ and $\omega_1/\omega_2 \notin \mathbb{R}$.

(a) Show that the lattices $\Lambda' \subset \mathbb{C}$ satisfying $\Lambda' \supset \Lambda$ and $[\Lambda' : \Lambda] = p$ are:

- $\mathbb{Z}\frac{\omega_1 + b\omega_2}{p} + \mathbb{Z}\omega_2$ with $b = 0, 1 \ldots, p-1$
- $\mathbb{Z}\omega_1 + \mathbb{Z}\frac{\omega_2}{p}$,

and that these constitute $p+1$ distinct lattices.

(b) Provide all details to the claim made in the first sentence of the proof of Proposition 4.5.

Exercise 4.8. Let p be a prime and consider the lattice $\Lambda := \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$ where $\omega_1, \omega_2 \in \mathbb{C}^*$ and $\omega_1/\omega_2 \notin \mathbb{R}$.

(a) Show that there are exactly $p^2 + p + 1$ lattices $\Lambda' \subset \mathbb{C}$ satisfying $\Lambda' \supset \Lambda$ and $[\Lambda' : \Lambda] = p^2$, and give a list of these.

(b) Try to generalize part (a) (e.g. replace $[\Lambda' : \Lambda] = p^2$ by $[\Lambda' : \Lambda] = p^k$ with $k \in \mathbb{Z}_{>0}$).

Exercise 4.9. Calculate the matrix of the Hecke operator T_2 on the space $\mathrm{S}_{24}(\mathrm{SL}_2(\mathbb{Z}))$ with respect to a basis of your choice. Show that the characteristic polynomial of T_2 is $x^2 - 1080x - 20468736$. (You may use a computer, but not a package in which this exercise can be solved with a one-line command.)

Exercise 4.10. Consider the formal (so we do not worry about convergence) generating function of the Hecke operators T_n on $M_k(\Gamma_1(N))$

$$g(s) := \sum_{n=1}^{\infty} T_n n^{-s}.$$

Deduce the following formal product expansion (over all primes p):

$$g(s) = \prod_p \left(\mathrm{id} - T_p p^{-s} + \langle p \rangle p^{k-1-2s}\right)^{-1}.$$

Exercise 4.11. Let $k, N \in \mathbb{Z}_{>0}$, and let χ be a Dirichlet character modulo N.

(a) For $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, denote by d_γ the lower-right entry of γ. Show that

$$M_k(N, \chi) = \{f \in M_k(\Gamma_1(N)) : f|_k\gamma = \chi(d_\gamma)f \text{ for all } \gamma \in \Gamma_0(N)\}$$

and

$$S_k(N, \chi) = \{f \in S_k(\Gamma_1(N)) : f|_k\gamma = \chi(d_\gamma)f \text{ for all } \gamma \in \Gamma_0(N)\}.$$

(b) Let 1_N denote the trivial character modulo N. Show that

$$M_k(N, 1_N) = M_k(\Gamma_0(N)) \quad \text{and} \quad S_k(N, 1_N) = S_k(\Gamma_0(N)).$$

Exercise 4.12. Let $k \in \mathbb{Z}_{>0}$, let $f \in \mathrm{M}_k(\mathrm{SL}_2(\mathbb{Z}))$ be an eigenform, normalised such that $a_1(f) = 1$, and let p be a prime number. Let $\alpha, \beta \in \mathbb{C}$ be the roots of the polynomial $t^2 - a_p(f)t + p^{k-1}$. *Note:* You may use without proof that $a_p(f)$ is real. This fact will be proved in Exercise 5.1.

(a) Prove the formula

$$a_{p^r}(f) = \sum_{j=0}^{r} \alpha^j \beta^{r-j} \quad \text{for all } r \geq 0.$$

(b) Show that the following conditions are equivalent: (1) $|a_p(f)| \leq 2p^{(k-1)/2}$; (2) α and β are complex conjugates of absolute value $p^{(k-1)/2}$.

(c) Show that if the equivalent conditions of part (b) hold for all prime numbers p, then the q-expansion coefficients of f satisfy the bound

$$|a_n(f)| \leq \sigma_0(n) n^{(k-1)/2} \quad \text{for all } n \geq 1,$$

where $\sigma_0(n)$ is the number of (positive) divisors of n.

Note: If f is a cusp form, then the conditions of part (b) do in fact hold. This follows from two very deep theorems proved by P. Deligne in 1968 and 1974.

Exercise 4.13. Play around with the functions in SageMath for some of your favorite choices of congruence subgroup, modular forms space, etc.

Exercise 4.14. In this exercise you are supposed to make (partly) use of SageMath. Please attach your code when handing in the exercise, preferably by using the print button on the top right in the SageMath worksheet with which you can generate a PDF file. (Any comments can also be written down in the SageMath worksheet.)

(a) Compute a basis for $S_2(\Gamma_0(26))$.

(b) Find a basis B for $S_2(\Gamma_0(26))$ such that all the basis elements are eigenvectors for the Hecke operator T_2.

(c) Check that all the basis elements in B are eigenvectors for the Hecke operators T_n with $1 \leq n \leq 101$.

Note: In fact, these basis elements are Hecke eigenforms.

(d) The following equation defines a curve in the plane:

$$E_1 \colon y^2 + xy + y = x^3. \tag{4.2}$$

(It is a so-called affine equation for an elliptic curve.) Tell SageMath about this object by typing `E1=EllipticCurve([1,0,1,0,0])` .
For a prime number p, the number of solutions to (4.2) with $x, y \in \mathbb{F}_p := \mathbb{Z}/p\mathbb{Z}$ *plus one* is denoted by $N_p(E_1)$, i.e.

$$N_p(E_1) = \#\{(x,y) \in \mathbb{F}_p^2 : y^2 + xy + y = x^3\} + 1.$$

(The +1 is there, because it is more natural to count solutions in the so-called *projective closure*, which boils down to one more solution 'at infinity'.) SageMath can compute these numbers by typing `E1.Np(prime)` where `prime` is some prime number (e.g. `prime=79`).

Find an explicit relation between $N_p(E_1)$ and $a_p(f)$ for one of the basis elements f in B and all primes $p < 1000$.

(e) Now consider

$$E_2 \colon y^2 + xy + y = x^3 - x^2 - 3x + 3,$$

which is given in SageMath by `E2=EllipticCurve([1,-1,1,-3,3])` .
Similarly as in (d), we set

$$N_p(E_2) = \#\{(x,y) \in \mathbb{F}_p^2 : y^2 + xy + y = x^3 - x^2 - 3x + 3\} + 1$$

and this can be computed in SageMath by typing `E2.Np(prime)` .

Find an explicit relation between $N_p(E_2)$ and $a_p(g)$ for one of the basis elements g in B and all primes $p < 1000$.

Note: This illustrates the *modularity* of the elliptic curves E_1 and E_2.

Chapter 5

The theory of newforms

5.1 The Petersson inner product

Let Γ be a congruence subgroup and $k \in \mathbb{Z}$. We have constructed the finite-dimensional $\mathbb{C}$-vector space $\mathrm{M}_k(\Gamma)$ of modular forms and its subspace $\mathrm{S}_k(\Gamma)$ of cusp forms. If Γ is of the form $\Gamma_1(N)$, we have also constructed commutative $\mathbb{C}$-algebras $\mathbb{T}(\mathrm{M}_k(\Gamma))$ and $\mathbb{T}(\mathrm{S}_k(\Gamma))$ acting on these spaces. We will now define an additional structure, namely a Hermitean inner product on $\mathrm{S}_k(\Gamma_1(N))$.

Lemma 5.1. *Let U be a subset of $\mathbb{H}$ whose boundary consists of finitely many line segments and circle arcs. Let $f\colon U \to \mathbb{C}$ be a continuous function, and let $\gamma \in \mathrm{SL}_2(\mathbb{R})$. Then we have*

$$\int_{z\in U} f(z)\frac{dx\,dy}{y^2} = \int_{z\in\gamma^{-1}U} f(\gamma z)\frac{dx\,dy}{y^2} \quad (z = x+iy).$$

Proof. We view $\mathbb{H}$ as an open subset of $\mathbb{R}^2$ with coordinates (x,y) and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ as a real differentiable map $\mathbb{H} \to \mathbb{H}$. We write

$$\gamma_1(x,y) = \Re\gamma(x+iy), \quad \gamma_2(x,y) = \Im\gamma(x+iy).$$

The Jacobian matrix of this map at a point $z = x+iy$ is

$$J_\gamma(x,y) = \begin{pmatrix} \partial\gamma_1/\partial x & \partial\gamma_1/\partial y \\ \partial\gamma_2/\partial x & \partial\gamma_2/\partial y \end{pmatrix}.$$

Since γ is holomorphic, it satisfies the Cauchy–Riemann equations

$$\frac{\partial\gamma_2}{\partial y} = \frac{\partial\gamma_1}{\partial x}, \quad \frac{\partial\gamma_2}{\partial x} = -\frac{\partial\gamma_1}{\partial y},$$

and its derivative can be expressed as

$$\gamma'(z) = \frac{\partial\gamma_1}{\partial x} + i\frac{\partial\gamma_2}{\partial x}.$$

Therefore we have

$$\begin{aligned} \det J_\gamma(x,y) &= \frac{\partial\gamma_1}{\partial x}\frac{\partial\gamma_2}{\partial y} - \frac{\partial\gamma_1}{\partial y}\frac{\partial\gamma_2}{\partial x} \\ &= \left(\frac{\partial\gamma_1}{\partial x}\right)^2 + \left(\frac{\partial\gamma_2}{\partial x}\right)^2 \\ &= |\gamma'(x+iy)|^2. \end{aligned}$$

On the other hand, by Proposition 1.2 part (i), we have

$$\Im(\gamma z) = \frac{\Im z}{|cz+d|^2}.$$

Furthermore, one computes

$$\gamma'(z) = \frac{1}{(cz+d)^2},$$

so that

$$|\det J_\gamma(z)| = |\gamma'(z)|^2 = \frac{\Im(\gamma z)^2}{(\Im z)^2}.$$

This implies

$$\begin{aligned}\int_{z\in U} f(z)\frac{dx\,dy}{y^2} &= \int_{z\in\gamma^{-1}U} f(\gamma z)|\det J_\gamma(z)|\frac{dx\,dy}{(\Im\gamma z)^2}\\ &= \int_{z\in\gamma^{-1}U} f(\gamma z)\frac{dx\,dy}{y^2}.\end{aligned}$$

This proves the lemma. □

Remark. In the language of differential forms, we have proved that the differential 2-form $\frac{dx\wedge dy}{y^2}$ is $\mathrm{SL}_2(\mathbb{R})$-invariant.

Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. We recall the following notation (see Proposition 3.2). Let R be a system of representatives for the quotient $\Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$. We write

$$\mathcal{D}_\Gamma = \bigcup_{\gamma\in R}\gamma\mathcal{D},$$

where $\mathcal{D}$ is the standard fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$. Note that $\mathcal{D}_\Gamma$ depends on the choice of R.

Let $F\colon \mathbb{H}\to\mathbb{C}$ be a continuous function that is Γ-*invariant* in the sense that

$$F(\gamma z) = F(z)\quad\text{for all } \gamma\in\Gamma, z\in\mathbb{H}.$$

We consider the integral

$$\int_{z\in\mathcal{D}_\Gamma} F(z)\frac{dx\,dy}{y^2}. \tag{5.1}$$

By Lemma 5.1, the value of this integral does not depend on the choice of the system of representatives R.

In the next two results, we will consider subsets of $\mathcal{D}_\Gamma$ that can be considered as "regions around the cusps" and are defined as follows. For $Y>0$, let U_Y be the following subset of $\mathbb{H}$:

$$U_Y = \{z = x+iy \mid -1/2\le x\le 1/2 \text{ and } y\ge Y\}$$

Then $\mathcal{D}_\Gamma$ is the union of some compact set $K\subset\mathbb{H}$ and the sets

$$\gamma U_Y = \{\gamma z \mid \gamma\in U_Y\}$$

for $\gamma\in R$.

Lemma 5.2. *Suppose that for all $\gamma\in\mathrm{SL}_2(\mathbb{Z})$ there exist real numbers $c_\gamma>0$ and $e_\gamma<1$ such that*

$$|F(\gamma z)|\le c_\gamma(\Im z)^{e_\gamma}\quad \textit{for } \Im z \textit{ sufficiently large.}$$

Then the integral (5.1) *converges.*

Proof. The integral (5.1) restricted to K converges because K is compact. It therefore remains to show that the integral converges on each of the sets γU_Y for $\gamma \in R$. By Lemma 5.1, we have

$$\begin{aligned}\left|\int_{z\in\gamma U_Y} F(z)\frac{dx\,dy}{y^2}\right| &= \left|\int_{z\in U_Y} F(\gamma z)\frac{dx\,dy}{y^2}\right| \\ &\le c_\gamma \int_{z\in U_Y} y^{e_\gamma}\frac{dx\,dy}{y^2} \\ &= c_\gamma \int_{y=Y}^{\infty} y^{e_\gamma-2}dy.\end{aligned}$$

This converges since $e_\gamma - 2 < -1$ by assumption. □

Let k be an integer. Let f, g be two modular forms of weight k for our congruence subgroup Γ. We apply the results above to the continuous (but in general non-holomorphic) function

$$F(z) = f(z)\overline{g(z)}(\Im z)^k.$$

Lemma 5.3. *The function $F(z)$ is Γ-invariant. If $k = 0$ or if for each cusp $\mathfrak{c}$ at least one of the forms f and g vanishes at $\mathfrak{c}$, then F is bounded.*

Proof. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Gamma$. By the modularity of f and g and by Proposition 1.2 part (i), we have

$$\begin{aligned} f(\gamma z) &= (cz+d)^k f(z), \\ \overline{g(\gamma z)} &= (\overline{cz+d})^k\overline{g(z)}, \\ (\Im\gamma z)^k &= |cz+d|^{-2k}(\Im z)^k.\end{aligned}$$

Multiplying these three equations yields $F(\gamma z) = F(z)$, so F is Γ-invariant.

Since F is Γ-invariant, proving that F is bounded means proving that it is bounded on $\mathcal{D}_\Gamma$. As in the proof of Lemma 5.2, we consider the regions around the γU_Y around the cusps. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ and let $\mathfrak{c}$ be the corresponding cusp, i.e. the Γ-orbit of $\gamma\infty \in \mathbb{P}^1(\mathbb{Q})$ in $\mathrm{Cusps}(\Gamma) = \Gamma\backslash\mathbb{P}^1(\mathbb{Q})$. Then by the definition of the q-expansion of f at $\mathfrak{c}$, we have

$$\begin{aligned}(cz+d)^{-k}f(\gamma z) &= (f|_k\gamma)(z) \\ &= \sum_{n=0}^{\infty} a_{n,\mathfrak{c}}(f)\exp(2\pi i n z/h_{\mathfrak{c}}),\end{aligned}$$

where $h_{\mathfrak{c}}$ is the width of the cusp $\mathfrak{c}$, and similarly for g. Therefore

$$\begin{aligned}F(\gamma z) &= f(\gamma z)\overline{g(\gamma z)}(\Im\gamma z)^k \\ &= (f|_k\gamma)(z)(cz+d)^k\overline{(g|_k\gamma)(z)}(\overline{cz+d})^k|cz+d|^{-2k}(\Im z)^k \\ &= \left(\sum_{n=0}^{\infty} a_{n,\mathfrak{c}}(f)\exp(2\pi i n z/h_{\mathfrak{c}})\right)\overline{\left(\sum_{n=0}^{\infty} a_{n,\mathfrak{c}}(g)\exp(2\pi i n z/h_{\mathfrak{c}})\right)}(\Im z)^k.\end{aligned}$$

If $a_{0,\mathfrak{c}}(f) = 0$ or $a_{0,\mathfrak{c}}(g) = 0$, then the absolute value of the expression above is bounded by a constant multiple of $|\exp(2\pi i z/h_{\mathfrak{c}})|y^k$ for $y \ge Y$. In particular, $F(\gamma z)$ is bounded on γU_Y for any $Y > 0$. Since the complement of all the γU_Y in $\mathcal{D}_\Gamma$ is a compact subset of $\mathbb{H}$, the function F is bounded on all of $\mathcal{D}_\Gamma$. □

For $f, g \in \mathrm{M}_k(\Gamma)$, we define

$$\langle f, g\rangle_\Gamma = \int_{z\in\mathcal{D}_\Gamma} f(z)\overline{g(z)}y^k\frac{dx\,dy}{y^2}, \tag{5.2}$$

where as usual we write the complex coordinate z in terms of the real coordinates x and y as $z = x+iy$. This is independent of the choice of the set of representatives used to define $\mathcal{D}_\Gamma$; however, to ensure that the integral converges, we need additional conditions. Combining Lemma 5.2 and Lemma 5.3, we observe that the integral above does converge whenever at least one of f and g is a cusp form. In particular, it gives rise to a well-defined map

$$\mathrm{M}_k(\Gamma) \times \mathrm{S}_k(\Gamma) \to \mathbb{C}.$$

Furthermore, it makes sense to introduce the following definition.

Definition. Let Γ be a congruence subgroup, and let k be an integer. The *Petersson inner product* on the $\mathbb{C}$-vector space $\mathrm{S}_k(\Gamma)$ is the Hermitean inner product $\langle\ ,\ \rangle_\Gamma$ defined by (5.2).

Unfortunately, the Petersson inner product on $\mathrm{S}_k(\Gamma)$ does not extend to an inner product on the whole space $\mathrm{M}_k(\Gamma)$, since $\langle f, f\rangle_\Gamma$ diverges for every $f \in \mathrm{M}_k(\Gamma)$ that is not a cusp form.

Definition. Let Γ be a congruence subgroup, and let k be an integer. The *Eisenstein subspace* (or *space of Eisenstein series*) in $\mathrm{M}_k(\Gamma)$, denoted by $\mathcal{E}_k(\Gamma)$, is the space

$$\mathcal{E}_k(\Gamma) = \bigl\{f \in \mathrm{M}_k(\Gamma) \bigm| \langle f, g\rangle_\Gamma = 0 \text{ for all } g \in \mathrm{S}_k(\Gamma)\bigr\}.$$

The Eisenstein subspace can be regarded as the “orthogonal complement” of $\mathrm{S}_k(\Gamma)$ with respect to $\langle\ ,\ \rangle_\Gamma$, even though $\langle\ ,\ \rangle_\Gamma$ does not define an inner product on all of $\mathrm{M}_k(\Gamma)$.

5.2 The adjoints of the Hecke operators

We would like to apply the spectral theorem from linear algebra (Theorem A.9) to the spaces of cusp forms $\mathrm{S}_k(\Gamma_1(N))$, equipped with the Petersson inner product, to obtain decompositions of these spaces into smaller spaces. We will show in this section that the diamond operators $\langle d\rangle$ are normal, as are the Hecke operators T_m with $\gcd(m, N) = 1$. For general T_m, we will need a more sophisticated theory, which we will develop in §5.3.

To compute the adjoints of the various operators in the Hecke algebra $\mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$, we begin by looking at a general congruence subgroup Γ. Let α be an element of $\mathrm{GL}_2^+(\mathbb{Q})$, and let T_α be the operator defined in §4.1. We recall that T_α is defined by

$$T_\alpha f = \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} f|_k\alpha\gamma,$$

where Γ_α is the subgroup $\Gamma \cap \alpha^{-1}\Gamma\alpha$ of Γ. (Earlier we denoted Γ_α by Γ', but below we will need to distinguish T_α for different α.)

Notation. For $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, we write

$$\alpha^* = (\det\alpha)\alpha^{-1} \in \mathrm{GL}_2^+(\mathbb{Q}).$$

More concretely, if $\alpha = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$, then $\alpha^* = \left(\begin{smallmatrix} d & -b \\ -c & a\end{smallmatrix}\right)$.

Notation. For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$ and $z \in \mathbb{H}$, we write

$$j(\gamma, z) = cz + d. \tag{5.3}$$

With this notation, we have

$$\begin{aligned}(f|_k\gamma)(z) &= \frac{(\det\gamma)^k}{j(\gamma,z)^k} f(\gamma z),\\ \Im(\gamma z) &= \frac{\det\gamma}{|j(\gamma,z)|^2}\Im z,\\ \frac{d}{dz}(\gamma z) &= \frac{\det\gamma}{j(\gamma,z)^2}.\end{aligned}$$

Lemma 5.4. *For all $\gamma, \delta \in \mathrm{GL}_2^+(\mathbb{Q})$ and all $z \in \mathbb{H}$, we have*

$$\begin{aligned} j(\gamma\delta, z) &= j(\gamma, \delta z) \cdot j(\delta, z), \\ j(\gamma, z) \cdot j(\gamma^{-1}, \gamma z) &= 1, \\ j(\mathrm{id}, z) &= 1. \end{aligned}$$

Proof. We prove the first identity; the third identity is immediate and the second easily follows from the other two.

We write $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $\delta = \left(\begin{smallmatrix} e & f \\ g & h \end{smallmatrix}\right)$. Then the bottom row of $\gamma\delta$ is given by

$$\gamma\delta = \begin{pmatrix} * & * \\ ce + dg & cf + dh \end{pmatrix}.$$

This implies

$$\begin{aligned} j(\gamma\delta, z) &= (ce + dg)z + (cf + dh) \\ &= c(ez + f) + d(gz + h) \\ &= \left(c\frac{ez + f}{gz + h} + d\right)(gz + h) \\ &= j(\gamma, \delta z)j(\delta, z). \end{aligned}$$

This is what we had to prove. □

Remark. The first equation in Lemma 5.4 is called the *cocycle condition.* Similar equations occur in other contexts involving group actions.

Proposition 5.5. *Let Γ be a congruence subgroup, and let k be an integer. Then for all $f, g \in \mathrm{M}_k(\Gamma)$ such that at least one of f and g is a cusp form, and for all $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, we have*

$$\langle T_\alpha f, g\rangle_\Gamma = \langle f, T_{\alpha^*} g\rangle_\Gamma.$$

Corollary 5.6. *With the notation above, the adjoint of the operator T_α on $\mathrm{S}_k(\Gamma)$ with respect to the inner product $\langle\ ,\ \rangle_\Gamma$ is the operator T_{α^*}.*

Proof of Proposition 5.5. We note that

$$T_\alpha f = \sum_{\gamma \in \Gamma_\alpha \backslash \Gamma} \frac{(\det \alpha\gamma)^k}{j(\alpha\gamma, z)^k} f(\alpha\gamma z)$$

This implies that we can compute $\langle T_\alpha f, g\rangle$ as

$$\langle T_\alpha f, g\rangle_\Gamma = \int_{z \in \mathcal{D}_\Gamma} \sum_{\gamma \in \Gamma_\alpha \backslash \Gamma} \frac{(\det \alpha\gamma)^k}{j(\alpha\gamma, z)^k} f(\alpha\gamma z)\overline{g(z)}(\Im z)^k \frac{dx\,dy}{y^2}.$$

The next step is to use the identities

$$\begin{aligned} \det(\alpha\gamma) &= \det(\alpha), \\ j(\alpha\gamma, z)^{-k} &= j(\alpha, \gamma z)^{-k} j(\gamma, z)^{-k}, \\ g(z) &= j(\gamma, z)^{-k} g(\gamma z), \\ (\Im z)^k &= |j(\gamma, z)|^{2k} (\Im \gamma z)^k; \end{aligned}$$

here we have used that $\det\gamma = 1$ (because $\gamma \in \Gamma \subset \mathrm{SL}_2(\mathbb{Z})$) and that g is modular of weight k for Γ. Substituting these identities in the equation above for $\langle T_\alpha f, g\rangle$ gives

$$\langle T_\alpha f, g\rangle_\Gamma = (\det\alpha)^k \int_{z\in\mathcal{D}_\Gamma} \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} j(\alpha,\gamma z)^{-k} f(\alpha\gamma z)\overline{g(\gamma z)}(\Im\gamma z)^k \frac{dx\,dy}{y^2}.$$

The expression that is being summed and integrated is a function of γz. Instead of integrating over $z \in \mathcal{D}_\Gamma$ and summing over $\gamma \in \Gamma_\alpha\backslash\Gamma$, we can integrate directly over $z \in \mathcal{D}_{\Gamma_\alpha}$. This gives

$$\langle T_\alpha f, g\rangle_\Gamma = (\det\alpha)^k \int_{z\in\mathcal{D}_{\Gamma_\alpha}} j(\alpha,z)^{-k} f(\alpha z)\overline{g(z)}(\Im z)^k \frac{dx\,dy}{y^2.}$$

We need to prove that this is equal to $\langle f, T_{\alpha^*} g\rangle_\Gamma$. We note that

$$T_{\alpha^*} g = (\det\alpha)^k T_{\alpha^{-1}} g.$$

We now apply the formula for $\langle T_\alpha f, g\rangle_\Gamma$ proved above to $\langle T_{\alpha^{-1}} g, f\rangle_\Gamma$. This gives

$$\begin{aligned}
\langle f, T_{\alpha^*} g\rangle_\Gamma &= (\det\alpha)^k \langle f, T_{\alpha^{-1}} g\rangle_\Gamma \\
&= (\det\alpha)^k \overline{\langle T_{\alpha^{-1}} g, f\rangle_\Gamma} \\
&= (\det\alpha)^k (\det\alpha^{-1})^k \overline{\int_{z\in\mathcal{D}_{\Gamma_{\alpha^{-1}}}} j(\alpha^{-1},z)^{-k} g(\alpha^{-1}z)\overline{f(z)}(\Im z)^k \frac{dx\,dy}{y^2}} \\
&= \int_{z\in\mathcal{D}_{\Gamma_{\alpha^{-1}}}} \overline{j(\alpha^{-1},z)^{-k}} f(z)\overline{g(\alpha^{-1}z)}(\Im z)^k \frac{dx\,dy}{y^2}.
\end{aligned}$$

We make the change of variables $z = \alpha w$. We note that

$$\begin{aligned}
\Gamma_{\alpha^{-1}} &= \Gamma \cap \alpha\Gamma\alpha^{-1} \\
&= \alpha\Gamma_\alpha\alpha^{-1}, \\
j(\alpha^{-1}, \alpha w) &= j(\alpha, w)^{-1}, \\
\Im(\alpha w) &= \frac{(\det\alpha)^k}{|j(\alpha,w)|^{2k}} (\Im w)^k.
\end{aligned}$$

Furthermore, letting z range over $\mathcal{D}_{\Gamma_{\alpha^{-1}}}$ has the effect of letting w range over some fundamental domain for Γ_α. Putting all of this together, we get

$$\langle f, T_{\alpha^*} g\rangle_\Gamma = (\det\alpha)^k \int_{w\in\mathcal{D}_{\Gamma_\alpha}} j(\alpha,w)^{-k} f(\alpha w)\overline{g(w)}(\Im w)^k \frac{dx'\,dy'}{y'^2} \quad (w = x' + iy').$$

This is the same as the expression we found above for $\langle T_\alpha f, g\rangle_\Gamma$, so the claim is proved. □

Remark. At the expense of slightly more abstraction and proving some more general facts first, one can reduce the calculation in the proof above to

$$\begin{aligned}
\langle T_\alpha f, g\rangle_\Gamma &= \langle f|_k\alpha, g\rangle_{\Gamma_\alpha} \\
&= \langle f, g|_k\alpha^*\rangle_{\Gamma_{\alpha^*}} \\
&= \langle f, T_{\alpha^*} g\rangle.
\end{aligned}$$

Corollary 5.7. *Let Γ be a congruence subgroup, and let k be an integer. Then all operators T_α on $\mathrm{M}_k(\Gamma)$, for $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$, preserve the Eisenstein subspace $\mathcal{E}_k(\Gamma)$ of $\mathrm{M}_k(\Gamma)$.*

Proof. Let $f \in \mathcal{E}_k(\Gamma)$, and let $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$. Then for all $g \in \mathrm{S}_k(\Gamma)$, Proposition 5.5 implies

$$\langle T_\alpha f, g\rangle_\Gamma = \langle f, T_{\alpha^*} g\rangle_\Gamma = 0,$$

since $T_{\alpha^*} g$ is still in $\mathrm{S}_k(\Gamma)$. Therefore $T_\alpha f$ is orthogonal to $\mathrm{S}_k(\Gamma)$, and hence is in $\mathcal{E}_k(\Gamma)$. □

The following lemma will be used to prove Proposition 5.9 below.

Lemma 5.8. *Let Γ be a congruence subgroup, and let k be an integer. Let $\alpha, \beta \in \mathrm{GL}_2^+(\mathbb{Q})$ be such that at least one of them normalises Γ. Then we have*

$$T_{\alpha\beta} = T_\beta T_\alpha \quad \textit{and} \quad T_{\beta\alpha} = T_\alpha T_\beta$$

as operators on $\mathrm{M}_k(\Gamma)$.

Proof. By symmetry, we may assume that β normalises Γ. Then we have

$$\Gamma_\beta = \Gamma \cap \beta^{-1}\Gamma\beta = \Gamma$$

and

$$\begin{aligned}\Gamma_{\beta\alpha} &= \Gamma \cap \alpha^{-1}\beta^{-1}\Gamma\beta\alpha \\ &= \Gamma \cap \alpha^{-1}\Gamma\alpha \\ &= \Gamma_\alpha.\end{aligned}$$

Furthermore, conjugation by β gives isomorphisms

$$\begin{aligned}\Gamma &\xrightarrow{\sim} \Gamma \\ \alpha^{-1}\Gamma\alpha &\xrightarrow{\sim} \beta^{-1}\alpha^{-1}\Gamma\alpha\beta \\ \Gamma_\alpha &\xrightarrow{\sim} \Gamma_{\alpha\beta}\end{aligned}$$

where all maps are defined by $\gamma \mapsto \beta^{-1}\gamma\beta$ and the last isomorphism is obtained from the first two by taking intersections.

Let f be an element of $\mathrm{M}_k(\Gamma)$. We have

$$\begin{aligned}T_\alpha f &= \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} f|_k\alpha\gamma, \\ T_\beta f &= f|_k\beta, \\ T_{\beta\alpha} f &= \sum_{\gamma\in\Gamma_{\beta\alpha}\backslash\Gamma} f|_k\beta\alpha\gamma \\ &= \sum_{\gamma\in\Gamma_\alpha\backslash\Gamma} (f|_k\beta)|_k\alpha\gamma \\ &= T_\alpha(T_\beta f), \\ T_{\alpha\beta} f &= \sum_{\gamma\in\Gamma_{\alpha\beta}\backslash\Gamma} f|_k\alpha\beta\gamma \\ &= \sum_{\delta\in\Gamma_\alpha\backslash\Gamma} f|_k\alpha\beta(\beta^{-1}\delta\beta) \\ &= \sum_{\delta\in\Gamma_\alpha\backslash\Gamma} f|_k\alpha\delta\beta \\ &= \sum_{\delta\in\Gamma_\alpha\backslash\Gamma} (f|_k\alpha\delta)|_k\beta \\ &= T_\beta(T_\alpha f),\end{aligned}$$

where $\delta = \beta\gamma\beta^{-1}$. This proves that $T_{\beta\alpha} = T_\alpha T_\beta$ and $T_{\alpha\beta} = T_\beta T_\alpha$, as claimed. □

Remark. The assumption that α or β normalises Γ in Lemma 5.8 is necessary. For instance, take $\Gamma = \Gamma_1(N)$, let p a prime number not dividing N, and consider $\alpha = \left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)$ and $\beta = \left(\begin{smallmatrix}p & 0\\ 0 & 1\end{smallmatrix}\right)$. If the lemma would hold for this choice of α and β, i.e.

$$T_{\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)} T_{\left(\begin{smallmatrix}p & 0\\ 0 & 1\end{smallmatrix}\right)} \overset{?}{=} T_{\left(\begin{smallmatrix}p & 0\\ 0 & p\end{smallmatrix}\right)},$$

then Proposition 5.9 would imply

$$p^2\langle p\rangle^{-1}T_pT_p \overset{?}{=} p^k\mathrm{id},$$

which is in general false, for example because T_p is not invertible in general.

We now apply Proposition 5.5 to the congruence groups $\Gamma_1(N)$ and to the matrices $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ defining the diamond and Hecke operators on $\mathrm{M}_k(\Gamma_1(N))$.

Proposition 5.9. *Let $N \geq 1$, and let k be an integer. In the Hecke algebra $\mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$, consider the diamond operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and the Hecke operators T_m for $m \geq 1$ with* $\gcd(m, N) = 1$*. The adjoints of these operators with respect to the Petersson inner product are*

$$\begin{aligned} \langle d\rangle^\dagger &= \langle d\rangle^{-1} && \textit{for all } d \in (\mathbb{Z}/N\mathbb{Z})^\times, \\ T_m^\dagger &= \langle m\rangle^{-1}T_m && \textit{if } \gcd(m, N) = 1. \end{aligned}$$

Proof. We first prove the formula for the diamond operators. Let $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be an element of $\Gamma_0(N)$, so that T_α is the diamond operator $\langle d\rangle$. We have

$$\alpha^* = \alpha^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \in \Gamma_0(N),$$

and this matrix defines the operator $\langle a\rangle$. From the fact that N divides c it follows that

$$1 = \det\alpha = ad - bc \equiv ad \bmod N,$$

so $\langle a\rangle = \langle d\rangle^{-1}$. We conclude that the adjoint of $\langle d\rangle$ is $\langle d\rangle^{-1}$.

To prove the formula for the Hecke operators, we start with the T_p for p a prime number not dividing N. By Proposition 5.5, we have

$$T_p = \frac{1}{p}T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)}, \quad T_p^\dagger = \frac{1}{p}T_{\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)}.$$

We therefore have to prove the identity

$$T_{\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)} = \langle p\rangle^{-1}T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)}.$$

By the Chinese remainder theorem and the fact that $\gcd(p, N) = 1$, we can choose $d \in \mathbb{Z}$ such that

$$\begin{cases} d \equiv 1 \pmod{N}, \\ d \equiv 0 \pmod{p}. \end{cases}$$

Then we have $\gcd(d, N) = 1$, so we can choose $a, b \in \mathbb{Z}$ such that $ad - bN = 1$. Since $d \equiv 1 \pmod{N}$, the matrix $\left(\begin{smallmatrix} a & b \\ N & d \end{smallmatrix}\right)$ is in $\Gamma_1(N)$. This implies $T_{\left(\begin{smallmatrix} a & b \\ N & d \end{smallmatrix}\right)} = \mathrm{id}$. We note that

$$\begin{pmatrix} a & b \\ N & d \end{pmatrix}\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} ap & b \\ Np & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}\begin{pmatrix} ap & b \\ N & d/p \end{pmatrix},$$

and since $d \equiv 0 \pmod{p}$, the matrix $\left(\begin{smallmatrix} ap & b \\ N & d/p \end{smallmatrix}\right)$ is in $\Gamma_0(N)$. By Lemma 5.8, we therefore get

$$\begin{aligned} T_{\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)} &= T_{\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)}T_{\left(\begin{smallmatrix} a & b \\ N & d \end{smallmatrix}\right)} \\ &= T_{\left(\begin{smallmatrix} a & b \\ N & d \end{smallmatrix}\right)\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)} \\ &= T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)\left(\begin{smallmatrix} ap & b \\ N & d/p \end{smallmatrix}\right)} \\ &= T_{\left(\begin{smallmatrix} ap & b \\ N & d/p \end{smallmatrix}\right)}T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)}. \end{aligned}$$

Again using $d \equiv 1 \pmod{N}$, we obtain

$$T_{\left(\begin{smallmatrix} ap & b \\ N & d/p \end{smallmatrix}\right)} = \langle d/p \rangle = \langle p \rangle^{-1}.$$

We conclude that $T_{\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)} = \langle p \rangle^{-1} T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)}$, as claimed.

To prove the claim that $T_m^\dagger = \langle m \rangle^{-1} T_m$ in the case where m is a power of a prime number $p \nmid N$, say $m = p^r$, we use induction on r. The claim is trivially true for $r = 0$, and we proved it above for $r = 1$. Let $r \geq 2$, and assume that the claim holds for $m = 1, p, \ldots, p^{r-1}$. We have by definition

$$T_{p^r} = T_p T_{p^{r-1}} - p^{k-1} \langle p \rangle T_{p^{r-2}}.$$

This implies

$$\begin{aligned} T_{p^r}^\dagger &= T_{p^{r-1}}^\dagger T_p^\dagger - p^{k-1} T_{p^{r-2}}^\dagger \langle p \rangle^\dagger \\ &= \langle p^{r-1} \rangle^{-1} T_{p^{r-1}} \langle p \rangle^{-1} T_p - p^{k-1} \langle p^{r-2} \rangle^{-1} T_{p^{r-2}} \langle p \rangle^{-1}. \end{aligned}$$

By the commutativity of the Hecke algebra, we can rewrite this as

$$\begin{aligned} T_{p^r} &= \langle p^r \rangle^{-1} T_p T_{p^{r-1}} - p^{k-1} \langle p^r \rangle^{-1} \langle p \rangle T_{p^{r-2}} \\ &= \langle p^r \rangle^{-1} T_{p^r}, \end{aligned}$$

which proves the claim for $m = p^r$.

Finally, for m, n coprime, we have

$$\begin{aligned} T_{mn}^\dagger &= T_n^\dagger T_m^\dagger \\ &= \langle n \rangle^{-1} T_n \langle m \rangle^{-1} T_m \\ &= \langle mn \rangle^{-1} T_{mn}. \end{aligned}$$

The claim for general m with $\gcd(m, N) = 1$ now follows by induction on the number of prime factors of N. □

Corollary 5.10. *The operators T_m for $\gcd(m, N) = 1$ and $\langle d \rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ form a commuting system of normal operators.*

Corollary 5.11. *The space $\mathrm{S}_k(\Gamma_1(N))$ admits a basis consisting of simultaneous eigenvectors for the operators T_m for $\gcd(m, N) = 1$ and $\langle d \rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$.*

Unfortunately, this result cannot be generalised to *all* Hecke operators T_m. For this reason, we will introduce the concept of *oldforms* and *newforms*.

5.3 Oldforms and newforms (Atkin–Lehner theory)

Recall that if $\Gamma' \subset \Gamma$ are congruence subgroups, then any modular form for Γ is also a modular form for Γ, so for every $k \in \mathbb{Z}$ we have an inclusion

$$\begin{aligned} \mathrm{M}_k(\Gamma) &\rightarrowtail \mathrm{M}_k(\Gamma') \\ f &\mapsto f. \end{aligned}$$

Recall also that for $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ and $\Gamma_\alpha = \Gamma \cap \alpha^{-1}\Gamma\alpha$, we have an inclusion

$$\begin{aligned} \mathrm{M}_k(\Gamma) &\to \mathrm{M}_k(\Gamma_\alpha) \\ \alpha &\mapsto f|_k\alpha. \end{aligned}$$

Now consider two positive integers M, N such that M divides N. Let e be a divisor of N/M. We take

$$\alpha = \begin{pmatrix} e & 0 \\ 0 & 1 \end{pmatrix}.$$

As a special case of Exercise 3.3(b), we have the inclusion

$$\Gamma_1(M) \cap \alpha^{-1}\Gamma_1(M)\alpha \supseteq \Gamma_1(eM) \supseteq \Gamma_1(N).$$

Now

$$(f|_k\alpha)(z) = e^k f(ez) \quad \text{for all } f \in \mathrm{M}_k(\Gamma_1(M)).$$

This implies that we have a well-defined map

$$\begin{aligned} i_e = i_e^{M,N} \colon \mathrm{M}_k(\Gamma_1(M)) &\longrightarrow \mathrm{M}_k(\Gamma_1(N)) \\ f &\longmapsto (z \mapsto f(ez)). \end{aligned}$$

These maps restrict to spaces of cusp forms.

Definition. Let $N \geq 1$ and $k \in \mathbb{Z}$. The space of *oldforms* in the space $\mathrm{S}_k(\Gamma_1(N))$ of cusp forms, denoted by $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}$, is the $\mathbb{C}$-linear subspace of $\mathrm{S}_k(\Gamma_1(N))$ spanned by the images of all the maps

$$i_e^{M,N} \colon \mathrm{S}_k(\Gamma_1(M)) \longrightarrow \mathrm{S}_k(\Gamma_1(N))$$

for all $M \mid N$ with $M \neq N$, and all $e \mid (N/M)$, i.e.

$$\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}} = \sum_{\substack{e|M|N \\ M\neq N}} i_e^{M,N}\bigl(\mathrm{S}_k(\Gamma_1(M))\bigr).$$

The space of *newforms* in $\mathrm{S}_k(\Gamma_1(N))$, denoted by $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$, is the orthogonal complement of $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}$ with respect to the Petersson inner product, i.e.

$$\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}} = \bigl\{f \in \mathrm{S}_k(\Gamma_1(N)) \bigm| \langle f, g\rangle_{\Gamma_1(N)} = 0 \quad \text{for all } g \in \mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}\bigr\}.$$

Note that every strict divisor $M \mid N$ is also a divisor of N/l for some prime divisor l of N, and that each of the images of $i_e^{M,N}$ with $e \mid (N/M)$ is contained either in the image of $i_1^{N/l,N}$ or in the image of $i_l^{N/l,N}$ for some prime number $l \mid N$. This means that we could have used the equivalent definition

$$\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}} = \sum_{\substack{l \text{ prime} \\ l|N}} \mathrm{S}_k(\Gamma_1(N))_{l\text{-old}},$$

where

$$\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}} = i_1^{N/l,N}\bigl(\mathrm{S}_k(\Gamma_1(N/l))\bigr) + i_l^{N/l,N}\bigl(\mathrm{S}_k(\Gamma_1(N/l))\bigr).$$

Analogously, for every prime divisor l of N, we can define $\mathrm{S}_k(\Gamma_1(N))_{l\text{-new}}$ as the orthogonal complement of $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$. Then we have

$$\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}} = \bigcap_{\substack{l \text{ prime} \\ l|N}} \mathrm{S}_k(\Gamma_1(N))_{l\text{-new}}.$$

Proposition 5.12. *All the subspaces of* $\mathrm{S}_k(\Gamma_1(N))$ *defined above are stable under the action of the Hecke algebra* $\mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$.

Proof. We will prove below that for every prime divisor $l \mid N$, the subspace $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}} \subset \mathrm{S}_k(\Gamma_1(N))$ is stable under the diamond operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$, the Hecke operators T_p for p prime, and the *adjoints* of all these operators. It will then follow that the whole old space $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}$ is stable under these operators. Furthermore, if T is in $\mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$ and $V \subset \mathrm{S}_k(\Gamma_1(N))$ is a subspace that is stable under T, then the orthogonal complement of V is stable under the adjoint of T

Let l be a prime divisor of N, and write $i_1 = i_1^{N/l,N}$ and $i_l = i_l^{N/l,N}$. We have to prove that for all operators T in the list above and all $f \in \mathrm{S}_k(\Gamma_1(N/l))$, the forms $T(i_1 f)$ and $T(i_l f)$ in $\mathrm{S}_k(\Gamma_1(N))$ are actually in $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$.

Consider a matrix

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N).$$

The map $f \mapsto f|_k\alpha$ defines the diamond operator $\langle d\rangle$ on both $\mathrm{S}_k(\Gamma_1(N))$ and $\mathrm{S}_k(\Gamma_1(N/l))$. We have

$$\langle d\rangle(i_1 f) = (i_1 f)|_k\alpha = f|_k\alpha = \langle d\rangle f = i_1(\langle d\rangle f).$$

Next, we note that

$$\begin{pmatrix} l & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & bl \\ c/l & d \end{pmatrix}\begin{pmatrix} l & 0 \\ 0 & 1 \end{pmatrix}.$$

Since c/l is an integer divisible by N/l, the matrix $\left(\begin{smallmatrix} a & bl \\ c/l & d \end{smallmatrix}\right)$ defines the operator $\langle d\rangle$ on $\mathrm{S}_k(\Gamma_1(N/l))$. We apply this to f and recall that $f|_k\left(\begin{smallmatrix} l & 0 \\ 0 & 1 \end{smallmatrix}\right) = l^k i_l(f)$. This gives

$$\langle d\rangle(i_l f) = i_l(\langle d\rangle f).$$

In particular, $\langle d\rangle(i_1 f)$ and $\langle d\rangle(i_l f)$ are again in $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$. Thus $\langle d\rangle$ preserves the subspace $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$ of $\mathrm{S}_k(\Gamma_1(N))$.

Now let p be a prime number not dividing N. Then the operator T_p is given by the same formula on $\mathrm{S}_k(\Gamma_1(N/l))$ as on $\mathrm{S}_k(\Gamma_1(N))$. From this, it follows immediately that

$$T_p(i_1 f) = i_1(T_p f).$$

It is also not hard to check that

$$T_p(i_l f) = i_l(T_p f).$$

Hence $T_p(i_1 f)$ and $T_p(i_2 f)$ are in $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$ and T_p preserves the space $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$. The same argument works if p does divide N but is not equal to l.

Next we consider the operator T_l. First suppose l divides N exactly once. Then l does not divide N/l, so the formula for the action of T_l on q-expansions is different on $\mathrm{S}_k(\Gamma_1(N/l))$ and $\mathrm{S}_k(\Gamma_1(N))$, respectively. In $\mathrm{S}_k(\Gamma_1(N))$, we have

$$T_l(i_1 f) = \sum_{n=1}^{\infty} a_{nl}(f)q^n$$

and

$$T_l(i_l f) = i_1 f.$$

In $\mathrm{S}_k(\Gamma_1(N/l))$, Theorem 4.4 gives

$$\begin{aligned} T_l f &= \sum_{n=1}^{\infty} a_{nl}(f)q^n + l^{k-1}\sum_{n=1}^{\infty} a_{n/l}(\langle l\rangle f)q^n \\ &= T_l(i_1 f) + l^{k-1}\sum_{n=1}^{\infty} a_n(\langle l\rangle f)q^{nl} \\ &= T_l(i_1 f) + l^{k-1} i_l(\langle l\rangle f). \end{aligned}$$

This implies

$$T_l(i_1 f) = i_1(T_l f) - l^{k-1} i_l(\langle l\rangle f).$$

In particular, $T_l(i_1 f)$ and $T_l(i_l f)$ are both in $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$.

If on the other hand $l^2 \mid N$, then the effect of T_l on both $\mathrm{S}_k(\Gamma_1(N/l))$ and $\mathrm{S}_k(\Gamma_1(N))$ is given by the formula

$$T_l f = \sum_{n=1}^{\infty} a_{nl}(f)q^n.$$

From this one deduces that

$$T_l(i_1 f) = i_1(T_l f) \quad \text{and} \quad T_l(i_l f) = i_1 f.$$

This shows that T_l preserves $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$.

By Proposition 5.9, the adjoints of the operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$, and of the T_p for $p \nmid N$ prime, are in the Hecke algebra $\mathbb{T}(\mathrm{S}_k(\Gamma_1(N)))$, and therefore they preserve the subspace $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$.

It remains to show that every prime number p dividing N, the adjoint $T_p^\dagger$ of T_p preserves $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$. Since there is no simple formula for $T_p^\dagger$, we introduce a new operator called the *Fricke operator* or *Atkin–Lehner operator*. Consider the matrix

$$\alpha_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix},$$

and let w_N be operator T_{α_N} on $\mathrm{S}_k(\Gamma_1(N))$.

For all $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2^+(\mathbb{Q})$, we have

$$\alpha_N \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha_N^{-1} = \begin{pmatrix} d & -c/N \\ -bN & a \end{pmatrix}. \tag{5.4}$$

In particular, α_N normalises $\Gamma_1(N)$. Because of this, we have

$$\begin{aligned}(w_N f)(z) &= (f|_k \alpha_N)(z) \\ &= \frac{(\det \alpha_N)^k}{(Nz)^k} f(\alpha_N z) \\ &= z^{-k} f(-1/(Nz)).\end{aligned}$$

Now let p be a prime number dividing N. Applying (5.4) to the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$ defining T_p, we get

$$\alpha_N \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \alpha_N^{-1} = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Again using the fact that α_N normalises $\Gamma_1(N)$, we see that

$$\begin{aligned}T_p^\dagger &= \frac{1}{p} T_{\left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)} \\ &= \frac{1}{p} T_{\alpha_N \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right) \alpha_N^{-1}} \\ &= \frac{1}{p} T_{\alpha_N^{-1}} T_{\left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)} T_{\alpha_N} \\ &= w_N^{-1} T_p w_N.\end{aligned}$$

We now compute the effect of w_N on $i_1 f$ as follows:

$$\begin{aligned}(w_N(i_1 f))(z) &= z^{-k}(i_1 f)\left(-\frac{1}{Nz}\right) \\ &= z^{-k} f\left(-\frac{1}{Nz}\right) \\ &= l^k (lz)^{-k} f\left(-\frac{1}{(N/l)lz}\right) \\ &= l^k (w_{N/l} f)(lz) \\ &= l^k (i_l(w_{N/l} f))(z).\end{aligned}$$

Similarly, the effect of w_N on $i_l f$ is

$$\begin{aligned}(w_N(i_l f))(z) &= z^{-k}(i_l f)\left(-\frac{1}{Nz}\right)\\ &= z^{-k} f\left(-\frac{l}{Nz}\right)\\ &= z^{-k} f\left(-\frac{1}{(N/l)z}\right)\\ &= (w_{N/l} f)(z)\\ &= (i_1(w_{N/l} f))(z).\end{aligned}$$

This implies that w_N preserves the space $\mathrm{S}_k(\Gamma_1(N))_{l\text{-old}}$, and therefore so does $T_p^\dagger = w_N^{-1} T_p w_N$ for every prime divisor p of N. This finishes the proof of the proposition. □

Recall that a *Hecke eigenform* is a modular form $f \in \mathrm{M}_k(\Gamma_1(N))$ which is an eigenvector for the diamond operators $\langle d\rangle$ with $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and for the Hecke operators T_n with $n \geq 1$. Recall also that if f is a Hecke eigenform, then after scaling f we may assume that f is *normalised*, i.e. satisfies $a_1(f) = 1$.

Definition. Let $N \geq 1$ and $k \in \mathbb{Z}$. A *primitive form* of weight k for $\Gamma_1(N)$ is a normalised eigenform in $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$.

Theorem 5.13 (Atkin, Lehner (1970); Li (1975))**.** *Let $N \geq 1$ and $k \in \mathbb{Z}$.*

(a) If $f \in \mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$ is an eigenform for the operators d with $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and the T_m with $\gcd(m, N) = 1$, then f is an eigenform for all T_m with $m \geq 1$, and f is a scalar multiple of a primitive form.

(b) If $f, g \in \mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$ are two eigenforms for the operators d with $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and the T_m with $\gcd(m, N) = 1$, with the same eigenvalues for these operators, then f, g are scalar multiples of the same primitive form.

(c) The set of primitive forms of weight k for $\Gamma_1(N)$ is an orthogonal basis for the $\mathbb{C}$-vector space $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$ with respect to the Petersson inner product.

We will give a partial proof in the sense that we will use Proposition 5.14 below, which we shall not prove. For a proof we refer to [4, Theorem 5.7.1]. That proof uses some representation theory, which is not considered a prerequisite for this course (and explaining this from scratch would take us too far afield). There exist more elementary proofs, but those are quite long and probably less insightful.

Proposition 5.14. *Let $f \in \mathrm{S}_k(\Gamma_1(N))$ be a form whose q-expansion coefficients $a_m(f)$ satisfy $a_m(f) = 0$ for all $m \geq 1$ such that $\gcd(m, N) = 1$. Then there exist forms $f_l \in \mathrm{S}_k(\Gamma_1(N/l))$, with l ranging over the prime divisors of N, such that*

$$f = \sum_{\substack{l \text{ prime}\\ l \mid N}} i_l^{N/l,N}(f_l).$$

Proof of Theorem 5.13. Let f be an eigenform for the operators d with $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and the T_m with $\gcd(m, N) = 1$. By definition, we have $f \neq 0$. We have seen in Proposition 4.7 that

$$a_1(T_m f) = a_m(f) \quad \text{for all } m \geq 1.$$

For all m with $\gcd(m, N) = 1$, by assumption there exists $\lambda_m \in \mathbb{C}$ such that

$$T_m f = \lambda_m f.$$

Combining the two equations above, we obtain

$$\lambda_m a_1(f) = a_m(f) \quad \text{for all } m \geq 1 \text{ with } \gcd(m, N) = 1.$$

If $a_1(f) = 0$, then $a_m(f) = 0$ for all $m \geq 1$ with $\gcd(m, N) = 1$, so the fact above implies $f \in \mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}$. But f is also in the orthogonal complement of $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}$ by assumption, so $f = 0$, contradiction. We conclude that $a_1(f) \neq 0$, and we may scale f such that $a_1(f) = 1$. We claim that f is a primitive form. Namely, for all $m \geq 1$, put

$$g_m = T_m f - a_m(f) f.$$

Then g_m is in $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$ because this space is preserved by the operators T_m. Furthermore, g_m is an eigenform for all the $\langle d \rangle$ ($d \in (\mathbb{Z}/N\mathbb{Z})^\times$) and the T_n with $\gcd(n, N) = 1$. This means that g_m satisfies the same conditions as f, except we do not know that $g_m \neq 0$. In fact, we have

$$\begin{aligned} a_1(g_m) &= a_1(T_m f) - a_1(a_m(f) f) \\ &= a_m(f) - a_m(f) a_1(f) \\ &= 0. \end{aligned}$$

Applying the same argument as above to g_m shows that $g_m = 0$, i.e. $T_m f = a_m(f) f$. Hence f is an eigenform for all T_m with $m \geq 1$ and therefore a primitive form. This proves part (a).

To prove part (b), we may assume that f and g are normalised (and hence primitive) by part (a). Then $f - g$ satisfies $a_m(f - g) = 0$ for all m with $\gcd(m, N) = 1$; applying the same argument again shows that $f - g = 0$, so $f = g$.

It remains to prove (c). Since the operators $\langle d \rangle$ and T_m with $\gcd(m, N) = 1$ form a commuting family of normal operators on the space $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$, there exists an orthogonal basis of eigenforms for all these operators. By part (a), we may scale these eigenforms such that they become primitive. What is left to prove is that these are all the primitive forms, i.e. that there is no linear relation between them. Suppose that there exists a linear relation

$$\sum_{i=1}^n c_i f_i = 0$$

with f_i distinct primitive forms and $c_i \neq 0$. We may assume that there is no linear relation with fewer terms. For any m, applying $T_m - a_m(f_1)$ to the relation above gives

$$\sum_{i=2}^n c_i (a_m(f_i) - a_m(f_1)) f_i = 0,$$

which is a linear relation with fewer terms. Hence $a_m(f_i) = a_m(f_1)$ for all $m \geq 1$, which implies $f_i = f_1$ for all $i \geq 2$. Since the f_i are distinct, we must have $i = 1$, but then the linear relation reads $f_1 = 0$, a contradiction. □

5.4 Exercises

Throughout these exercises, N and k are positive integers.

Exercise 5.1. Let $f \in \mathrm{S}_k(\Gamma_1(N))$ be a normalised Hecke eigenform with q-expansion $\sum_{n=1}^\infty a_n q^n$ (at the cusp ∞) and character $\chi\colon (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$.

(a) Prove the identity

$$\overline{a_m} = \chi(m)^{-1} a_m \quad \text{for all } m \geq 1 \text{ with } \gcd(m, N) = 1.$$

Deduce that the quantity $a_m^2/\chi(m)$ is real for all $m \geq 1$ such that $\gcd(m, N) = 1$.

(b) Prove the following statement, which you were allowed to use without proof in Exercise 4.12: Let $f \in \mathrm{M}_k(\mathrm{SL}_2(\mathbb{Z}))$ be a normalised eigenform, and let p be a prime number. Then $a_p(f)$ is real. (*Hint:* treat Eisenstein series and cusp forms separately.)

Exercise 5.2. Let V be be the space $\mathrm{S}_2(\Gamma_1(16))$ of cusp forms of weight 2 for $\Gamma_1(16)$. You may use the following fact without proof: a basis for V, expressed in q-expansions at the cusp ∞, is

$$\begin{aligned} f_1 &= q - 2q^3 - 2q^4 + 2q^6 + 2q^7 + 4q^8 - q^9 + O(q^{10}), \\ f_2 &= q^2 - q^3 - 2q^4 + q^5 + 2q^7 + 2q^8 - q^9 + O(q^{10}). \end{aligned}$$

(a) Show that $\mathrm{S}_2(\Gamma_1(8)) = \{0\}$ and $V = \mathrm{S}_2(\Gamma_1(16))_{\mathrm{new}}$. (*Hint:* consider the map $i_2^{8,16}$ on q-expansions.)

(b) Compute the matrix of the Hecke operator T_2 on V with respect to the basis (f_1, f_2).

(c) Compute a basis (g_1, g_2) of V consisting of eigenforms for T_2.

(Do the computations by hand; you may use a computer to check your results.)

Exercise 5.3. Let M and e be positive integers, let l be a prime number not dividing M, and let $N = l^e M$. Let f be a Hecke eigenform in $\mathrm{S}_k(\Gamma_1(M))$ with character χ. Let V_f be the $\mathbb{C}$-linear subspace of $\mathrm{S}_k(\Gamma_1(N))$ spanned by the forms $f_j = i_{l^j}^{M,N}(f)$ for $0 \le j \le e$.

(a) Prove that the forms $f_0, \ldots, f_e$ are $\mathbb{C}$-linearly independent.

(b) Show that the Hecke operator T_l on $\mathrm{S}_k(\Gamma_1(N))$ preserves the subspace V_f, and compute the matrix of T_l on V_f with respect to the basis $(f_0, \ldots, f_e)$.

Answer:
$$\begin{pmatrix} a_l & 1 & 0 & 0 & \cdots & 0 \\ -\chi(l)l^{k-1} & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

Exercise 5.4. Suppose that $\mathrm{S}_k(\Gamma_0(N))$ contains some normalised eigenform f. Write $g = f^2 \in \mathrm{S}_{2k}(\Gamma_0(N))$. Calculate the first two terms of the q-expansions of g and $T_2 g$, and deduce that the dimension of $\mathrm{S}_{2k}(\Gamma_0(N))$ is at least 2.

Exercise 5.5. Let Γ be a congruence subgroup, and let f be a modular form of weight k for Γ. Define a function $f^* \colon \mathbb{H} \to \mathbb{C}$ by

$$f^*(z) = \overline{f(-\bar{z})}.$$

(a) Prove that f^* is a modular form of weight k for the group $\sigma^{-1}\Gamma\sigma$, where $\sigma = \left(\begin{smallmatrix} -1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$.

(b) Suppose (for simplicity) that both Γ and $\sigma^{-1}\Gamma\sigma$ contain the subgroup $\left\{\left(\begin{smallmatrix} 1 & b \\ 0 & 1 \end{smallmatrix}\right) \mid b \in \mathbb{Z}\right\}$. Show that the standard q-expansions of f and f^* in the variable $q = \exp(2\pi i z)$ are related by

$$a_n(f^*) = \overline{a_n(f)} \quad \text{for all } n \ge 0.$$

(c) Show that if $\Gamma = \Gamma_0(N)$ or $\Gamma = \Gamma_1(N)$ for some $N \ge 1$, then $\sigma^{-1}\Gamma\sigma = \Gamma$.

Bonus problem: Give an example of a congruence subgroup Γ such that $\sigma^{-1}\Gamma\sigma \ne \Gamma$.

Exercise 5.6. Let g_1 and g_2 be the eigenforms for the operator T_2 on $\mathrm{S}_2(\Gamma_1(16))$ found in Exercise 5.2.

(a) Prove that g_1 and g_2 are in fact eigenforms for the full Hecke algebra $\mathbb{T}(\mathrm{S}_2(\Gamma_1(16)))$. (*Hint:* first show that $\mathrm{S}_2(\Gamma_1(16))$ admits a basis of eigenforms for the full Hecke algebra.)

(b) Compute the eigenvalues of the diamond operator $\langle 3\rangle$ on g_1 and g_2. (*Hint:* use T_3 and T_9.)

(c) Prove that the characters of g_1 and g_2 are given by

$$\langle d\rangle g_j = \chi_j(d) g_j \quad \text{for all } d \in (\mathbb{Z}/16\mathbb{Z})^\times \qquad (j = 1, 2),$$

where χ_1, χ_2 are the two group homomorphisms $(\mathbb{Z}/16\mathbb{Z})^\times \to \mathbb{C}^\times$ with kernel $\{\pm 1\}$.

(Do the computations by hand; you may use a computer to check your results.)

Exercise 5.7.

(a) Use the SageMath command `Newforms` to show that there is exactly one primitive form f of weight 6 for the group $\Gamma_1(4)$. Determine the q-expansion coefficients $a_n(f)$ for $n \le 20$.

(b) Prove that $a_n(f) = 0$ for all even integers n.

(c) Give a formula expressing the modular form θ^{12} (see §3.8) as a linear combination of $E_6(z)$, $E_6(2z)$, $E_6(4z)$ and f.

(d) Deduce that for all *even* integers $n \ge 2$, the number of representations of n as a sum of 12 squares is given by the formula

$$r_{12}(n) = 8\sum_{d|n} d^5 - 512 \sum_{d\,|\,n/4} d^5.$$

(Cf. Theorem 3.19; the sums are taken over all positive divisors of n and $n/4$, respectively, and the last sum is omitted if $4 \nmid n$.)

(As in the lecture, a *primitive form* is an eigenform f in the new subspace, normalised such that $a_1(f) = 1$. These are often also called *newforms*, which explains the name of the SageMath command `Newforms`.)

Exercise 5.8. For $f \in \mathrm{S}_k(\Gamma_1(N))$, let $f^* \in \mathrm{S}_k(\Gamma_1(N))$ be the form defined by $f^*(z) = \overline{f(-\bar{z})}$ (see Exercise 5.5).

(a) Show that the map $\mathrm{S}_k(\Gamma_1(N)) \to \mathrm{S}_k(\Gamma_1(N))$ sending f to f^* preserves the subspaces $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{old}}$ and $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$.

(b) Let $f \in \mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$ be a primitive form. Show that the form f^*, which by part (a) is in $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$, is also a primitive form, and determine the eigenvalues of the operators $\langle d\rangle$ (for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$) and T_m (for $m \ge 1$) on f^*.

Exercise 5.9. Recall that the Fricke (or Atkin–Lehner) operator w_N on $\mathrm{S}_k(\Gamma_1(N))$ is the operator T_{α_N} with $\alpha_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$.

(a) Show that $w_N^2 = (-N)^k \cdot \mathrm{id}$ and that the adjoint of w_N equals $(-1)^k w_N$.

(b) Show that for every $d \in (\mathbb{Z}/N\mathbb{Z})^\times$, the operator $\langle d\rangle$ on $\mathrm{S}_k(\Gamma_1(N))$ satisfies $w_N^{-1}\langle d\rangle w_N = \langle d\rangle^{-1}$.

(c) Show that for every positive integer m such that $\gcd(m, N) = 1$, the Hecke operator T_m satisfies $w_N^{-1} T_m w_N = \langle m\rangle^{-1} T_m$.

Exercise 5.10. Let w_N be the Fricke operator on $\mathrm{S}_k(\Gamma_1(N))$; recall that this preserves the new subspace $\mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$. Let $f \in \mathrm{S}_k(\Gamma_1(N))_{\mathrm{new}}$ be a primitive form.

(a) Show that the form $w_N f$ is an eigenform for the operators $\langle d\rangle$ for $d \in (\mathbb{Z}/N\mathbb{Z})^\times$ and T_m for $m \ge 1$ with $\gcd(m, N) = 1$, and determine the eigenvalues of these operators on $w_N f$.

(b) Deduce that $w_N f = \eta_f f^*$ for some $\eta_f \in \mathbb{C}$, with f^* as in Exercise 5.5. (*Hint:* use Exercise 5.1 as one ingredient.)

(c) Prove the identities $\eta_f \eta_{f^*} = (-N)^k$, $\eta_{f^*} = (-1)^k \bar{\eta}_f$ and $|\eta_f| = N^{k/2}$. (*Hint:* consider $\langle w_N f, f^* \rangle_{\Gamma_1(N)}$.)

You may use results from earlier exercises.

(The complex number η_f is called the *Atkin–Lehner pseudo-eigenvalue* of f.)

Chapter 6

L-functions

In modern number theory, L-functions are a fundamental tool for studying various kinds of arithmetic objects and the relations between them. The prototypical examples of an L-functions are the Riemann ζ-function and Dirichlet L-functions, i.e. L-functions attached to Dirichlet characters.

Modular forms are a very important "source" of L-functions. One reason for their importance is that the transformation properties of modular forms imply that the L-functions associated to them, which are a priory only defined on a suitable half-plane in $\mathbb{C}$, can in fact be continued to entire functions that satisfy a certain kind of functional equation. An even stronger reason is perhaps the way in which L-functions link elliptic curves to modular forms. The proof of Fermat's last theorem by Andrew Wiles in 1994 relies in an essential way on this connection.

In the context of modular forms and L-functions, we should also mention one of the most famous open problems in number theory, namely the conjecture of Birch and Swinnerton-Dyer. This conjecture, formulated in the 1960s based on some of the first computer calculations in number theory, is about L-functions of elliptic curves over $\mathbb{Q}$. It links various objects related to such an elliptic curve in a way comparable to the analytic class number formula from algebraic number theory. All results that have been obtained so far in the direction of the conjecture of Birch and Swinnerton-Dyer strongly depend on techniques related to modular forms.

6.1 The Mellin transform

We start by studying a kind of integral transform, related to the Fourier transform, that expresses the relationship between a modular form and the L-function associated to it.

Lemma 6.1. *Let $g\colon (0,\infty) \to \mathbb{C}$ be a continuous function such that for some real numbers $a < b$ we have*

$$|g(t)| \ll t^{-a} \quad \text{as } t \to 0$$

and

$$|g(t)| \ll t^{-b} \quad \text{as } t \to \infty.$$

Then the integral

$$\mathcal{M}g(s) = \int_0^\infty g(t) t^s \frac{dt}{t}$$

converges absolutely and uniformly on compact subsets of the strip $\{s \in \mathbb{C} \mid a < \Re s < b\}$.

Proof. Let α and β be real numbers with $a < \alpha < \beta < b$. Then we have

$$\begin{aligned}\int_0^\infty |g(t)t^s|\frac{dt}{t} &\ll \int_0^1 t^{-a}t^{\Re s}\frac{dt}{t} + \int_1^\infty t^{-b}t^{\Re s}\frac{dt}{t}\\ &\le \int_0^1 t^{\alpha-a}\frac{dt}{t} + \int_1^\infty t^{\beta-b}\frac{dt}{t}\\ &= \frac{1}{\alpha-a} + \frac{1}{b-\beta}.\end{aligned}$$

This implies that the integral defining $\mathcal{M}g(s)$ converges absolutely and uniformly on the strip $\{s \in \mathbb{C} \mid \alpha \le \Re s \le \beta\}$, from which the claim follows. □

Definition. For a function $g\colon \mathbb{H} \to \mathbb{C}$ satisfying the assumptions of the lemma above, the function $\mathcal{M}g$ is called the *Mellin transform* of g.

Remark. The *Mellin inversion formula* expresses $g(t)$ in terms of $\mathcal{M}g(s)$ as

$$g(t) = \frac{1}{2\pi i}\int_{\Re s = c} \mathcal{M}g(s)t^{-s}ds,$$

where c is any real number with $a < c < b$, and the integral is taken over the vertical line $\Re s = c$ in the upward direction. For details, see Exercise 6.2.

Example. The Γ*-function* is defined as the Mellin transform of the function $t \mapsto \exp(-t)$:

$$\Gamma(s) = \int_0^\infty \exp(-t)t^s\frac{dt}{t},$$

where the integral converges absolutely if and only if $\Re s > 0$. Using integration by parts, one shows that

$$\Gamma(s+1) = s\Gamma(s) \quad \text{for } \Re s > 0,$$

and this relation can be used to extend the Γ-function to a meromorphic function on $\mathbb{C}$ with poles in the non-positive integers and no other poles.

6.2 The L-function of a modular form

We will now define the L-function of a modular form and prove the basic properties of this L-function. We start with an explicit growth condition for the Fourier coefficients of a modular form. We are mainly interested in cusp forms, so we provide all details of the proof of the proposition below for cusp forms only. For the sake of completeness, we also incorporate the result for the space of Eisenstein series in the proposition, but we will not go into the details of the proof in that case.

Proposition 6.2. *Let Γ be a congruence subgroup, $k \in \mathbb{Z}_{>0}$, and $f \in M_k(\Gamma)$ with Fourier expansion at the cusp ∞ given by*

$$f(z) = \sum_{n=0}^\infty a_n(f)\exp\left(\frac{2\pi i n z}{h}\right) \ (\textit{with } h = \tilde{h}_\Gamma([\infty]) \in \mathbb{Z}_{>0}).$$

Then there exists a $C \in \mathbb{R}_{>0}$ such that for all $n \in \mathbb{Z}_{>0}$

$$|a_n(f)| \le Cn^{k/2} \ \textit{if } f \in S_k(\Gamma);$$

$$|a_n(f)| \le Cn^{k-1} \ \textit{if } f \in \mathcal{E}_k(\Gamma).$$

Proof. Let $f \in S_k(\Gamma)$. Consider the associated function $\tilde{f}$ on the unit disc $\mathbb{D}$ given by

$$\tilde{f}(q_h) = \sum_{n=1}^{\infty} a_n(f) q_h^n,$$

so

$$\tilde{f}\left(\exp\left(\frac{2\pi i z}{h}\right)\right) = f(z).$$

Note that $a_n(f)$ is the coefficient of q_h^{-1} in the Laurent expansion of $\tilde{f}(q_h)/q_h^{n+1}$ around $q_h = 0$. So Cauchy's formula gives

$$a_n(f) = \frac{1}{2\pi i} \oint_{C_r} \frac{\tilde{f}(q_h)}{q_h^n} \frac{dq_h}{q_h} \tag{6.1}$$

for any positively oriented circle C_r with centre O and radius r where $0 < r < 1$. Write $r = \exp(-2\pi y/h)$ with $y \in \mathbb{R}_{>0}$ and parametrize C_r by

$$q_h = \exp\left(\frac{2\pi i(x+iy)}{h}\right), \quad 0 \le x \le h$$

to rewrite (6.1) as

$$a_n(f) = \frac{1}{h} \int_0^h f(x+iy) \exp\left(-\frac{2\pi i n(x+iy)}{h}\right) dx. \tag{6.2}$$

Choosing $y = 1/n$ in (6.2) yields

$$a_n(f) = \frac{\exp(2\pi/h)}{h} \int_0^h f(x+i/n) \exp\left(-\frac{2\pi i n x}{h}\right) dx. \tag{6.3}$$

By Lemma 5.3, $z \mapsto |f(z)|^2 \Im(z)^k$ is bounded on $\mathbb{H}$. So in particular, there is a $C' \in \mathbb{R}_{>0}$ such that

$$|f(x+i/n)| \le C' n^{k/2}.$$

Together with (6.3) we get

$$|a_n(f)| \le \exp(2\pi/h) C' n^{k/2},$$

which proves the result for cusp forms.

For $f \in \mathcal{E}_k(\Gamma)$ it is possible to simply read off the result from an explicit basis for $\mathcal{E}_k(\Gamma)$. For the latter, see [4, Chapter 4] when $\Gamma = \Gamma(N)$, from which the growth condition for general Γ can be deduced. □

Let Γ be a congruence subgroup. Let f be a Hecke eigenform of weight k for Γ, normalised so that $a_1(f) = 1$.

One way to define the L-function $L(f,s)$ is as follows. Suppose r is a real number satisfying

$$|a_n(f)| \ll n^r \quad \text{as } n \to \infty.$$

According to Proposition 6.2, if f is any modular form of weight $k \in \mathbb{Z}_{>0}$, we can take $r = \max(k-1, k/2)$ (which is $k-1$ if $k \neq 1$, and $1/2$ otherwise); and if f is a cusp form of weight $k \in \mathbb{Z}_{>0}$, we can alternatively take $r = k/2$. From the bound above on the coefficients $a_n(f)$, it follows that for any $a > r+1$, the Dirichlet series

$$L(f,s) = \sum_{n \ge 1} a_n(f) n^{-s}$$

converges uniformly on the half-plane $\{s \in \mathbb{C} \mid \Re s \ge a\}$. This Dirichlet series therefore defines a holomorphic function on the right half-plane $\{s \in \mathbb{C} \mid \Re s > r+1\}$. (Notice that even though the constant term $a_0(f)$ in the q-expansion of f may be non-zero (since f is not necessarily a cusp form), this term does not appear in the definition of $L(f,s)$.) However, it is not immediately clear that this function can be continued to a meromorphic function on all of $\mathbb{C}$ that satisfies a functional equation.

The analytic continuation and functional equation will follow from the transformation properties of f under the Fricke (or Atkin–Lehner) operator w_N. Assume that there exists a normalised Hecke eigenform f^* and a complex number η_f satisfying the relation

$$w_N f = \eta_f f^*. \tag{6.4}$$

We note that due to the fact that $\alpha_N^2 = \left(\begin{smallmatrix} -N & 0 \\ 0 & -N \end{smallmatrix}\right)$, we then also have

$$w_N f^* = \eta_{f^*} f,$$

where η_f and η_{f^*} are related by

$$\eta_f \eta_{f^*} = (-N)^k. \tag{6.5}$$

We define the *completed L-function* attached to f as

$$\Lambda(f,s) = N^{s/2} \frac{\Gamma(s)}{(2\pi)^s} L(f,s).$$

Proposition 6.3. *Let $f \in \mathrm{S}_k(\Gamma_1(N))$ be a primitive form. Then there exist a primitive form $f^* \in \mathrm{S}_k(\Gamma_1(N))$ and a complex number $\eta_f \neq 0$ satisfying* (6.4).

Proof. Let a_p and $\epsilon(d)$ denote the eigenvalues of the Hecke operators T_p and the diamond operators $\langle d \rangle$ on f. We write $g = w_N f$. For every prime number p, we use the fact that the matrix $\alpha_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$ normalises $\Gamma_1(N)$, the identity

$$\alpha_N \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \alpha_N^{-1} = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}^*,$$

and Lemma 5.8 to deduce the identity

$$w_N^{-1} T_p w_N = T_p^\dagger.$$

If furthermore p does not divide N, then Proposition 5.9 gives

$$T_p^\dagger = \langle p \rangle^{-1} T_p.$$

This implies that g is an eigenform for T_p; more precisely, we have

$$T_p g = \epsilon(p)^{-1} a_p g.$$

Therefore g is an eigenform for all Hecke operators T_m with $\gcd(m,N) = 1$. Similarly, one shows that g is an eigenform for the diamond operators. By Theorem 5.13(a), g is a scalar multiple of a primitive form, i.e. we can write $g = \eta_f f^*$ as claimed. □

Remark. In fact, f^* is given by $f^*(z) = \overline{f(-\bar{z})}$; this follows from Exercise 5.10.

Theorem 6.4. *Let f be a normalised Hecke eigenform of weight k for $\Gamma_1(N)$, and assume that there exist a normalised Hecke eigenform f^* and a complex number η_f such that* (6.4) *holds.*

(a) The function $\Lambda(f,s)$ can be continued to a meromorphic function on $\mathbb{C}$ with at most simple poles in $s = 0$ and $s = k$, and no other poles. If f is a cusp form, then $\Lambda(f,s)$ is holomorphic on all of $\mathbb{C}$.

(b) The functions $\Lambda(f,s)$ and $\Lambda(f^*,s)$ are related by the functional equation

$$\Lambda(f, k-s) = \epsilon_f \Lambda(f^*, s), \tag{6.6}$$

where

$$\epsilon_f = i^k \eta_f N^{-k/2}.$$

Proof. We rewrite the transformation formula (6.4) as

$$f\biggl(-\frac{1}{Nz}\biggr) = \eta_f z^k f^*(z) \quad \text{for all } z \in \mathbb{H}.$$

Because $f^*(it)$ is bounded as $t \to \infty$, this formula implies that $|f(it)| \ll t^{-k}$ as $t \to 0$. Furthermore, we have $|f(it) - a_0(f)| \ll \exp(-2\pi t)$ as $t \to \infty$. This implies that the integral defining the Mellin transform of $f(it) - a_0(f)$ converges uniformly on compact subsets of the right half-plane $\{s \in \mathbb{C} \mid \Re s > k\}$.

If moreover r is a real number such that $|a_n(f)| \ll n^r$ as $n \to \infty$, then we can compute the Mellin transform of $f(it) - a_0(f)$ for $\Re s > \max\{k, r+1\}$ as

$$\begin{aligned}\int_0^\infty (f(it) - a_0(f))t^s \frac{dt}{t} &= \int_0^\infty \sum_{n\ge1} a_n(f)\exp(-2\pi nt)t^s\frac{dt}{t}\\ &= \sum_{n\ge1} a_n(f) \int_0^\infty \exp(-2\pi nt)t^s\frac{dt}{t}\\ &= \sum_{n\ge1} a_n(f)(2\pi n)^{-s}\int_0^\infty \exp(-u)u^s\frac{du}{u}\\ &= \frac{\Gamma(s)}{(2\pi)^s}\sum_{n\ge1} a_n(f)n^{-s}\\ &= \frac{\Gamma(s)}{(2\pi)^s}L(f,s).\end{aligned}$$

On the other hand, we can rewrite the integral on the left-hand side for $\Re s > k$ as

$$\int_0^\infty (f(it) - a_0(f))t^s\frac{dt}{t} = \int_{1/\sqrt{N}}^\infty (f(it) - a_0(f))t^s\frac{dt}{t} + \int_0^{1/\sqrt{N}} (f(it) - a_0(f))t^s\frac{dt}{t}.$$

The second term is

$$\int_0^{1/\sqrt{N}} (f(it) - a_0(f))t^s\frac{dt}{t} = \int_0^{1/\sqrt{N}} f(it)t^s\frac{dt}{t} - a_0(f)N^{-s/2}\frac{1}{s}.$$

By (6.4), the integral on the right-hand side is

$$\int_0^{1/\sqrt{N}} f(it)t^s\frac{dt}{t} = i^k\eta_f N^{-s}\int_{1/\sqrt{N}}^\infty f^*(it)t^{k-s}\frac{dt}{t}.$$

Splitting off the constant term from $f^*(it)$, we get

$$\int_{1/\sqrt{N}}^\infty f^*(it)t^{k-s}\frac{dt}{t} = \int_{1/\sqrt{N}}^\infty (f^*(it) - a_0(f^*))t^{k-s}\frac{dt}{t} + a_0(f^*)N^{(s-k)/2}\frac{1}{s-k}.$$

Putting everything together, we obtain for $\Re s > k$ the identity

$$\begin{aligned}\int_0^\infty (f(it) - a_0(f))t^s\frac{dt}{t} = &\int_{1/\sqrt{N}}^\infty (f(it) - a_0(f))t^s\frac{dt}{t}\\ &+ i^k\eta_f N^{-s}\int_{1/\sqrt{N}}^\infty (f^*(it) - a_0(f^*))t^{k-s}\frac{dt}{t}\\ &- a_0(f)N^{-s/2}\frac{1}{s} + i^k\eta_f a_0(f^*)N^{-(s+k)/2}\frac{1}{s-k}.\end{aligned}$$

Because $|f(it)-a_0(f)|$ and $|f^*(it)-a_0(f^*)|$ are bounded by a constant times $\exp(-2\pi t)$ as $t\to\infty$, both integrals on the right-hand side converge uniformly for s in compact subsets of $\mathbb{C}$. Combining the two expressions for the Mellin transform of $f(it)-a_0(f)$ computed above, we get

$$\begin{aligned}\Lambda(f,s) = N^{s/2}\int_{1/\sqrt{N}}^{\infty}(f(it)-a_0(f))t^s\frac{dt}{t} + i^k\eta_f N^{-s/2}\int_{1/\sqrt{N}}^{\infty}(f^*(it)-a_0(f^*))t^{k-s}\frac{dt}{t}\\ - a_0(f)\frac{1}{s} + i^k\eta_f N^{-k/2}a_0(f^*)\frac{1}{s-k}.\end{aligned}$$

This proves (a). Comparing the formula above to the analogous formula for $\Lambda(f^*,s)$ and using (6.5), we see that $\Lambda(f,s)$ and $\Lambda(f^*,s)$ are related by the functional equation (6.6). □

6.3 Exercises

Throughout these exercises, N and k are positive integers.

Exercise 6.1. Let f be a normalised eigenform of weight k for $\Gamma_1(N)$. Let a_p and $\chi(d)$ denote the eigenvalues of the Hecke operators T_p and the diamond operators $\langle d\rangle$ on f, respectively. Prove that $L(f,s)$ can be expressed as an *Euler product*: for S a sufficiently large real number, we have the identity

$$L(f,s) = \prod_{p \text{ prime}}\left(1-a_p p^{-s}+\chi(p)p^{k-1-2s}\right)^{-1} \quad \text{for } \Re s > S,$$

where the infinite product converges on compact subsets of the right half-plane $\{s\in\mathbb{C}\mid \Re s > S\}$ and defines a holomorphic function on this half-plane.

Exercise 6.2. A *Schwartz function* is an infinitely continuously differentiable function $f\colon\mathbb{R}\to\mathbb{C}$ such that for all $m,n\ge 0$ the function $x^m f^{(n)}(x)$ (where $f^{(n)}$ is the n-th derivative of x) tends to zero as $|x|\to\infty$. The Fourier transform of a Schwartz function f is defined as

$$\hat f(t) = \int_{-\infty}^{\infty} f(x)\exp(-2\pi itx)dx.$$

It is known that $\hat f$ is again a Schwartz function and that f can be recovered from $\hat f$ using the Fourier inversion formula,

$$f(x) = \int_{-\infty}^{\infty}\hat f(t)\exp(2\pi ixt)dt.$$

Use this to give a proof of the Mellin inversion formula (see §6.1) for functions $g(t)$ that satisfy the assumptions of Lemma 6.1 and are in addition infinitely continuously differentiable.

Exercise 6.3.

(a) Suppose k is even and $k\ge 4$. Prove that the L-function of the Eisenstein series E_k admits the factorisation

$$L(E_k,s)=\zeta(s)\zeta(s-k+1),$$

where $\zeta(s)$ is the Riemann ζ-function.

(b) (This part is optional and depends on the optional exercises 3.24, 3.25 and 3.26.) Let α, β be primitive Dirichlet characters modulo M and N, respectively, satisfying $\alpha(-1)\beta(-1)=(-1)^k$. Prove that the L-function of the Eisenstein series $E_k^{\alpha,\beta}\in \mathrm{M}_k(\Gamma_1(MN))$ associated to the pair (α,β) admits the factorisation

$$L(E_k^{\alpha,\beta},s)=L(\alpha,s)L(\beta,s-k+1).$$

Note: The fact that the L-function of an Eisenstein series has such a factorisation is one of the manifestations of the rule of thumb that Eisenstein series are ‘easier’ than cusp forms.

Exercise 6.4. Let $f \in \mathrm{S}_k(\Gamma_1(N))$ be an eigenform such that all coefficients $a_n(f)$ for $n \geq 1$ are real.

(a) Show that the complex number ϵ_f defined in Theorem 6.4 is either $+1$ or -1.

(b) Let r be the order of vanishing of the holomorphic function $L(f, s)$ in $s = k/2$. Prove that r is even if $\epsilon_f = +1$ and that r is odd if $\epsilon_f = -1$. (*Hint:* expand the completed L-function $\Lambda(f, s)$ in a power series around $s = k/2$.)

Chapter 7

Elliptic curves, modularity and the conjecture of Birch and Swinnerton-Dyer

There exists an important connection between modular forms and elliptic curves by means of their L-functions. Since knowledge of elliptic curves is not a prerequisite for this course, we will now introduce some background on elliptic curves that will enable us to explain this connection. Two illustrations of the connection between L-functions and modular forms are the modularity theorem for elliptic curves over $\mathbb{Q}$ and the conjecture of Birch and Swinnerton-Dyer.

The modularity theorem predicts that the L-function attached to any elliptic curve over $\mathbb{Q}$ is also the L-function of some primitive cusp form. This was stated as a conjecture by Shimura and Taniyama in the 1950s. Around 1995, Andrew Wiles (with the help of Richard Taylor) proved enough of the modularity theorem to deduce Fermat's last theorem. The proof of the modularity theorem was completed in 2001 by the work of Breuil, Conrad, Diamond and Taylor.

An elliptic curve E over $\mathbb{Q}$ has an L-function $L(E,s)$, which is defined purely in terms of *local* data (the reductions of E modulo prime numbers). The conjecture of Birch and Swinnerton-Dyer predicts the behaviour of the function $L(E,s)$ at the point $s=1$ in terms of *global* data (the rank of the group of rational points). This conjecture can be compared to the analytic class number formula from algebraic number theory, which links the residue at $s=1$ of the Dedekind ζ-function $\zeta_K(s)$ of a number field K to various arithmetic invariants attached to K.

7.1 Elliptic curves

We first give a 'canonical' definition of elliptic curves, and then say somewhat more concretely what an equation for an elliptic curve looks like.

Definition. An *elliptic curve* over a field K is a smooth cubic curve E in the projective plane over K together with a K-rational point $O \in E(K)$.

After a suitable choice of coordinates, every elliptic curve over K is given by a *Weierstrass equation.* This is a homogeneous equation of the form

$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3$$

with $a_1, \ldots, a_6 \in K$ (subject to some condition expressing the smoothness of E), with the point O having coordinates $(0:1:0)$. This is usually written in affine coordinates as

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6.$$

We write $E(K)$ for the set of all K-rational points of E. This can be thought of as the set of all pairs $(x, y) \in K \times K$ satisfying the above equation, together with the 'infinite' point O.

One of the most fundamental facts about elliptic curves is that the set $E(K)$ has the structure of an Abelian group with identity element O. The group structure is determined uniquely by the property that three points add up to O if and only if they are the three intersection points (counted with multiplicities) of E with a line.

For elliptic curves over $\mathbb{Q}$ (or more general number fields), the structure of $E(K)$ has been studied very extensively. The basic result about $E(K)$ in this case is the *Mordell–Weil theorem.*

Theorem 7.1 (Mordell–Weil)**.** *If $K = \mathbb{Q}$ (or more generally any number field), then the Abelian group $E(K)$ is finitely generated.*

If E is an elliptic curve over a number field K, then $E(K)$ is called the *Mordell–Weil group* of E. The above theorem implies that if E is an elliptic curve over $\mathbb{Q}$, we can write

$$E(\mathbb{Q}) = T \times \mathbb{Z}^r,$$

where T is a finite Abelian group (the torsion subgroup $E(\mathbb{Q})_{\mathrm{tor}}$ of $E(\mathbb{Q})$) and r is some non-negative integer, called the *(algebraic) rank* of E.

Given an elliptic curve E over $\mathbb{Q}$, there is a straightforward way of computing $E(\mathbb{Q})_{\mathrm{tor}}$, and there are only finitely many possibilities for the group $E(\mathbb{Q})_{\mathrm{tor}}$ up to isomorphism. However, it is in general much harder to determine r, and it is not known whether r can be arbitrarily large. Furthermore, it is in general 'hard' to determine a finite set of points $P_1, \ldots, P_r \in E(\mathbb{Q})$ such that $P_1, \ldots, P_r$ together with $E(\mathbb{Q})_{\mathrm{tor}}$ is a generating set of $E(\mathbb{Q})$.

Let E be an elliptic curve over $\mathbb{Q}$ given by a Weierstrass equation as above. We assume in addition that the a_i are in $\mathbb{Z}$ and that the equation is *minimal* (a notion that we will not make precise). Then for every prime number p, we consider the equation over $\mathbb{F}_p$ obtained by reducing the equation modulo p. The curve $E_{\mathbb{F}_p}$ defined by this equation is called the *reduction of E modulo p.*

We say that E has *good reduction* at p if $E_{\mathbb{F}_p}$ is a smooth curve; otherwise we say that E has *bad reduction* at p. There exists a positive integer N (called the *conductor* of E, and related to the *minimal discriminant* of E) such that E has bad reduction at p if and only if $p \mid N$. In particular, E only has bad reduction at finitely many prime numbers p.

For every prime number, the set $E(\mathbb{F}_p)$ is finite. We define integers a_p for p prime by

$$a_p = p + 1 - \#E(\mathbb{F}_p).$$

Remark. This is the correct definition even when the reduction of E modulo p is singular.

Theorem 7.2 (Hasse)**.** *The integers a_p satisfy*

$$|a_p| \le 2\sqrt{p}.$$

Furthermore, we define

$$\epsilon(p) = \begin{cases} 1 & \text{if } p \nmid N, \\ 0 & \text{if } p \mid N. \end{cases}$$

(In other words, ϵ is the trivial Dirichlet character modulo N.)

We now define

$$L(E, s) = \prod_{p \text{ prime}} \frac{1}{1 - a_p p^{-s} + \epsilon(p) p^{1-2s}}.$$

It follows from Hasse's theorem that this infinite product converges absolutely and uniformly for $\Re s \ge \sigma$, for any $\sigma > 3/2$. This implies that the product defines a holomorphic function on $\{s \in \mathbb{C} \mid \Re s > 3/2\}$. However, unlike in the setting of modular forms, where we have the Mellin transform at our disposal, there seems to be no easy way to prove that $L(E, s)$ has an analytic continuation and functional equation.

7.2 The modularity theorem

To be able to say anything really interesting about the L-function of an elliptic curve over $\mathbb{Q}$, we will need the modularity theorem. This is the following statement.

Theorem 7.3 (Modularity of elliptic curves over $\mathbb{Q}$)**.** *Let E be an elliptic curve over $\mathbb{Q}$. Then there exist $N \geq 1$ and a primitive form $f \in \mathrm{S}_2(\Gamma_0(N))$ of weight 2 such that $L(f, s)$ is equal to $L(E, s)$.*

Example. Let E be the elliptic curve defined by the equation

$$E\colon y^2 + y = x^3 - x^2.$$

Then the conductor N of E equals 11. (This is the smallest possible conductor of an elliptic curve over $\mathbb{Q}$.) There exists exactly one primitive cusp form of weight 2 for $\Gamma_0(11)$, namely

$$q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 + O(q^{10}).$$

Hence this is the primitive form attached to E by the modularity theorem.

The modularity theorem (formerly the Taniyama–Shimura conjecture) is a very deep result. One of its most important consequences is that it implies that L-functions of elliptic curves over $\mathbb{Q}$ admit an analytic continuation to all of $\mathbb{C}$ and satisfy a functional equation relating $L(E, s)$ and $L(E, 2 - s)$ (again via completed L-functions, like in the case of L-functions of modular forms). This is not at all obvious, and the modularity theorem is currently the only known way to prove the analytic continuation and functional equation for L-functions of elliptic curves. For elliptic curves over other number fields than $\mathbb{Q}$, only very partial results are known.

The modularity theorem was proved first for an important class of elliptic curves (the *semistable* ones, corresponding to square-free N) in 1995 by work of Wiles [11], completed by Taylor and Wiles [9], from which Fermat's last theorem follows. (More about Fermat's last theorem will be said in the last lecture of this course.) The modularity theorem was then proved in more generality by Diamond (1996), Conrad, Diamond and Taylor (1999) and finally for arbitrary E by Breuil, Conrad, Diamond and Taylor (2001).

The proof of the modularity theorem combines many different techniques, which we cannot explain here. Just to mention one key part of the proof that is related to this course: one ingredient is to show that certain Hecke algebras are isomorphic to so-called *deformation rings*. This is then used to prove a *modularity lifting theorem*, which is a statement of the type that if the coefficients a_n of $L(E, s)$ are congruent to the coefficients of a modular form modulo some prime number l, then the coefficients a_n are actually *equal* to the coefficients of a modular form.

Remark. We have phrased the modularity theorem in terms of L-functions, but there are many other formulations; see Diamond and Shurman's book for alternative versions.

7.3 The conjecture of Birch and Swinnerton-Dyer

Around 1958, Birch and Swinnerton-Dyer performed some of the first computer calculations in number theory. Given an elliptic curve over $\mathbb{Q}$ given by a Weierstrass equation with coefficients in $\mathbb{Z}$, they studied the way in which the number of points $\#E(\mathbb{F}_p)$ of the reduction depends on the rank of E over $\mathbb{Q}$. Based on their calculations, they stated a conjecture that can be formulated in one way as follows.

Conjecture 7.4 (Birch, Swinnerton-Dyer)**.** *Let E be an elliptic curve over $\mathbb{Q}$. Then the order of vanishing of function $L(E, s)$ in $s = 1$ (which is defined thanks to the modularity theorem) is equal to the rank of E.*

Remark. The order of vanishing of $L(E, s)$ in $s = 1$ is called the *analytic rank* of E; hence the Birch–Swinnerton-Dyer conjecture claims that the analytic rank of an elliptic curve is equal to its algebraic rank.

The conjecture of Birch and Swinnerton-Dyer is as yet unproved. It is in fact one of the "Millennium Prize Problems"; a proof is therefore worth one million dollars. The only general result known so far is the following.

Theorem 7.5 (Gross, Zagier; Kolyvagin)**.** *Let E be an elliptic curve over $\mathbb{Q}$. If the analytic rank of E is either* 0 *or* 1*, then it is equal to the algebraic rank of E.*

The proof of this theorem heavily relies on modular forms, modular curves, L-functions and related techniques.

In addition, Bhargava and Shankar proved in recent years that a positive proportion (in a precise sense) of all elliptic curves over $\mathbb{Q}$ have analytic rank 0, which implies that the conjecture of Birch and Swinnerton-Dyer holds for a positive proportion of all elliptic curves over $\mathbb{Q}$. (It seems, however, that they have not been awarded the corresponding fraction of the prize money.)

Remark. Let r denote the analytic rank of E. We consider the non-zero complex number

$$L^*(E,s) = \lim_{s\to 1}(s-1)^{-r}L(E,s).$$

There is a refined variant of the conjecture of Birch and Swinnerton-Dyer, which predicts the exact value of $L^*(E,s)$ in terms of certain analytic and arithmetic invariants of E (namely *periods*, *Tamagawa numbers*, the *regulator*, the order of the torsion group $E(\mathbb{Q})_{\mathrm{tor}}$, and the order of the *Tate–Shafararevich group*).

7.4 The congruent number problem

In this last section, we discuss how elliptic curves, modular forms and the conjecture of Birch and Swinnerton-Dyer can be applied to a classical Diophantine problem.

A positive rational number n is called *congruent* if there exists a right-angled triangle with rational side lengths and area n. The *congruent number problem* is the question which n are congruent. This comes down to the question for which n the system of equations

$$a^2 + b^2 = c^2 \quad \text{and} \quad ab = 2n$$

has a solution in non-zero rational numbers a, b, c.

Proposition 7.6. *A positive rational number n is congruent if and only if the equation*

$$y^2 = x^3 - n^2 x \tag{7.1}$$

has a solution $(x, y) \in \mathbb{Q}\times\mathbb{Q}$ with $y \neq 0$.

It suffices to study the case where n is a square-free positive integer. The equation (7.1) defines an elliptic curve E_n over $\mathbb{Q}$. The congruent number problem was solved by Tunnell [10] assuming the Birch–Swinnerton-Dyer conjecture for elliptic curves of the form E_n. We define integers c_n for $n \ge 1$ by the following identities of holomorphic functions $\mathbb{H} \to \mathbb{C}$ (or of power series in q):

$$\begin{aligned} h &= \eta(8z)\eta(16z) \\ &= q\prod_{m\ge 1}\left((1-q^{8m})(1-q^{16m})\right), \\ f &= h\theta(2z), \\ g &= h\theta(4z), \\ c_n &= \begin{cases} a_n(f) & \text{if } n \text{ is odd,} \\ a_{n/2}(g) & \text{if } n \text{ is even.} \end{cases} \end{aligned}$$

Here η and θ are the Dedekind η-function and the Jacobi θ-function defined in the course. (The functions f and g are in fact "modular forms of weight 3/2".)

Theorem 7.7 (Tunnell [10])**.** *Let n be a square-free positive integer. If n is congruent, then $c_n = 0$. The converse is true if the Birch–Swinnerton-Dyer conjecture holds for the elliptic curve E_n.*

Tunnell's proof of the first implication relies on partial results on the Birch–Swinnerton-Dyer conjecture due to Coates and Wiles [2].

Appendix A

Appendix: analysis and linear algebra

A.1 Uniform convergence

Let S be any set, and let $\{f_n\}_{n\geq 0}$ be a sequence of complex-valued functions on S. We say that the sequence *converges uniformly* if it is a Cauchy sequence with respect to the supremum norm on the $\mathbb{C}$-vector space of all functions $S \to \mathbb{C}$. In other words, $\{f_n\}_{n\geq 0}$ converges uniformly if for all $\epsilon > 0$ there exists $N \geq 0$ such that for all $m, n \geq N$ and all $s \in S$ we have $|f_m(s) - f_n(s)| < \epsilon$. From the fact that $\mathbb{C}$ is complete, it follows that any uniformly convergent sequence of functions has a unique limit.

When S is a subset of $\mathbb{C}$, uniform convergence allows us to interchange limits as follows. Let $\{f_n\}_{n\geq 0}$ be a sequence of continuous functions $S \to \mathbb{C}$ that converges uniformly on S with limit f. Then f is again continuous, i.e.

$$\lim_{z\to a}\lim_{n\to\infty} f_n(z) = \lim_{z\to a} f(z) = f(a) = \lim_{n\to\infty} f_n(a).$$

In particular, for a uniformly convergent sum of continuous functions, we may interchange the sum with limits:

$$\lim_{z\to a}\sum_{n=1}^{\infty} g_n(z) = \sum_{n=1}^{\infty} g_n(a).$$

There are many variants. For example, if $\{f_n\}_{n\geq 0}$ is a sequence of continuous functions $\mathbb{R} \to \mathbb{C}$ that converges uniformly on some interval of the form $[M, \infty)$ with limit, and if $\lim_{x\to\infty} f_n(x)$ exists for all n, then $\lim_{x\to\infty} f(x)$ also exists, and we have

$$\lim_{x\to\infty} f(x) = \lim_{n\to\infty}\lim_{x\to\infty} f_n(x).$$

In particular, for a sum that converges uniformly on some interval $[M, \infty)$, we have

$$\lim_{x\to\infty}\sum_{n=1}^{\infty} g_n(x) = \sum_{n=1}^{\infty}\lim_{x\to\infty} g_n(x).$$

A.2 Uniform convergence of holomorphic functions

Let U be an open subset of $\mathbb{C}$, and let $\{f_n\}_{n\geq 0}$ be a sequence of complex-valued functions on U. We say that the sequence *converges uniformly on compact subsets of* U if for every compact subset $K \subset U$ the sequence of functions $\{f_n|_K\}_{n\geq 0}$ converges uniformly.

Theorem A.1. *Let U be an open subset of $\mathbb{C}$, and let $\{f_n\}_{n\geq 0}$ be a sequence of holomorphic functions that converges uniformly on compact subsets of U. Then the limit function $f\colon U \to \mathbb{C}$ is holomorphic.*

The result above applies in particular to sums of holomorphic functions (consider the sequence of partial sums).

A.3 Orders and residues

Let f be a meromorphic function on an open subset $U \subset \mathbb{C}$, and w is a point of U. Then we can expand z in a Laurent series around w:

$$f(z) = c_n(z-w)^n + c_{n+1}(z-w)^{n+1} + \dots \quad (n \in \mathbb{Z}, c_j \in \mathbb{C}, c_n \neq 0).$$

The integer n is called the *order* or *valuation* of f at w and denoted by $\operatorname{ord}_w f$. The *residue* of f at w is c_{-1} (which is defined to be 0 if $n \geq 0$). From the series expansion of f above, one deduces

$$\frac{f'(z)}{f(z)} = \frac{n}{z-w} + b_0 + b_1(z-w) + \cdots$$

In particular, the function f'/f has a simple pole precisely at the points where f has a zero or a pole, and

$$\operatorname{Res}_w(f'/f) = n = \operatorname{ord}_w f.$$

A *contour* is a simple closed piecewise C^1 path in $\mathbb{C}$ with anti-clockwise orientation. For a contour C, the complement $\mathbb{C} - C$ consists of 2 connected components, exactly one of which is bounded; we denote this connected component by $\operatorname{interior}(C)$.

Theorem A.2 (Cauchy's integral formula)**.** *Let g be holomorphic on an open subset $U \subset \mathbb{C}$, let C be a contour in U, and let $w \in \operatorname{interior}(C)$. Then we have*

$$\oint_C \frac{g(z)}{z-w}dz = 2\pi i g(w).$$

Theorem A.3 (argument principle)**.** *Let f be meromorphic on an open subset $U \subset \mathbb{C}$, and let C be a contour in U not passing through any zeroes or poles of f. Then we have*

$$\oint_C \frac{f'(z)}{f(z)}dz = 2\pi i \sum_{z\in\operatorname{interior}(C)} \operatorname{ord}_z f.$$

A variant of this is the following: let C be an arc around w with angle α and radius r (not necessarily a contour). Then if g is holomorphic at w, we have

$$\lim_{r\to 0} \int_C \frac{g(z)}{z-w}dz = \alpha i g(w),$$

and if f is meromorphic at w, we have

$$\lim_{r\to 0} \int_C \frac{f'(z)}{f(z)}dz = \alpha i \operatorname{ord}_w f.$$

A.4 Cotangent formula and maximum modulus principle

For all $z \in \mathbb{C} - \mathbb{Z}$ we have

$$\pi \frac{\cos(\pi z)}{\sin(\pi z)} = \frac{1}{z} + \sum_{n=1}^{\infty} \left(\frac{1}{z-n} + \frac{1}{z+n} \right). \tag{A.1}$$

One way to prove this formula is by the so-called Herglotz trick; see e.g. Chapter 20 of *Proofs from THE BOOK* by Aigner and Ziegler (and generalize from $\mathbb{R}$ to $\mathbb{C}$).

Theorem A.4. *Let $U \subset \mathbb{C}$ be connected and open, and let $f \colon U \to \mathbb{C}$ be holomorphic. If $|f|$ attains a maximum on U, then f is constant.*

A.5 Infinite products

Theorem A.5. *Let U be an open subset of $\mathbb{C}$ and let $\{f_n\}_{>0}$ be a sequence of holomorphic functions on $\mathbb{C}$ such that*

$$\sum_{n=1}^{\infty} |f_n|$$

converges uniformly on compact subsets of U. Then the following holds.

- *The sequence of partial products*

$$F_N := \prod_{n=1}^{N} (1 + f_n)$$

converges uniformly on compact subsets of U. In particular, the limit F is holomorphic on U.

- *For all $z \in U$ we have $\operatorname{ord}_z(F) = \sum_{n=1}^{\infty} (\operatorname{ord}_z(1 + f_n))$.*
- *For every $z \in U$ such that $F(z) \neq 0$ we have*

$$\frac{F'(z)}{F(z)} = \sum_{n=1}^{\infty} \frac{f_n'(z)}{1 + f_n(z)}.$$

A.6 Fourier analysis and the Poisson summation formula

We recall that every function $F \colon \mathbb{R} \to \mathbb{C}$ that is infinitely often continuously differentiable and is periodic with period 1 admits a *Fourier series*

$$F(x) = \sum_{n \in \mathbb{Z}} c_n \exp(2\pi i n x).$$

The constants c_n satisfy

$$c_n = \int_0^1 F(x) \exp(-2\pi i n x) dx.$$

Let $f \colon \mathbb{R} \to \mathbb{C}$ be a function that is infinitely continuously differentiable and such that all derivatives $f^{(n)}$ for $n \geq 0$ have the property that $f^{(n)}(x)$ tends to zero exponentially fast as $|x| \to \infty$. We recall that the *Fourier transform* of such a function f is defined as

$$\hat{f}(t) = \int_{-\infty}^{\infty} f(x) \exp(-2\pi i x t) dx.$$

Theorem A.6 (Poisson summation formula)**.** *Let $f\colon \mathbb{R} \to \mathbb{C}$ be a function as above. Then we have*

$$\sum_{m\in\mathbb{Z}} f(m) = \sum_{n\in\mathbb{Z}} \hat{f}(n).$$

Proof. The function

$$F(x) = \sum_{m\in\mathbb{Z}} f(x+m)$$

is periodic with period 1 and therefore admits a Fourier series

$$F(x) = \sum_{n\in\mathbb{Z}} c_n \exp(2\pi i n x).$$

The coefficients c_n are given by

$$\begin{aligned} c_n &= \int_0^1 F(x)\exp(-2\pi i n x)dx \\ &= \int_0^1 \sum_{m\in\mathbb{Z}} f(x+m)\exp(-2\pi i n(x+m))dx \\ &= \int_{-\infty}^{\infty} f(x)\exp(-2\pi i n x)dx \\ &= \hat{f}(n). \end{aligned}$$

Substituting $x = 0$ in the Fourier series for $F(x)$, we obtain

$$\sum_{m\in\mathbb{Z}} f(m) = \sum_{n\in\mathbb{Z}} \hat{f}(n),$$

as claimed. □

Proposition A.7. *The function*

$$f(x) = \exp(-\pi x^2)$$

is its own Fourier transform.

Proof. We compute

$$\begin{aligned} \hat{f}(t) &= \int_{-\infty}^{\infty} \exp(-\pi x^2 - 2\pi i t x)dx \\ &= \int_{-\infty}^{\infty} \exp(-\pi(x+it)^2 - \pi t^2)dx \\ &= \exp(-\pi t^2)\int_{-\infty}^{\infty} \exp(-\pi(x+it)^2)dx \\ &= \exp(-\pi t^2)\int_{-\infty}^{\infty} \exp(-\pi x^2 dx) \\ &= \exp(-\pi t^2), \end{aligned}$$

where we have used contour integration and the identity $\int_{-\infty}^{\infty} \exp(-\pi x^2)dx = 1$. □

Corollary A.8. *For all $a > 0$, the Fourier transform of the function*

$$f_a(x) = \exp(-\pi a x^2)$$

is

$$\hat{f}_a(t) = \frac{1}{\sqrt{a}}\exp(-\pi t^2/a).$$

Proof. We note that $f_a(x) = f(\sqrt{a}x)$. We compute

$$\begin{aligned}\hat{f}_a(t) &= \int_{-\infty}^{\infty} f_a(x)\exp(-2\pi ixt)dx \\ &= \int_{-\infty}^{\infty} f(\sqrt{a}x)\exp(-2\pi ixt)dx \\ &= \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} f(u)\exp\bigl(-2\pi iut/\sqrt{a}\bigr)du \\ &= \frac{1}{\sqrt{a}}\hat{f}\bigl(t/\sqrt{a}\bigr) \\ &= \frac{1}{\sqrt{a}}f\bigl(t/\sqrt{a}\bigr).\end{aligned}$$

This proves the claim. □

A.7 The spectral theorem

We recall some definitions and facts from linear algebra.

Let V be a finite-dimensional $\mathbb{C}$-vector space, equipped with a Hermitean inner product $\langle\ ,\ \rangle$. If A is an endomorphism of V, then there exists a unique endomorphism $A^\dagger$ of V, called the *adjoint* of A with respect to $\langle\ ,\ \rangle$, such that

$$\langle Av, w\rangle = \langle v, A^\dagger w\rangle \quad \text{for all } v, w \in V.$$

With respect to any $\mathbb{C}$-basis of V that is orthonormal with respect to $\langle\ ,\ \rangle$, the matrix of $A^\dagger$ is the conjugate tranpose of the matrix of A. The operator A is called *normal* if it commutes with its adjoint $A^\dagger$.

Theorem A.9 (spectral theorem)**.** *Let V be a finite-dimensional $\mathbb{C}$-vector space, and let $\Phi \subset \operatorname{End}_{\mathbb{C}} V$ be a family of normal, pairwise commuting endomorphisms of V. Then there exists a $\mathbb{C}$-basis of V consisting of simultaneous eigenvectors for all the $A \in \Phi$.*

Bibliography

[1] J.H. Bruinier, G. van der Geer, G. Harder and D. Zagier, *The 1-2-3 of Modular Forms*. Universitext, Springer-Verlag, Berlin/Heidelberg, 2008.

[2] J. Coates and A. Wiles, On the conjecture of Birch and Swinnerton-Dyer. *Inventiones mathematicae* **39** (1977), no. 3, 223–251.

[3] H. Cohen and F. Strömberg, *Modular Forms. A Classical Approach*. American Mathematical Society, Providence, RI, 2017.

[4] F. Diamond and J. Shurman, *A First Course in Modular Forms*. Springer-Verlag, Berlin/Heidelberg/New York, 2005.

[5] J. S. Milne, Modular Functions and Modular Forms. Course notes, `http://www.jmilne.org/math/CourseNotes/mf.html`.

[6] T. Miyake, *Modular Forms*. Springer-Verlag, Berlin/Heidelberg, 1989.

[7] J-P. Serre, *Cours d'arithmétique*. Presses universitaires de France, Paris, 1970. (4$^{\text{e}}$ édition : 1995.) English translation: *A Course in Arithmetic*. Springer-Verlag, Berlin/Heidelberg/New York, 1973. (5th edition: 1996.)

[8] W. A. Stein, *Modular Forms, a Computational Approach*. With an appendix by P. E. Gunnells. American Mathematical Society, Providence, RI, 2007.

[9] R. Taylor and A. Wiles, Ring-theoretic properties of certain Hecke algebras. *Annals of Mathematics (2)* **141** (1995), no. 3, 553–572.

[10] J. B. Tunnell, A classical Diophantine problem and modular forms of weight 3/2. *Inventiones mathematicae* **72** (1983), 323–334.

[11] A. Wiles, Modular elliptic curves and Fermat's last theorem. *Annals of Mathematics (2)* **141** (1995), no. 3, 443–551.