Part C Modular Forms HT 2025*

Alan Lauder

January 13, 2025

1 Introduction

1.1 Overview: what you will learn about

Let

$$\mathfrak{H} := \{z \in \mathbb{C} \,|\, \mathrm{Im}(z) > 0\}.$$

denote the complex upper half plane. Modular forms are holomorphic functions

$$f : \mathfrak{H} \to \mathbb{C} \text{ with } |f(z)| \text{ bounded as } \mathrm{Im}(z) \to \infty$$

which have certain very strong "invariance properties". To describe these invariance properties one must first fix a choice of *weight* $k \in \mathbb{Z}_{\geq 0}$, *level* $N \in \mathbb{N}$ and *Dirichlet character* $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$. Having chosen such a k, N and χ, the set of modular forms with those particular "invariance properties" forms a finite-dimensional $\mathbb{C}$-vector space denoted $M_k(N, \chi)$.

Most of the modular forms f we shall study are invariant under the transformation $z \mapsto z + 1$ and thus have a Fourier expansion in terms of the function $q(z) := \exp(2\pi i z)$:

$$f(z) = \sum_{n=0}^{\infty} a_n(f) q^n,\ a_n(f) \in \mathbb{C}.$$

Note that the Fourier coefficients $a_n(f) = 0$ for $n < 0$ by the boundedness condition as $\mathrm{Im}(z) \to \infty$. It is a deep arithmetic fact—which we shall prove for many spaces of modular forms—that one may take as a basis for the space $M_k(N, \chi)$ rather special modular forms f whose Fourier coefficients $a_n(f)$ are algebraic numbers. These special basis elements are called the *eigenforms*. (I am glossing over some technical points here and elsewhere.)

*These notes are based upon those from MT 2013, but very heavily revised. I have changed notation and now talk about spaces of modular forms of "weight k (k even)" rather than"weight $2k$". Past examination papers from 2012 and 2013 follow the former notation, so beware.

For level $N = 1$ and trivial Dirichlet character χ_1, each eigenform f in this canonical basis is a simultaneous eigenvector for an infinite set of commuting linear maps, called the *Hecke operators*:

$$T_n : M_k(1, \chi_1) \to M_k(1, \chi_1) \text{ for } n \in \mathbb{N}.$$

Moreover, different eigenforms f and g in this canonical basis are orthogonal with respect to an inner product $\langle \cdot, \cdot \rangle$, called the *Petersson inner product*. (The same is true for general level N and character χ, with some caveats.)

The eigenforms may be used to construct representations of the absolute Galois group of $\mathbb{Q}$. This important fact explains why the coefficients $a_n(f)$ of the eigenforms are arithmetically interesting, but we will not have time to develop this point of view.

The purpose of this course is to develop as much of this beautiful theory as possible in sixteen lectures.

1.2 Examples of modular forms

Before fully defining modular forms let me give some of my favourite examples of eigenforms.

Example 1.1 Let $\zeta(z)$ be the Riemann zeta function (Part C Analytic Number Theory). Then for $k \geq 2$ and even we have

$$\zeta(k) := \sum_{n=1}^{\infty} \frac{1}{n^k} = \frac{b_k 2^{k-1} \pi^k}{k!}$$

where $b_k \in \mathbb{Q}$ is called the kth *Bernoulli number*. For example

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$$

and so $b_2 = \frac{1}{6}$. For $k \geq 2$ and even define

$$E_k(z) := 1 + \frac{(-1)^{k/2} 2k}{b_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n$$

where

$$\sigma_{k-1}(n) := \sum_{d|n} d^{k-1}.$$

Then for $k \geq 4$ and even

$$E_k(z) \in M_k(1, \chi_1)$$

where χ_1 is the trivial character. The $E_k(z)$ are called the *Eisenstein series* in level 1, and we shall study them in detail. (Note that $E_2(z) \in \mathbb{Z}[[q]]$ but it is not (quite) a modular form.)

Example 1.2 Consider the elliptic curve (see Part C Elliptic Curves)

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

For p prime with $p \neq 37$ define

$$a_p(E) := p - \#\{(x, y) \in \mathbb{F}_p^2 : y^2 + y = x^3 - x\} \in \mathbb{Z}.$$

Let $a_1(E) := 1$, $a_{37^r}(E) := (-1)^r$ $(r \geq 1)$ and

$$\begin{aligned} a_{p^{r+1}}(E) &:= a_p(E) \cdot a_{p^r}(E) - p \cdot a_{p^{r-1}}(E) \text{ for } r \geq 1 \text{ and } p \neq 37, \\ a_{mn}(E) &:= a_m(E)a_n(E) \text{ when } \gcd(m, n) = 1. \end{aligned}$$

Then

$$f_E := \sum_{n=1}^{\infty} a_n(E)q^n = q - 2q^2 - 3q^3 + 2q^4 - 2q^5 + 6q^6 - q^7 + 6q^9 + 4q^{10} - 5q^{11} + \cdots \in M_2(37, \chi_1)$$

is an eigenform.

For any elliptic curve E one may follow the same recipe to get an $f_E \in \mathbb{Z}[[q]]$ and it turns out f_E is always a modular form. This is the "Modularity theorem", proved (under mild hypotheses) by Wiles and Taylor-Wiles: it implies Fermat Last theorem, so needless to say we shall not prove this!

Example 1.3 Let $K := \mathbb{Q}(\sqrt{-23})$. This is the imaginary quadratic field of smallest discriminant (in absolute value) whose class number is 3 (see Part B Algebraic Number Theory). For p prime

$$a_p(K) := \begin{cases} 0 & \text{if } p \text{ is inert in } K \\ -1 & \text{if } p \text{ is split in } K \\ 1 & \text{if } p \text{ ramifies in } K. \end{cases}$$

Note that only 23 ramifies in K. Let $a_1(K) := 1$ and

$$\begin{aligned} a_{p^{r+1}}(K) &:= a_p(K) \cdot a_{p^r}(K) - \left(\tfrac{p}{23}\right) \cdot a_{p^{r-1}}(K) \text{ for } r \geq 1, \\ a_{mn}(K) &:= a_m(K)a_n(K) \text{ when } \gcd(m, n) = 1. \end{aligned}$$

Then

$$f_K := \sum_{n=1}^{\infty} a_n(K)q^n = q - q^2 - q^3 + q^6 + q^8 - q^{13} - q^{16} + q^{23} - q^{24} + \cdots \in M_1(23, \chi)$$

where $\chi : (\mathbb{Z}/23\mathbb{Z})^\times \to \{\pm 1\} \subset \mathbb{C}$ is given by the Legendre symbol $\left(\frac{\cdot}{23}\right)$ (Part A Number Theory).

This association of modular forms of weight one with imaginary quadratic fields is due to Hecke, from (roughly) around 1930. Modular forms of weight one are particularly mysterious, and not amenable to many of the methods in this course.

1.3 The definition of a modular form

Let $\mathrm{GL}_2^+(\mathbb{R})$ denote the group of 2×2 real matrices with positive determinant, and $\mathrm{SL}_2(\mathbb{Z})$ the subgroup of integer matrices with determinant 1. For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R})$ and $z \in \mathfrak{H}$ define

$$\gamma(z) = \frac{az+b}{cz+d} \in \mathfrak{H}.$$

(You should check yourself that indeed $\mathrm{Im}(\gamma(z)) > 0$, or look at the proof of Lemma 2.5.)

Definition 1.4. *For $f : \mathfrak{H} \to \mathbb{C}$ and $k \in \mathbb{Z}$ define the* **weight k action** *of $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2^+(\mathbb{R})$ on f by*

$$f|_k\gamma(z) := \det(\gamma)^{k/2} \cdot f(\gamma(z)) \cdot (cz+d)^{-k}.$$

That is, $f|_k\gamma : \mathfrak{H} \to \mathbb{C}$ via this formula. (Here $\det(\gamma)^{k/2}$ is the positive square-root when k is odd.)

Exercise 1.5 Check this a group action of $\mathrm{GL}_2^+(\mathbb{R})$ on the set of functions $f : \mathfrak{H} \to \mathbb{C}$, and moreover it preserves the property of f being holomorphic.

Let Γ be a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ of finite index and $k \in \mathbb{Z}$ be an integer.

Definition 1.6. *A holomorphic function $f : \mathfrak{H} \to \mathbb{C}$ is called a* **modular form** *of weight k for Γ if*

(1) (Invariance) For every $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $z \in \mathfrak{H}$ we have

$$f(\gamma(z)) = (cz+d)^k \cdot f(z).$$

(2) (Holomorphy at the cusps) For each $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ we have

$$|f|_k\gamma(z)| \text{ is bounded as } \mathrm{Im}(z) \to \infty.$$

The set of all such modular forms is denoted $M_k(\Gamma)$.

If $f \in M_k(\Gamma)$ further satisfies for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$,

$$|f|_k\gamma(z)| \to 0 \text{ as } \mathrm{Im}(z) \to \infty$$

we say it is a **cusp form**. *The set of all cusps forms is denoted $S_k(\Gamma)$.*

Equivalently one has f is a modular form of weight k for Γ if and only if

(1′) For every $\gamma \in \Gamma$ we have

$$f|_k\gamma = f.$$

(2′) Let $\gamma_1, \cdots, \gamma_r$ be right coset representatives for Γ in $\mathrm{SL}_2(\mathbb{Z})$ (so $\mathrm{SL}_2(\mathbb{Z}) = \cup_i \Gamma\gamma_i$). Then

$$|f|_k\gamma_i(z)| \text{ is bounded as } \mathrm{Im}(z) \to \infty$$

for each $1 \leq i \leq r$.

Exercise 1.7 Show (1) $\Leftrightarrow$ (1′), (2) $\Rightarrow$ (2′) and ((1′) and (2′)) $\Rightarrow$ (2). Prove $M_k(\Gamma)$ and $S_k(\Gamma)$ are $\mathbb{C}$-vector spaces.

Observe that provided our subgroup Γ contains the matrix

$$T := \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

then any $f \in M_k(\Gamma)$ will satisfy $f(z+1) = f(z)$ and so have a Fourier expansion

$$f(q) = \sum_{n=0}^{\infty} a_n(f) q^n,\ q := e^{2\pi i z}.$$

Note that as $\mathrm{Im}(z) \to \infty$ we have $f(z) \to a_0(f)$. So if f is a cusp form then $a_0(f) = 0$. The converse is true when $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ but not in general: the condition $a_0(f) = 0$ tells us nothing about the behaviour of $|f|_k\gamma(z)|$ as $\mathrm{Im}(z) \to \infty$ for $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma \notin \Gamma$.

Let $N \in \mathbb{N}$. The most important subgroups for us are the **congruence subgroups**:

$$\begin{aligned}
\Gamma(N) &:= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : a \equiv d \equiv 1 \bmod N,\ b \equiv c \equiv 0 \bmod N \right\} \\
\Gamma_1(N) &:= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : a \equiv d \equiv 1 \bmod N,\ c \equiv 0 \bmod N \right\} \\
\Gamma_0(N) &:= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \bmod N \right\}.
\end{aligned}$$

Matrices in these groups have the shape, respectively

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \bmod N,\ \begin{pmatrix} 1 & \star \\ 0 & 1 \end{pmatrix} \bmod N,\ \begin{pmatrix} \star & \star \\ 0 & \star \end{pmatrix} \bmod N.$$

Note that

$$\Gamma(N) \leq \Gamma_1(N) \trianglelefteq \Gamma_0(N) \leq \mathrm{SL}_2(\mathbb{Z}),\ \Gamma(N) \trianglelefteq \mathrm{SL}_2(\mathbb{Z})$$

and so for each $k \in \mathbb{Z}$ we have

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

The subgroup $\Gamma(N)$ has the very nice property of being normal in $\mathrm{SL}_2(\mathbb{Z})$. This makes many computations about the space $M_k(\Gamma(N))$ easier. Unfortunately though $T \notin \Gamma(N)\,(N > 1)$ and so modular forms in this biggest space do not (all) have a Fourier expansion.

Let $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ be a Dirichlet character; that is, a homomorphism between these two multiplicative groups. We define

$$M_k(N,\chi) := \left\{ f \in M_k(\Gamma_1(N)) : f(\gamma(z)) = \chi(d)\cdot(cz+d)^k \cdot f(z) \text{ for every } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \right\}.$$

This is a $\mathbb{C}$-vector subspace of $M_k(\Gamma_1(N))$. Note that $M_k(N,\chi_1) = M_k(\Gamma_0(N))$ and we shall just denote this space $M_k(N)$.

Theorem 1.8. *We have*

$$M_k(\Gamma_1(N)) = \bigoplus_\chi M_k(N,\chi)$$

where χ runs over all Dirichlet characters $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$.

Proof. Omitted. See Pages 169-170 of [6]. (The spaces $M_k(N,\chi)$ are the different "simultaneous" eigenspaces for the action of a set of commuting and diagonalisable linear operators $D := \{\langle d\rangle \mid d \in (\mathbb{Z}/N\mathbb{Z})^\times\}$ on $M_k(\Gamma_1(N))$: each character χ describes the eigenvalues $\{\chi(d) : \mid d \in (\mathbb{Z}/N\mathbb{Z})^\times\}$ of the operators in D on the eigenspace $M_k(N,\chi)$.) □

The spaces $M_k(N,\chi)$ have an incredibly rich structure and are of profound interest to number theorists (and mathematicians in general). We shall describe some of this, often though focussing on more tractable examples such as $M_k(\mathrm{SL}_2(\mathbb{Z}))$ or $M_k(\Gamma(N))$ to illustrate parts of the general theory in technically simpler settings. In particular, no mention shall be made again of Dirichlet characters.

We shall discuss

- The geometry of modular curves, especially for $\Gamma(N)$, leading to a more geometric understanding of modular forms.
- The finiteness of the dimension for spaces of modular forms, in more tractable cases.
- Eisenstein series in $M_k(\mathrm{SL}_2(\mathbb{Z}))$, as explicit examples of modular forms.
- The Petersson inner product on spaces of modular forms.
- Modular forms in $M_k(\mathrm{SL}_2(\mathbb{Z}))$ as functions on lattices, leading to the definition of the Hecke operators.
- The relationship between the Petersson inner product and Hecke operators for $M_k(1)$, leading to the construction of the canonical basis of eigenforms.

2 Geometry of modular curves

2.1 A fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$

We consider the action of $\mathrm{SL}_2(\mathbb{Z})$ on the upper half plane $\mathfrak{H}$ via linear fractional transformations (LFTs). Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ and recall that for $z \in \mathfrak{H}$,

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

Notice that γ and $-\gamma$ define the same LFT.

Definition 2.1. *For $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ define*

$$\overline{\Gamma} := \Gamma/(\Gamma \cap \{\pm I\}).$$

When $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ we call $\overline{\Gamma}$ the projective special linear group and sometimes denote it $\mathrm{PSL}_2(\mathbb{Z})$.

Working with $\mathrm{PSL}_2(\mathbb{Z})$ gets rid of this problem:

Exercise 2.2 Check that $\mathrm{PSL}_2(\mathbb{Z})$ acts faithfully on $\mathfrak{H}$ via LFTs.

Thus we may identifty $\mathrm{PSL}_2(\mathbb{Z})$ with a group of LFTs.

Suppose that $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma \neq \pm I_2$ fixes some point $z \in \mathfrak{H}$ (one calls such a transformation *elliptic*). We find

$$\frac{az+b}{cz+d} = z, \text{ so } cz^2 + (d-a)z - b = 0.$$

Since $\mathrm{Im}(z) > 0$ we have $\bar{z} \neq z$ and this equation has distinct complex roots. Taking the discriminant we find

$$(d-a)^2 + 4bc < 0$$

and since $ad - bc = 1$ this gives us $(a+d)^2 < 4$. Hence $a + d = 0, \pm 1$. When $a + d = 0$ we get $d = -a$ and it follows $\gamma^2 = -I$. Likewise a direct calculation shows that when $a + d = \pm 1$ we have $\gamma^3 = \pm I$.

Let $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$. We wish to understand the geometry of the quotient space $\mathfrak{H}/\Gamma$.

Definition 2.3. *Two points $z, z' \in \mathfrak{H}$ are equivalent under Γ if there exists $\gamma \in \Gamma$ such that $z' = \gamma(z)$. (This is an equivalence relation.) A fundamental domain for Γ is an open set $D \subset \mathfrak{H}$ which does not contain any pair of distinct equivalent points and whose closure $\overline{D} \subset \mathfrak{H}$ contains at least one point from each equivalence class.*

Here the closure refers to the usual topology coming from the Euclidean metric on $\mathbb{C}$.

Our first task is to understand the quotient $\mathfrak{H}/\mathrm{SL}_2(\mathbb{Z})$ by finding a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$.

Lemma 2.4. *For a given point $z = x + iy \in \mathfrak{H}$ there are only finitely many pairs of integers (c, d) such that $|cz + d| \leq 1$.*

Proof. Let (c, d) be such a pair; then

$$|cz + d|^2 = (cx + d)^2 + c^2y^2,$$

so that

$$c^2y^2 \leq (cx + d)^2 + c^2y^2 \leq 1.$$

Since $z \in \mathfrak{H}$, $y > 0$; then

$$|c| \leq \frac{1}{y}$$

and hence there are only a finite number of possible values for c. For each such value of c the equation

$$(cx + d)^2 + c^2y^2 \leq 1$$

shows there are only finitely many possible values of d. □

But which matrices in $\mathrm{SL}_2(\mathbb{Z})$ have a fixed (c, d) as their bottom row? (Note we need $\gcd(c, d) = 1$.)

If $(c, d) = (0, \pm 1)$ we just have $\pm T^k$ for some $k \in \mathbb{Z}$. Given $(c, d) \in \mathbb{Z}^2 \backslash (0, \pm 1)$ with $\gcd(c, d) = 1$ we can find unique $0 \leq a < |c|$ with $ad \equiv 1 \bmod |c|$, and for this a we have a unique $b \in \mathbb{Z}$ with $ad - bc = 1$. All matrices $\begin{pmatrix} A & B \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ can then be described as

$$\begin{pmatrix} A & B \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = T^k \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

for some $k \in \mathbb{Z}$. Note that T^k just translates a point horizontally by k.

It is convenient to call $y = \mathrm{Im} z$ the **height** of $z = x + iy$.

Lemma 2.5. *For $z \in \mathfrak{H}$ the set of heights $\{\mathrm{Im}(\gamma z) \,|\, \gamma \in \mathrm{SL}_2(\mathbb{Z})\}$ has a maximum.*

Proof. For any $z \in \mathfrak{H}$ and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$,

$$\gamma(z) = \frac{az + b}{cz + d} = \frac{az + b}{cz + d} \cdot \frac{c\bar{z} + d}{c\bar{z} + d} = \frac{\star + i(ad - bc)\mathrm{Im}(z)}{|cz + d|^2}$$

where $\star := ac|z|^2 + bd + x(ad + bc) \in \mathbb{R}$, so that

$$\mathrm{Im}(\gamma(z)) = \frac{\mathrm{Im}(z)}{|cz + d|^2}. \tag{1}$$

The desired result now follows from Lemma 2.4. □

Lemma 2.5 suggest we select from each equivalence class an element of maximum height, i.e., a point z such that $|cz+d| \geq 1$ for all integer pairs c, d. Since the translation $T : z \mapsto z+1$ is in $\mathrm{PSL}_2(\mathbb{Z})$, we can further assume the fundamental domain lies in the strip $|\Re(z)| = |x| \leq \frac{1}{2}$.

Theorem 2.6. *A fundamental domain for* $\mathrm{SL}_2(\mathbb{Z})$ *is the set*

$$\mathcal{D} := \left\{ z \in \mathfrak{H} \,:\, |\Re(z)| < \frac{1}{2} \text{ and } |z| > 1 \right\}.$$

Proof. We first show that D is the same set as

$$D_1 := \left\{ z \in \mathfrak{H} \,:\, |\Re(z)| < \frac{1}{2} \text{ and } |cz+d| > 1 \text{ for all } (c,d) \in \mathbb{Z}^2 \backslash \{(0,0),(0,\pm 1)\} \right\}.$$

Setting $c = 1, d = 0$ shows $D_1 \subseteq D$. Conversely, suppose $z \in D$. Then for $c \neq 0$,

$$|cz+d|^2 = (cx+d)^2 + c^2y^2 = c^2(x^2+y^2) + 2cdx + d^2 > c^2 - |cd| + d^2$$

and the latter is ≥ 1. (This is certainly true when $d = 0$. For $d \neq 0$, the final expression is invariant under replacing c by $-c$, so we may assume that $cd > 0$ and then $c^2 - |cd| + d^2 = c^2 - cd + d^2 > (c-d)^2 \geq 0$.) When $c = 0$ we have $|cz+d| = |d|$ which is > 1 provided $d \neq \pm 1$. Hence $z \in D_1$, and so $D = D_1$.

By our preceding remark the closure of D_1 contains at least one point from each equivalence class; that is, take a point of maximum height in the class, and shift it by a power of T so that $|\Re(z)| \leq \frac{1}{2}$. (Note that in the closure the second inequality becomes $|cz+d| \geq 1$.)

Suppose now that $z, z' \in D$ with $z' = \gamma(z)$ for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. Since $D = D_1$ then by (1) we have $\mathrm{Im}(z') < \mathrm{Im}(z)$. Since also $z = \gamma^{-1}(z_1)$ we have $\mathrm{Im}(z) < \mathrm{Im}(z')$, a contradiction. So no two distinct points in D are equivalent. □

We next show that the only pairs of points of the closure of D which are equivalent under $\mathrm{SL}_2(\mathbb{Z})$ are the pairs of points of the boundary which coincide upon reflection about the line $x = 0$. These points are identified by the transformations

$$T : z \mapsto z+1 \text{ and } S : z \mapsto -\frac{1}{z}.$$

Suppose $z, z' \in \overline{D}$ with $z' = \gamma(z)$ for some $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. Then $\mathrm{Im}(z) = \mathrm{Im}(\gamma(z))$. (For if $\mathrm{Im}(\gamma(z)) > \mathrm{Im}(z)$, say, then $|cz+d| < 1$, which contradicts the fact that $z \in \overline{D} = \overline{D_1}$.) So

$$1 = |cz+d|^2 \geq c^2 - |cd| + d^2 \geq 1 \text{ (middle and final inequalities from above).}$$

So these inequalities are equalities. Hence either $c = 0, d = \pm 1$ (any z) or $d = 0, c = \pm 1$ ($|z| = 1$ only), or $c = d = \pm 1$ (here $|z+1| = 1$ and so $z = e^{2\pi i/3}$), or $c = -d = \pm 1$ (here $|z-1| = 1$ and so $z = e^{\pi i/3}$). Our earlier discussion shows the first case gives $\gamma = T^k$ for some k, but such a γ can only identify points in $\overline{D}$ if $\gamma = T^{\pm 1}$ and the points z, z' are on the vertical boundary lines. In the second case $\gamma = T^k S$ for some k, but since $|z| = 1$ and $z' = \gamma(z) \in \overline{D}$ we must have $\gamma = S$

and so also $|z'| = 1$. (Or in fact $\gamma = TS$ works too when $z = e^{\pi i/3}$ and we have then $z' = z$.). Finally, in the last two cases we must have $z = e^{\pi i/3}$ or $e^{2\pi i/3}$ and thinking about heights we see then $z' = e^{2\pi i/3}$ or $e^{\pi i/3}$, respectively.

Our analysis has shown that the only fixed points of non-trivial transformations of $\mathrm{SL}_2(\mathbb{Z})$ which lie in $\overline{D}$ are the points

$$i, \rho \text{ and } \rho^2 \text{ where } \rho = e^{\pi i/3} \text{ and } i = \sqrt{-1}.$$

These points are fixed under the elliptic transformations

$$S \text{ (order 2) and } TS : z \mapsto \tfrac{z-1}{z} \text{ (order 3) and } ST : z \mapsto \tfrac{-1}{z+1} \text{ (order 3)}$$

respectively. They are called the **elliptic points** in $\overline{D}$. More precisely, one says that the group $\mathrm{SL}_2(\mathbb{Z})$ has two elliptic points (the equivalence classes of) i and ρ, respectively.

Note 2.7 The transformations S and T generate $\mathrm{PSL}_2(\mathbb{Z})$, satisfying the relations $S^2 = (TS)^3 = I$. See [2, Page 5 Figure 1] for a drawing of D and its translates under some elements in $\mathrm{PSL}_2(\mathbb{Z})$.

2.2 The modular curves $Y(1)$ and $X(1)$ (sketch)

We define the open and compact modular curves as Riemann surfaces.

Definition 2.8. *The (open) modular curve of level* 1 *is* $Y(1) := \mathfrak{H}/\mathrm{PSL}_2(\mathbb{Z})$ *with the quotient topology.*

Thus points in $Y(1)$ are equivalence classes of points in the upper half plane $\mathfrak{H}$ under the modular group $\mathrm{PSL}_2(\mathbb{Z})$. Let $\tau : \mathfrak{H} \to Y(1)$ be the natural map. Then by definition open sets $U \subseteq Y(1)$ are just those for which $\tau^{-1}(U)$ is open; put another way, we give $Y(1)$ the finest topology for which τ is continuous. Note that $Y(1)$ is connected since it is the continuous image of a connected set.

Pictorially we can just think of $Y(1)$ as the closed fundamental domain $\overline{D}$ with appropriate identification of points along the boundary. It is clear from this description that $Y(1)$ is Hausdorff; see [1, Proposition 2.1.1, Corollary 2.1.2] for a more formal proof.

Next we put a complex structure on $Y(1)$. To do this we need local coordinates at each point $\tau(z) \in Y(1)$, i.e., find a neighbourhood $\tilde{U}$ of $\tau(z)$ and a homeomorphism $\phi : \tilde{U} \to V \subset \mathbb{C}$ such that the transition maps between local coordinate systems are holomorphic.

Around any point $z \in \overline{D}$ which is not a fixed point we can draw a small disk U not containing any elliptic points which is mapped homeomorphically onto an open neighbourhood $\tau(U)$ of $\tau(z)$ in $Y(1) = \mathfrak{H}/\mathrm{PSL}_2(\mathbb{Z})$. The local inverse $\phi : \tau(U) \to U$ gives our local coordinate.

The elliptic points must be treated separately. For $z_0 \in \mathfrak{H}$ define $\delta_{z_0} : z \mapsto \frac{z-z_0}{z-\overline{z_0}}$. This maps z_0 to zero and $\overline{z_0}$ to ∞. For h a positive integer let $s_h : z \mapsto z^h$ be the hth power map. Around $i \in \overline{D}$ we consider the map

$$\phi := s_2 \circ \delta_i : z \mapsto \left(\frac{z-i}{z+i}\right)^2$$

on a small "half-disk" around i not containing any other elliptic points. See [2, Page 7 Figure 2] or [1, Page 50 Figure 2.2] for a picture. The image of this map is a true disk $V \subset \mathbb{C}$ around zero. This gives the required local coordinate map. Around $\rho = e^{\pi i/3}$ we use the same procedure, but using a "one third disk" and the cubing map. See [1, Page 48-52] for detailed proofs, including that the transition maps are holomorphic.

Definition 2.9. *The* **extended upper half plane** *is defined as* $\mathfrak{H}^\star := \mathfrak{H} \cup \mathbb{Q} \cup \{\infty\}$.

Geometrically, we think of the points $\mathbb{Q}$ as lying along the real axis, and the point ∞ as lying infinitely far up the imaginary axis.

Lemma 2.10. *The group* $\mathrm{SL}_2(\mathbb{Z})$ *acts on* $\mathfrak{H}^\star$ *via LFTs.*

Proof. We already know $\mathrm{SL}_2(\mathbb{Z})$ acts on $\mathfrak{H}$. So let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ and consider a point $z = m/n \in \mathbb{Q} \cup \infty$ with $\gcd(m,n) = 1$ and $(m,n) \neq (0,0)$. Observe that here we are formally writing $\infty = 1/0$. Then

$$\gamma(z) = \frac{a(m/n)+b}{c(m/n)+d} = \frac{am+bn}{cm+dn} \in \mathbb{Q} \cup \{\infty\}.$$

Again this is a group action, giving an action on $\mathfrak{H}^\star$ in which both $\mathfrak{H}$ and $\mathbb{Q} \cup \{\infty\}$ are fixed as sets. □

It is easy to show that $\mathbb{Q} \cup \{\infty\}$ forms a single orbit under this action. We put a topology on $\mathfrak{H}^\star$ by taking a basis of open sets around ∞ to be $S_\varepsilon := \{z \,|\, \mathrm{Im}(z) > 1/\varepsilon\} \cup \{\infty\}$ $(\varepsilon > 0)$, and around a point P to be $\gamma(S_\varepsilon)$ where $\gamma(\infty) = P$.

Definition 2.11. *The (compact) modular curve of level* 1 *is* $X(1) := \mathfrak{H}^\star/\mathrm{PSL}_2(\mathbb{Z})$ *with the quotient topology.*

Observe that $\mathrm{PSL}_2(\mathbb{Z})$ acts transitively on $\mathbb{Q} \cup \{\infty\}$ so we have $X(1) = Y(1) \cup [\infty]$ where $[\infty] := \mathbb{Q} \cup \{\infty\}$ here is the orbit of ∞. Following convention we shall just write $X(1) = Y(1) \cup \{\infty\}$, and we call ∞ the infinite **cusp**. (Beware we are using "∞" in two different ways here: as an element in $\mathfrak{H}^\star$ and the equivalence class it lies in.)

Thinking again of $Y(1)$ as $\overline{D}$ with edges identified, a basis of open sets around ∞ can be taken to be its intersections with the sets S_ε.

To put a complex structure on $X(1)$ we must define a local coordinate function at the cusp. The set $\{z \in \mathfrak{H} : \mathrm{Im}(z) > 1\}$ is mapped by $\phi : z \mapsto q := e^{2\pi i z}$ onto the punctured disk $0 < |q| <$

$e^{-2\pi}$. Notice for fixed x as y tends to $+\infty$, $\arg(\phi(z))$ remains constant while $|\phi(z)|$ approaches 0. Moreover, two points z, z' of the set are mapped into the same point only if $z' = z + m$ for some integer m; but these points are the same in $Y(1) = \mathfrak{H}/\mathrm{PSL}_2(\mathbb{Z})$. Thus this maps descends (and extends) to a well-defined map on an open neighbourhood of ∞ in $X(1)$ to a open disk around zero in $\mathbb{C}$, as needed.

The resulting Riemann surface is denoted $X(1)$ is easily seen to be a sphere; for example, by using the natural triangulation of $\overline{D}$ in [2, Page 5 Figure 1], with vertices ρ, i, ρ^2 and ∞.

Thus we have given all the key ideas behind the proof of the following theorem.

Theorem 2.12. *The quotient space $X(1) = \mathfrak{H}^\star/\mathrm{PSL}_2(\mathbb{Z})$ can be given a natural complex structure under which it is a compact Riemann surface of genus 0.*

For a detailed treatment and further pictures see [1, Chapter 2].

2.3 Riemann surfaces X_Γ and their genus

Let $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ be a subgroup of the modular group of *finite index*. We shall find a fundamental domain for Γ, which can be compactified and made into a Riemann surface.

2.3.1 Riemann surfaces Y_Γ and X_Γ (sketch)

We first describe our Riemann surfaces.

Theorem 2.13. *Let $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ with $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma] < \infty$ and define*

$$\mu := [\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}] = \begin{cases} [\mathrm{SL}_2(\mathbb{Z}) : \Gamma] & \text{if } -I \in \Gamma \\ \frac{1}{2}[\mathrm{SL}_2(\mathbb{Z}) : \Gamma] & \text{if } -I \notin \Gamma. \end{cases}$$

Select coset representatives $T_1, \cdots, T_\mu$ so that

$$\mathrm{PSL}_2(\mathbb{Z}) = \overline{\Gamma} T_1 \cup \cdots \cup \overline{\Gamma} T_\mu.$$

If D is a fundamental domain for $\mathrm{PSL}_2(\mathbb{Z})$ then

$$D_\Gamma := T_1 D \cup \cdots \cup T_\mu D$$

is a fundamental domain for Γ.

Proof. Exercise. □

Example 2.14 For $\Gamma(2)$ we have $\mathrm{PSL}_2(\mathbb{Z})/\overline{\Gamma(2)} \cong \mathrm{SL}(2,\mathbb{Z}_2)$ and coset representatives can be taken as

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

$$TST = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, TSTS = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

See [2, Page 16 Figure 5]. Gluing along edges can now be done using T^2, ST^2S and TST^2ST^{-1}, as in the picture, again giving visually a Riemann surface of genus 0.

The quotient $Y_\Gamma := \mathfrak{H}/\Gamma$ may be given a complex structure. Let $\tau_\Gamma(\mathfrak{H}) \to Y_\Gamma$ be the natural map.

Definition 2.15. *An* **elliptic point** $z \in \mathfrak{H}$ *for* Γ *is one for which the stabiliser* $\overline{\Gamma}_z := \{\gamma \in \overline{\Gamma} \,|\, \gamma(z) = z\}$ *is non-trivial. The corresponding point* $\tau_\Gamma(z) \in Y_\Gamma$ *is also called elliptic. The* **period** *of the elliptic point is defined as the order of the stabiliser (necessarily* 2 *or* 3*).*

Note elliptic points $z \in \mathfrak{H}$ for Γ must be $\mathrm{SL}_2(\mathbb{Z})$-translates of i or ρ. For such z, $|\overline{\Gamma}_z| = |\mathrm{PSL}_2(\mathbb{Z})_z \cap \overline{\Gamma}| = 2$ or 3, e.g., if $z = L(i)$ for some $L \in \mathrm{PSL}_2(\mathbb{Z})$ then $\overline{\Gamma}_z = L \cdot \mathrm{PSL}_2(\mathbb{Z})_i \cdot L^{-1} \cap \overline{\Gamma}$ has order dividing $2 = |\mathrm{PSL}_2(\mathbb{Z})_i| = |\langle S \rangle|$.

The local coordinate around images of points in $\mathfrak{H}$ is treated as before, depending upon whether or not the point is elliptic. The local coordinate around the cusp at infinity is defined using the map $z \mapsto e^{2\pi i z/h}$ where h, the **width** of the cusp, is pictorially the number of copies of D which meet at infinity. More formally:

Definition 2.16. *The width of the infinite cusp for* $\overline{\Gamma}$ *it is the least positive integer* h *such that the translation* $z \mapsto z + h$ *lies in* $\overline{\Gamma}$*, or equivalently the index* $[\mathrm{PSL}_2(\mathbb{Z})_\infty : \overline{\Gamma}_\infty]$ *of the stabiliser* $\overline{\Gamma}_\infty := \{L \in \overline{\Gamma} \,|\, L(\infty) = \infty\}$ *in* $\mathrm{PSL}_2(\mathbb{Z})_\infty = \langle T \rangle$*.*

To compactify one may need to add further finite cusps, which are rational points on the real axis, as well as the cusp at infinity. That is, when the fundamental domain contains a transform of D which touches the real axis. More formally with $\mathfrak{H}^\star = \mathfrak{H} \cup \mathbb{Q} \cup \{\infty\}$ we define

$$X_\Gamma := \mathfrak{H}^\star/\Gamma.$$

Definition 2.17. *A* **cusp** *is an orbit for the action of* Γ *on the set* $\mathbb{Q} \cup \{\infty\}$*.*

This fits with the intuitive description. The finite cusps $[\alpha]$ are treated in a similar manner to the infinite one. That is, one transforms the topology and complex structure (and notion of width) around the infinite cusp by $\gamma \in \mathrm{PSL}_2(\mathbb{Z})$ with $\gamma(\infty) = \alpha$. So again the width is the index of the stabiliser. (See [1, Chapter 2] for full details.) This makes X_Γ a compact Riemann surface. As before, pictorially it can be thought of as $\overline{D_\Gamma}$ with appropriate identification around the edges, cf. Example 2.14.

2.3.2 A triangulation of X_Γ

As mentioned before, there is a natural triangulation of $\overline{D}$ in which ρ, i, ρ^2 and ∞ are the vertices, see [2, Page 5 Figure 1]. Identifying edges appropriately gives a triangulation of $X(1)$: one triangle with vertices ρ, i and ∞. Our triangulation of $\overline{D}$ induces one of $\overline{D_\Gamma}$ by taking transformations under the coset representatives for $\overline{\Gamma}$ in $\mathrm{PSL}_2(\mathbb{Z})$, and thus one for the compact Riemann surface X_Γ by identifying edges appropriately, cf. Example 2.14.

We compute the genus of this surface by means of this triangulation and the Euler characteristic formula

$$\chi = 2 - 2g = \sigma_0 - \sigma_1 + \sigma_2 \tag{2}$$

where χ is the Euler characteristic, g is the genus (number of "holes" in the Riemann surface), and σ_k is the number of k-simplexes in the triangulation (vertices, edges or faces). (See Page 47 in the Part B Algebraic Curves notes for the formula "$\chi = 2 - 2g$ where $g \geq 0$ is the genus of a Riemann surface". That the Euler characteristic can be computed using any triangulation is stated without proof as Theorem 4.2 in Part B Geometry of Surfaces.)

Fixing a choice of coset representatives, in our natural triangulation of X_Γ, σ_0 is the number of images of elliptic and cuspidal points of $\mathrm{PSL}_2(\mathbb{Z})$. We write

$$\sigma_0 = \lambda_i + \lambda_\rho + \lambda_\infty$$

where λ_i ($\lambda_\rho, \lambda_\infty$ respectively) is the number of vertices equivalent to i (ρ, ∞ respectively) under $\mathrm{PSL}_2(\mathbb{Z})$.

Let $p_1, \cdots, p_{\sigma_0}$ be the vertices of the triangulation, the first λ_i being equivalent (under $\mathrm{PSL}_2(\mathbb{Z})$) to i, the next λ_ρ to ρ, and so on for ∞. To find out how many edges meet at a typical vertex we distinguish various cases:

(a) If p_k is equivalent to i then two or four edges meet at p_k according as whether it is a fixed point for $\overline{\Gamma}$ or not, i.e., is an elliptic point for $\overline{\Gamma}$ or not. (The stabiliser of i in $\mathrm{PSL}_2(\mathbb{Z})$ is $\langle S \rangle$, and hence the stabiliser of a translate of i by an element of $\mathrm{PSL}_2(\mathbb{Z})$ is a conjugate subgroup of $\langle S \rangle$ in $\mathrm{PSL}_2(\mathbb{Z})$. Note that when $\overline{\Gamma}$ is normal in $\mathrm{PSL}_2(\mathbb{Z})$, all points equivalent to i will or will not have a non-trivial stabiliser according to whether $S \in \overline{\Gamma}$ or not.).

(b) If p_k is equivalent to ρ then two or six edges meet at p_k, according to whether it is a fixed point for $\overline{\Gamma}$ or not, i.e., is an elliptic point for $\overline{\Gamma}$ or not, i.e., the appropriate conjugate of TS is in $\overline{\Gamma}$ or not.

(c) If p_k is equivalent to ∞ then, if it compactifies n transforms of the fundamental domain D (for $\mathrm{SL}_2(\mathbb{Z})$), then $2n$ edges meet there.

This analysis is clarified pictorial by drawing the triangulation for Example 2.14. Observe here $S \notin \overline{\Gamma(2)}$ and $TS \notin \overline{\Gamma(2)}$ and one can see that 4 and 6 edges emanate from i and ρ, respectively, and each cusp compactifies 2 transforms of D. In any case an even number $2n_k$, say, meet at p_k.

Theorem 2.18. *The genus of X_Γ is*

$$g = 1 + \frac{1}{2}(\mu - \sigma_0). \tag{3}$$

Proof. It is sufficient, from formula (2), to find σ_1 and σ_2. Since $\overline{\Gamma}$ is of index μ the fundamental domain D_Γ consists of $2\mu = \sigma_2$ faces in the standard triangulation (there are two in the triangulation of D). The number of edges is

$$\sigma_1 = \frac{1}{2}\sum_{k=1}^{\sigma_0}(2n_k) = \sum_{k=1}^{\sigma_0} n_k \tag{4}$$

i.e., the total number of edges emanating from vertices divided by 2 (since each is counted twice). We break (4) into three sums[1]:

$$\sigma_1 = \sum_{k=1}^{\lambda_i} n_k + \sum_{k=\lambda_i+1}^{\lambda_i+\lambda_\rho} n_k + \sum_{k=\lambda_i+\lambda_\rho+1}^{\sigma_0} n_k \tag{5}$$

corresponding to points equivalent to i, ρ and ∞. We claim that each sum is equal to μ, i.e., $\sigma_1 = 3\mu$. We show this for the first sum only, the argument being similar for the other two. In any face there is one vertex equivalent to i, and there are two edges of the triangle having that point as a common vertex. Each edge belongs to two faces, and there are 2μ faces, so that a total of 2μ edges emanate from points equivalent to i. But since no edge connects two points equivalent to i this number is also $\sum_{k=1}^{\lambda_i}(2n_k)$; hence $\mu = \sum_{k=1}^{\lambda_i} n_k$.

Substituting $\sigma_1 = 3\mu$ and $\sigma_2 = 2\mu$ in (2) now proves the theorem. □

2.3.3 The genus for normal subgroups

We now deduce a simpler formula in the case in which $\overline{\Gamma}$ is a **normal** subgroup of $\mathrm{PSL}_2(\mathbb{Z})$. In this case, all vertices of X_Γ equivalent to i under $\mathrm{PSL}_2(\mathbb{Z})$ have the *same* number of edges meeting there. (There are two edges meeting at each such vertex if $S \in \overline{\Gamma}$ and four edges otherwise.) The same is true for ρ and ∞.

Let $2n(i), 2n(\rho)$ and $2n(\infty)$ be the number of edges meeting at typical points equivalent, respectively, to i, ρ and ∞. Recall we have the following conditions

$$\begin{aligned} n(i) &= 1 \text{ or } 2; \\ n(\rho) &= 1 \text{ or } 3; \\ n(\infty) &= \text{any positive integer.} \end{aligned} \tag{6}$$

If we use the fact, verified during the proof of Theorem 2.18, that

$$\sum_{k=1}^{\lambda_i} n_k = \sum_{k=\lambda_i+1}^{\lambda_i+\lambda_\rho} n_k = \sum_{k=\lambda_i+\lambda_\rho+1}^{\sigma_0} n_k = \mu$$

[1] In fact, we can see $\sigma_1 = 3\mu$ directly by observing every face has three edges, every edge appears on two faces, and there are 2μ faces. This more detailed analysis is helpful later on though.

we obtain

$$\lambda_i n(i) = \lambda_\rho n(\rho) = \lambda_\infty n(\infty) = \mu \tag{7}$$

and

$$\sigma_0 = \lambda_i + \lambda_\rho + \lambda_\infty = \mu\left(\frac{1}{n(i)} + \frac{1}{n(\rho)} + \frac{1}{n(\infty)}\right).$$

This proves

Theorem 2.19. *Let $\overline{\Gamma}$ be a normal subgroup of finite index μ in $\mathrm{PSL}_2(\mathbb{Z})$. Then X_Γ is a compact Riemann surface of genus*

$$g = 1 + \frac{1}{2}\mu\left(1 - \frac{1}{n(i)} - \frac{1}{n(\rho)} - \frac{1}{n(\infty)}\right).$$

By Theorem 2.19 knowledge of $(n(i), n(\rho), n(\infty))$ is enough to determine the genus of X_Γ in this case. Formula (6) shows there are four possibilties

$$(n(i), n(\rho), n(\infty)) = (1,1,n), (1,3,n), (2,1,n), (2,3,n).$$

However, we have in fact.

Lemma 2.20.

$$(n(i), n(\rho), n(\infty)) = (1,1,1), (2,1,2), (1,3,3) \tag{8}$$

or $(2,3,n(\infty))$ for some positive integer $n(\infty)$.

Proof. It is easy to see that any $(n(i), n(\rho), n(\infty))$ not of the form $(2,3,n)$ must be those described in (8). For example if $(n(i), n(\rho), n(\infty)) = (2,1,n)$ then, by Theorem 2.19, $g = 1 - \frac{\mu}{4n}(n+2)$. Since all these quantities on the RHS are positive integers, we must have $g = 0$ and $n = \frac{2\mu}{4-\mu}$. Since n is a positive integer and divides μ, by (7), we must have $n = \mu = 2$. □

There are unique normal subgroups with each triple $(n(i), n(\rho), n(\infty))$ on the list (8); namely, $\mathrm{PSL}_2(\mathbb{Z})$, the subgroup of index 2 generated by squares, and that of index 3 generated by cubes, see our examples below and [2, Chapter 1, Theorem 7].

2.4 Examples

Example 2.21 We show there is a unique (necessarily) normal subgroup of $\mathrm{PSL}_2(\mathbb{Z})$ of index 2. Given such a subgroup $\overline{\Gamma}$ we have $\mathrm{PSL}_2(\mathbb{Z})/\overline{\Gamma} \cong \mathbb{Z}_2 (= \{0,1\})$. Any homomorphism $h : \mathrm{PSL}_2(\mathbb{Z}) \to \mathbb{Z}_2$ is determined by what it does to the generators S and T, and must be compatible with the relations $(TS)^3 = S^2 = I$. The only non-trivial one is thus $h(T) = h(S) = 1$, and the kernel is the unique normal subgroup, $\overline{\Gamma}_2$ say. We can take $D \cup TD$ as a fundamental domain, and the elements

$$T^2 : z \mapsto z+2$$
$$TS = (ST^{-1})^2 : z \mapsto \tfrac{z-1}{z}$$

lie in $\overline{\Gamma}_2$ and so can be used to "glue" the edges. See [2, Page 14 Figure 3]. From the picture we see that 4 edges emanate from i, and 2 from ρ and 2 fundamental domains meet at ∞. Thus $n(i) = 2$, $n(\rho) = 1$ and $n(\infty) = 2$ giving genus

$$1 + \frac{1}{2} \cdot 2 \left(1 - \frac{1}{2} - 1 - \frac{1}{2} \right) = 0$$

which should agree with your intuition here.

Example 2.22 By a similar analysis there is a unique normal subgroup $\overline{\Gamma}_3$ of $\mathrm{PSL}_2(\mathbb{Z})$ of index 3, with corresponding homomorphism to $\mathbb{Z}_3 = \{0, 1, 2\}$ given by $S \mapsto 0$ and $T \mapsto 1$. We can take $T^{-1}D \cup D \cup TD$ as a fundamental domain and since the transformations

$$T^3, (T^{-1}ST) = (T^{-1}ST)^3, S = S^3, TST^{-1} = (TST^{-1})^3$$

lie in $\overline{\Gamma}_3$ they can be used to "'glue" the edges. See [2, Page 15 Figure 4]. Thus $(n(i), n(\rho), n(\infty)) = (1, 3, 3)$ and the genus is zero.

2.5 The genus of the modular curves $X(N), X_1(N)$ and $X_0(N)$

We write

$$X(N) := X_{\Gamma(N)},\ X_1(N) := X_{\Gamma_1(N)},\ X_0(N) = X_{\Gamma_0(N)}.$$

2.5.1 Index computations

We compute $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)]$ for $N \in \mathbb{N}$.

Proposition 2.23. *The natural map*

$$\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$$

is surjective.

Proof. Let

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

Lifting $\alpha, \beta, \gamma, \delta$ arbitrarily to $a, b, c, d \in \mathbb{Z}$ we see that $ad - bc \equiv 1 \bmod N$. Observe first that $\gcd(c, d, N) = 1$ since any prime dividing c, d and N would divide 1. We claim that there exists $n \in \mathbb{Z}$ such that $\gcd(c, d + nN) = 1$. To see this, let n be a solution to the systems of congruences

$$\begin{array}{ll} n \equiv 0 \bmod p, & \text{for primes } p|c \text{ with } p \not| d \\ n \equiv 1 \bmod p & \text{for primes } p|c \text{ with } p|d. \end{array}$$

Let p be a prime with $p|c$. If $p \not| d$ then $p|n$ so $p \not| d + nN$. If $p|d$ then $p \not| n$ and so $p \not| d + nN$. (Note p cannot divide all of c, d and N.) Thus in either case $p \not| d + nN$, which proves the claim.

Now setting $D := d + nN$ we have $\gcd(c, D) = 1$ with $D \equiv d \bmod N$. We also know that $aD - bc \equiv 1 \bmod N$. We want to solve

$$(a + kN)D - c(b + \ell N) = 1$$

for some $k, \ell \in \mathbb{Z}$. Since $aD - bc = 1 + zN$ for some $z \in \mathbb{Z}$ we find we need $kND - c\ell N = -zN$. Cancelling, this gives $kD - c\ell = -z$, which indeed has a solution since $\gcd(c, D) = 1$.

Taking now $A := a + kN$ and $B = b + \ell N$ gives the required matrix

$$\begin{pmatrix} A & B \\ c & D \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$$

mapping to

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}).$$

□

Corollary 2.24.

$$[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)] = |\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})|.$$

Proof. Follows since $\Gamma(N)$ is the kernel of the reduction map. □

Proposition 2.25. *Let $N = \prod_{i=1}^{r} p_i^{\alpha_i}$ be the prime power factorisation. Then*

$$|\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})| = \prod_{i=1}^{r} |\mathrm{SL}_2(\mathbb{Z}/p^{\alpha_i}\mathbb{Z})|.$$

Proof. Consider the natural reduction map

$$\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \to \prod_{i=1}^{r} \mathrm{SL}_2(\mathbb{Z}/p^{\alpha_i}\mathbb{Z}),\ \gamma \to (\gamma \bmod p_1^{\alpha_1}, \cdots, \gamma \bmod p_r^{\alpha_r}).$$

This is a bijection of sets by the Chinese remainder theorem (and even a homomorphism) and so the two sides have the same size. (To see it is surjective, first solve 4 CRT problems to construct a 2×2 matrix over $\mathbb{Z}/N\mathbb{Z}$ with the correction reduction, and then observe this matrix has determinant which is 1 mod $p_i^{\alpha_i}$ for all $1 \le i \le r$, and hence 1 mod N and so lies in $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$.) □

Proposition 2.26. *For p prime and $\alpha \in \mathbb{N}$,*

$$|\mathrm{SL}_2(\mathbb{Z}/p^{\alpha}\mathbb{Z})| = p^{3\alpha}\left(1 - \frac{1}{p^2}\right).$$

Proof. By induction on α. First note that

$$|\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})| = (p^2-1)(p^2-p)$$

since for such matrices there are p^2-1 choices for the first row, and p^2-p for the second. The homomorphism $\det : \mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z}) \to (\mathbb{Z}/p\mathbb{Z})^\times$ is surjective with kernel $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ and so we find

$$|\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})| = \frac{(p^2-1)(p^2-p)}{p-1} = (p+1)p(p-1) = p^3-p = p^3\left(1-\frac{1}{p^2}\right)$$

as required.

For the induction step note that the reduction map $\mathrm{SL}_2(\mathbb{Z}/p^{\alpha+1}\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/p^\alpha\mathbb{Z})$ is surjective (since the map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/p^\alpha\mathbb{Z})$ is, by Proposition 2.23). Elements in the kernel have the form

$$\begin{pmatrix} 1+p^\alpha a & p^\alpha b \\ p^\alpha c & 1+p^\alpha d \end{pmatrix}$$

where

$$(1+p^\alpha a)(1+p^\alpha d) - p^{2\alpha}cd \equiv 1 \bmod p^{\alpha+1};$$

that is $a+d \equiv 0 \bmod p$. Thus the kernel has size p^3, from which the induction step follows. □

Theorem 2.27. *For $N \in \mathbb{N}$ we have*

$$[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)] = N^3 \prod_{p|N}\left(1-\frac{1}{p^2}\right).$$

Proof. From Propositions 2.25 and 2.26. □

Corollary 2.28. *For $N > 2$ we have*

$$[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(N)}] = \frac{1}{2}[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)] = \frac{1}{2}N^3 \prod_{p|N}\left(1-\frac{1}{p^2}\right),$$

and

$$[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(2)}] = [\mathrm{SL}_2(\mathbb{Z}) : \Gamma(2)] = 8 \cdot \left(1-\frac{1}{4}\right) = 6.$$

Proof. Follows since $-I \notin \Gamma(N)$ for $N > 2$, whereas $-I \in \Gamma(2)$ and $-I \in \mathrm{SL}_2(\mathbb{Z})$. □

2.5.2 The genus for modular curves $X(N)$

We now state our main theorem, giving the genus of the modular curves $X(N)$, the compactification of the open modular curves $Y(N) := \mathfrak{H}/\Gamma(N)$.

Theorem 2.29. *The genus of the modular curve $X(N)$ is*

(a) $g = 0$ *if* $N = 2$,

(b) $g = 1 + \frac{N^2(N-6)}{24} \prod_{p|N} \left(1 - \frac{1}{p^2}\right)$ *if* $N > 2$.

Proof. Let $N \geq 2$. Note that $\overline{\Gamma(N)}$ is normal in $\mathrm{PSL}_2(\mathbb{Z})$ and so it is enough to compute the number of edges meeting at i, ρ and ∞. We have $S, TS \notin \overline{\Gamma(N)}$. Hence by our analysis immediately prior to the statement of Theorem 2.18, 4 edges meet at i and 6 at ρ, and so our $(n(i), n(\rho), n(\infty))$ is of the form $(2, 3, n(\infty))$. By Theorem 2.19 the genus is

$$g = 1 + \frac{\mu(n(\infty) - 6)}{12n(\infty)}.$$

It remains to compute $n(\infty)$. This is the width of the cusp at infinity; that is, the number of inequivalent powers of T in the quotient $\mathrm{PSL}_2(\mathbb{Z})/\overline{\Gamma(N)}$, each power adding another translate of the fundamental domain D to $D_{\Gamma(N)}$. (That gives the number of translates of D meeting at the cusp ∞, the number of transformations of D meeting at the other cusps being the same by normality.) But for $s \geq 0$,

$$T^s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$$

and so a complete set of representatives is given by T^s for $s = 0, 1, \cdots, N-1$; thus $n(\infty) = N$. The result now follows from Corollary 2.28. □

2.5.3 The curves $X_1(N)$ and $X_0(N)$

The curves $X_1(N)$ and $X_0(N)$ are more difficult since the subgroups $\overline{\Gamma_1(N)}$ and $\overline{\Gamma_0(N)}$ are not normal in $\mathrm{PSL}_2(\mathbb{Z})$. Using the results from the problem sheets we have the following partial result.

Proposition 2.30. *Let* $p \equiv 11 \bmod 12$ *be prime. Then the genus of* $X_0(p)$ *is* $(p+1)/12$.

Proof. Let p be any odd prime. Then from Sheet 2 Question 2 we have that

$$[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma_0(N)}] = N \prod_{p|N} \left(1 + \frac{1}{p}\right)$$

and so

$$[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma_0(p)}] = p \left(1 + \frac{1}{p}\right) = p + 1.$$

When $p \equiv 11 \bmod 12$ we have by Sheet 1 Question 5(c) that $\Gamma_0(p)$ contains no elliptic transformations. We choose a fundamental domain for $\Gamma_0(p)$ and triangulate it in the manner above, with λ_i vertices equivalent (under $\mathrm{SL}_2(\mathbb{Z})$) to i etc. Then at every vertex equivalent to i there are 4 edges meeting, and at every vertex equivalent to ρ there are 6 edges meeting. So similar to (7) we find

$$2\lambda_i = 3\lambda_\rho = \mu$$

with here $\mu = [\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma_0(p)}] = (p+1)$. From Sheet 2 Question 3 (a) we know there are two cusps, so $\lambda_\infty = 2$. Hence

$$\sigma_0 = \lambda_i + \lambda_\rho + \lambda_\infty = \frac{p+1}{2} + \frac{p+1}{3} + 2$$

So by Theorem 2.18 we find the genus is

$$1 + \frac{1}{2}(\mu - \sigma_0) = 1 + \frac{1}{2}\left(p + 1 - \frac{p+1}{2} - \frac{p+1}{3} - 2\right) = 1 + \frac{1}{2} \cdot \frac{p-11}{6} = \frac{p+1}{12}.$$

□

For example, $X_0(11)$ has genus 1 and $X_0(23)$ has genus 2.

From Sheet 1 Question 5 (b) we know that $\Gamma_1(N)$ for $N > 3$ has no elliptic transformations, and Question 4 computes the index for this subgroup. Thus Theorem 2.18 again reduces the computation of the genus in that case to the computation of the number of cusps; that is, the number of orbits for the action of $\Gamma_1(N)$ on $\mathbb{Q} \cup \{\infty\}$.

3 Dimensions of spaces of modular forms

Throughout this section (unless stated otherwise) the weight k is **even** and $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ is a subgroup of finite index,

Let $f : \mathfrak{H} \to \mathbb{C}$ be a meromorphic function. Then for any point $Q \in \mathfrak{H}$ we can expand f locally around Q as

$$f(z) = (z-Q)^{\mathrm{ord}_Q(f)} \sum_{j=0}^{\infty} a_j (z-Q)^j,\ a_0 \neq 0$$

where the exponent $\mathrm{ord}_Q(f) \in \mathbb{Z}$ is called the *order* of f at the point Q. Thus if $\mathrm{ord}_Q(f) > 0$, then f vanishes (has a zero) at Q, while if $\mathrm{ord}_Q(f) < 0$ then f has a pole at Q.

Let us further assume that

$$f(\gamma(z)) = (cz+d)^k f(z)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. Let $Q = [\alpha]$ be some cusp for Γ, represented by an element $\alpha \in \mathbb{Q} \cup \{\infty\}$. We suppose that $f(z)$ is "meromorphic at Q", in the sense that after choosing $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with $\gamma(\infty) = \alpha$ we have locally around ∞ an expansion

$$f|_k\gamma(z) = \left(e^{\frac{2\pi i}{h}}\right)^{\mathrm{ord}_Q(f)} \sum_{j=0}^{\infty} a_j \left(e^{\frac{2\pi i z}{h}}\right)^j,\ a_0 \neq 0$$

where the exponent $\mathrm{ord}_Q(f) \in \mathbb{Z}$ is called the *order* of f at the cusp Q. (Here h is the width of the cusp Q.) We call such f a *meromorphic modular form* for Γ of weight k.

Observe that the rules of calculus insist that for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ we have

$$\gamma(dz) := d(\gamma(z)) = d\left(\frac{az+b}{cz+d}\right) = \frac{a(cz+d)-(az+b)c}{(cz+d)^2} \cdot dz = \frac{dz}{(cz+d)^2}.$$

So it makes sense to define

$$\gamma((dz)^{k/2}) := (d(\gamma(z))^{k/2} = (cz+d)^{-k}(dz)^{k/2}.$$

Thus for any meromorphic modular form f for Γ of weight k one should have

$$\gamma(f(dz)^{k/2}) := f(\gamma(z)) \cdot \gamma((dz)^{k/2}) = (cz+d)^k f(z) \cdot (cz+d)^{-k}(dz)^{k/2} = f(z)(dz)^{k/2}.$$

So $f(dz)^{k/2}$ is invariant under Γ. The element $f(dz)^{k/2}$ is called a "$k/2$-fold differential form" on $X_\Gamma = \mathfrak{H}^\star/\Gamma$.

Given any (meromorphic) $k/2$-fold differential ω on X_Γ and any point $P \in X_\Gamma$, one can use the local coordinate at P to define $\mathrm{ord}_P(\omega)$. That is, let t be a local parameter and write locally $\omega(t) = h(t)dt^{k/2}$ for some Laurent series $h(t)$, and now define $\mathrm{ord}_P(\omega)$ to be the order of vanishing of $h(t)$.

Observe that for a differential $\omega := f(z)(dz)^{k/2}$ the order of vanishing of f at a point $Q \in \mathfrak{H}^\star$ and of ω at its image $P \in X_\Gamma$ under the quotient map $\mathfrak{H}^\star \to X_\Gamma$ are *not* defined in the same way, thus need not be equal. (The difference is between the z coordinate on $\mathfrak{H}$, and the t coordinate on X_Γ.)

Lemma 3.1. *Let f be a meromorphic modular forms of even weight k for Γ, and let ω by the corresponding $k/2$-fold differential on X_Γ. Let $Q \in \mathfrak{H}^\star$ map to $P \in X_\Gamma$.*

(a) If Q is a elliptic point of period e then

$$\mathrm{ord}_Q(f) = e\mathrm{ord}_P(\omega) + (k/2)(e-1).$$

(b) If Q is a cusp then

$$\mathrm{ord}_Q(f) = \mathrm{ord}_p(\omega) + (k/2).$$

(c) For all remaining points

$$\mathrm{ord}_Q(f) = \mathrm{ord}_p(\omega).$$

Proof. Omitted and non-examinable: a local analysis at each point. (See [4, Pages 52-53].) □

This lemma is key to proving:

Theorem 3.2. *The dimension of the space of modular forms of* **even** *weight k for Γ is:*

$$\begin{cases} 0 & \text{if } k \le -2, \\ 1 & \text{if } k = 0, \\ (k-1)(g-1) + \sigma \cdot \frac{k}{2} + \sum_i \left[\frac{k}{2} \cdot \left(1 - \frac{1}{e_i}\right)\right] & \text{if } k \ge 2, \end{cases}$$

where g is the genus of X_Γ, σ is the number of cusps, and the sum runs through the elliptic points of X_Γ of period e_i. (Here $[\cdot]$ denotes the integer part.)

Proof. (Non-examinable sketch) Let ω_0 be *any* (non-zero) meromorphic $(k/2)$-fold differential on X_Γ.[2] If ω is any other $(k/2)$-fold differential on X_Γ then the quotient ω/ω_0 is a meromorphic *function* on X_Γ. In fact the map

$$\Theta : f(z) \to F(z) := \frac{f(z)(dz)^{k/2}}{\omega_0}$$

is a bijection between the space of meromorphic modular forms of weight k for Γ and the space of meromorphic functions on X_Γ. We are interested in the dimension of the space of functions $\Theta(M_k(\Gamma))$, and so insist that $F = \Theta(f)$ where f is *holomorphic* on $\mathfrak{H}$ and at the cusps. Applying Lemma 3.1 one finds that $\Theta(M_k(\Gamma))$ consists exactly of those meromorphic functions F on X_Γ such that for $P \in X_\Gamma$:

$$\begin{aligned} \mathrm{ord}_P(F) + \mathrm{ord}_P(\omega_0) + (k/2)\left(1 - \tfrac{1}{e}\right) &\geq 0 \text{ if } P \text{ is an elliptic point of period } e \\ \mathrm{ord}_P(F) + \mathrm{ord}_P(\omega_0) + (k/2) &\geq 0 \text{ if } P \text{ is a cusp} \\ \mathrm{ord}_P(F) + \mathrm{ord}_P(\omega_0) &\geq 0 \text{ at the remaining points.} \end{aligned}$$

The "Riemann-Roch theorem" computes dimensions of spaces of meromorphic functions on a compact Riemann surface defined in this sort of manner, and applying it gives our theorem. □

Note that if $-I \in \Gamma$ then taking $\gamma = -I$ we find that for $f \in M_k(\Gamma)$, with *any* $k \in \mathbb{Z}$, then $f(\gamma(z)) = (-1)^k f(z)$. But also $f(\gamma(z)) = f(z)$ since $\gamma(z) = z$. Hence $f(z) = (-1)^k f(z)$ which forces $f = 0$ when k is odd. Thus odd weight spaces are all zero in this case, e.g., for $\mathrm{SL}_2(\mathbb{Z})$ or $\Gamma_0(N)$.

A different application of Lemma 3.1 gives the useful:

Theorem 3.3 (Valence formula)**.** *Let f be a non-zero (meromorphic) modular form of even weight k for Γ and g be the genus of X_Γ. Then*

$$\sum_Q \left\{ \frac{\mathrm{ord}_Q(f)}{e_Q} - \frac{k}{2}\left(1 - \frac{1}{e_Q}\right)\right\} = k(g-1) + \frac{k}{2}\cdot\sigma$$

where the sum is over representatives Q in $\mathfrak{H}^\star$ for the points in $\mathfrak{H}^\star/\Gamma$, σ is the number of cusps of X_Γ, and for a regular point or cusp Q we define $e_Q := 1$ and an elliptic point $Q \in \mathfrak{H}$ write as usual e_Q for its period.

Proof. (Non-examinable) We shall use here that, counting multiplicity, the total number of zeros and poles of a meromorphic $(k/2)$-fold differential on a compact Riemann surface of genus g is $\frac{k}{2}\cdot(2g-2) = k(g-1)$. (This follows easily from the more fundamental facts that the total number of zeros and poles of a meromorphic function on such a surface is zero, and the total number of zero and poles of a meromorphic 1-fold differential is $2g-2$.) Letting ω be the differential attached to f we have from Lemma 3.1 that

$$\mathrm{ord}_P(\omega) = \frac{\mathrm{ord}_Q(f)}{e_Q} - \frac{k}{2}\left(1 - \frac{1}{e_Q}\right)$$

[2]Take any non-zero function g on $X(1)$—the Riemann sphere–differentiate it to get a differential form dg, and take $\omega_0 := (dg)^{k/2}$.

when Q is not a cusp, and for Q a cusp we find

$$\mathrm{ord}_P(\omega) = \mathrm{ord}_Q(f) - \frac{k}{2}.$$

Summing these equations and using that $\sum_P \mathrm{ord}_P(\omega) = k(g-1)$ gives the required result. $\square$

Example 3.4 For the full modular group $\mathrm{SL}_2(\mathbb{Z})$ we have $g = 0$, $\sigma = 1$ and "distinct" elliptic points have periods 2 (for i) and 3 (for ρ) (recall ρ^2 is just the translate of ρ by T). So $\dim(M_k(\Gamma)) = 0$ for $k < 0$ and for $k \geq 2$ and even the dimension of the space of weight k forms is

$$(k-1)(-1) + \frac{k}{2} + \left[\frac{k}{4}\right] + \left[\frac{k}{3}\right] = \begin{cases} \left[\frac{k}{12}\right] & \text{if } k \equiv 2 \bmod 12 \\ \left[\frac{k}{12}\right] + 1 & \text{if } k \not\equiv 2 \bmod 12 \end{cases}$$

Odd weight spaces all have dimension zero, since $-I \in \mathrm{SL}_2(\mathbb{Z})$.

Example 3.5 For the principal congruence subgroup $\Gamma(N)$ of level $N > 1$ we have $(n(i), n(\rho), n(\infty)) = (2, 3, N)$ (proof of Theorem 2.29). Now in equation (5) the final sum is equal to the index μ; in the notation of this chapter there are σ terms (cusps) and each "n_k" equals N. So $\sigma N = \mu$. Moreover, there are no elliptic points of period $e > 1$ (points equivalent under $\mathrm{SL}_2(\mathbb{Z})$ to the elliptic points i and ρ for $\mathrm{SL}_2(\mathbb{Z})$ have trivial stabiliser in $\overline{\Gamma(N)}$). So our formulas for the genus (Theorem 2.29) and index (Corollary 2.28) along with Theorem 3.2 yield the following. For $N = 2$ and $k \geq 1$ even we have

$$\dim(M_k(\Gamma(2))) = (k-1)(-1) + \frac{6}{2} \cdot \frac{k}{2} = \frac{k}{2} + 1.$$

For $N > 2$ and $k \geq 2$ even we have that

$$\sigma = \frac{1}{2} N^2 \prod_{p|N} \left(1 - \frac{1}{p^2}\right)$$

and

$$g = 1 + \frac{N^2(N-6)}{24} \prod_{p|N} \left(1 - \frac{1}{p^2}\right)$$

so

$$\begin{aligned} \dim(M_k(\Gamma(N))) &= N^2 \prod_{p|N} \left(1 - \frac{1}{p}\right) \left(\frac{(k-1)(N-6)}{24} + \frac{k}{4}\right) \\ &= \left(\frac{(k-1)N + 6}{24}\right) N^2 \prod_{p|N} \left(1 - \frac{1}{p^2}\right). \end{aligned}$$

Spaces of odd dimension $k \neq 1$ may be treated in a similar (though more involved) manner (and are zero for k negative). These methods though fail in weight $k = 1$ and no general explicit formulas are known, even for such simple cases as $\dim(M_1(\Gamma_1(p)))$ for p prime.

4 Examples of modular forms

4.1 Eisenstein series in level 1

For $k \in \mathbb{N}$ define

$$G_k(z) := \sum_{c,d\in\mathbb{Z}}{}' \frac{1}{(cz+d)^k}$$

where the dash indicates that the sum omits the term $(c,d) = (0,0)$. (Note that as yet it is not clear when this sum converges, or even makes sense since we have not specified an order of summation.) For $\gamma = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ we have

$$G_k\left(\frac{Az+B}{Cz+D}\right) = \sum_{c,d\in\mathbb{Z}}{}' \frac{1}{\left(c\cdot\frac{Az+B}{Cz+D}+d\right)^k}.$$

$$= \sum_{c,d\in\mathbb{Z}}{}' \frac{(Cz+D)^k}{((cA+dC)z+(cB+dD))^k} = (Cz+D)^k \sum_{c',d'\in\mathbb{Z}}{}' \frac{1}{(c'z+d')^k}.$$

Here we have used that as (c,d) varies over $\mathbb{Z}^2\backslash(0,0)$ so does $(c',d') := (cA+dC, cB+dD)$. So $G_k(z)$ looks like a modular form of weight k for $\mathrm{SL}_2(\mathbb{Z})$.

To establish this rigorously we first prove that the sum converges absolute and uniformly on compact subsets of $\mathfrak{H}$. Precisely, one needs the absolute convergence since the order of summation is not specified and so in fact without this the definition does not make sense. Then, given the absolute convergence, the uniform convergence (which can then be proved with respect to any choice of ordering) implies the function is holomorphic on $\mathfrak{H}$, by for example [5, V1.2 Lemma 1].

Proposition 4.1. *The series*

$$\sum_{c,d\in\mathbb{Z}}{}' \frac{1}{|cz+d|^k},$$

converges uniformly on compact subsets of $\mathfrak{H}$ whenever $k > 2$.

Proof. Let z be fixed, with $\mathrm{Im}(z) > 0$. Then $\{cz+d : c,d \in \mathbb{Z}\}$ is the integer lattice generated by 1 and z. For $r \in \mathbb{Z}_{>0}$, let π_r be the parallelogram

$$\{\pm rz + d,\ dz \pm r : -r \le d \le r\}$$

in the lattice, see [2, Chapter III Figure 7]. We sum the series over each parallelogram separately. On π_r there are $8r$ vertices. Let δ be the minimum distance of π_1 to the origin; then $r\delta$ is the minimum distance of π_r to the origin, so that $|cz+d| \ge r\delta$ if $cz+d \in \pi_r$. Thus

$$\sum_{(c,d)\in\pi_r} \frac{1}{|cz+d|^k} \le \frac{8r}{(\delta r)^k} = 8\delta^{-k}\frac{1}{r^{k-1}}$$

and

$$\sum_{c,d\in\mathbb{Z}}{}' \frac{1}{|cz+d|^k} = \sum_{r=1}^{\infty} \sum_{(c,d)\in\pi_r} \frac{1}{|cz+d|^k} \leq 8\delta^{-k} \sum_{r=1}^{\infty} \frac{1}{r^{k-1}} < \infty$$

if $k > 2$. For uniformity of convergence over compact subsets in $\mathfrak{H}$, we note that our estimate depends only upon δ. By making δ smaller, if necessary, the estimate holds uniformly for all z in any compact subset of $\mathfrak{H}$. □

Corollary 4.2. *For $k > 2$ we have $G_k(z)$ is a holomorphic function on $\mathfrak{H}$.*

Note that for $k \geq 3$ odd the function $G_k(z)$ is identically zero, since by taking $\gamma = -I$ we see $G(z) = (-1)^k G(z)$. The calculation above shows it has the correct invariance properties, and thus to prove $G_k(z) \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ we need to examine how it behaves as $\mathrm{Im}(z) \to \infty$. (When $k = 1, 2$ we do not get a modular form.) Note that $G_k(z)\,(k > 2)$ is a holomorphic function invariant under $z \mapsto z + 1$, and so has a Fourier expansion. We shall compute this, and see then that $G_k(z)$ is holomorphic at the cusp ∞.

Define for $k > 2$

$$E_k(z) := \frac{1}{2} \sum_{\gcd(c,d)=1} \frac{1}{(cz+d)^k}.$$

Notice that

$$\sum_{(c,d)\in\mathbb{Z}^2}{}' \frac{1}{(cz+d)^k} = \sum_{n=1}^{\infty} \sum_{\substack{(c,d)\\ \gcd(c,d)=n}} \frac{1}{(cz+d)^k}$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^k} \sum_{\gcd(c,d)=1} \frac{1}{(cz+d)^k} = \zeta(k) \sum_{\gcd(c,d)=1} \frac{1}{(cz+d)^k}.$$

This calculation is valid by absolute convergence. Thus

$$G_k(z) = 2\zeta(k) E_k(z).$$

We compute the q-expansion of $G_k(z)$.

Lemma 4.3.

$$\sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^r} = \frac{(-2\pi i)^r}{(r-1)!} \sum_{\nu=1}^{\infty} \nu^{r-1} e^{2\pi i \nu z}.$$

if $r > 1$ and $\mathrm{Im}(z) > 0$.

Proof. We start with the "well-known" partial fraction decomposition of the cotangent function

$$\pi \cot \pi z = \sum_{n=-\infty}^{\infty} \frac{1}{(z+n)} \left(:= \lim_{m\to\infty} \sum_{n=-m}^{m} \frac{1}{(z+n)} \right). \tag{9}$$

(This can be obtained from the more intuitive identity

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

by taking the logarithmic derivative, i.e., forming $d(\log(f)) = f'/f$.) The series converges (absolutely) uniformly on compact subsets on $\mathfrak{H}$, and so we can differentiate term by term to obtain

$$\frac{d}{dz}(\pi \cot \pi z) = -\sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^2},$$

and similarly

$$\frac{d^{r-1}}{dz^{r-1}}(\pi \cot \pi z) = (-1)^{r-1}(r-1)! \sum_{n=-\infty}^{\infty} \frac{1}{(z+n)^r}.$$

On the other hand we have

$$\begin{aligned} \pi \cot \pi z &= -\pi i \tfrac{1+e^{2\pi i z}}{1-e^{2\pi i z}} \\ &= -\pi i (1+e^{2\pi i z}) \textstyle\sum_{\nu=0}^{\infty} e^{2\pi i \nu z} \text{ for } \mathrm{Im}(z) > 0 \\ &= -\pi i \left(1 + 2\textstyle\sum_{\nu=1}^{\infty} e^{2\pi i \nu z}\right). \end{aligned}$$

The series above converges (absolutely) uniformly on compact subsets of $\mathfrak{H}$ and hence can be differentiated term by term to get

$$\frac{d^{r-1}}{dz^{r-1}}(\pi \cot \pi z) = -(2\pi i)^r \sum_{\nu=0}^{\infty} \nu^{r-1} e^{2\pi i \nu z}.$$

The lemma now follows by comparing these two expansions. □

Theorem 4.4. *For $k > 2$ and even,*

$$G_k(z) = 2\zeta(k) + 2\frac{(-1)^{k/2}(2\pi)^k}{(k-1)!} \sum_{\lambda=1}^{\infty} \sigma_{k-1}(\lambda) e^{2\pi i \lambda z}$$

where $\zeta(z)$ is the Riemann zeta function, and

$$\sigma_{k-1}(\lambda) = \sum_{d|\lambda} d^{k-1}.$$

Proof.

$$G_k(z) = \sum_{c,d\in\mathbb{Z}}{}' \frac{1}{(cz+d)^k} = \sum_{\substack{d=-\infty \\ d\neq 0}}^{\infty} \frac{1}{d^k} + \sum_{c=1}^{\infty}\sum_{d=-\infty}^{\infty} \frac{1}{(cz+d)^k} + \sum_{c=-\infty}^{-1}\sum_{d=-\infty}^{\infty} \frac{1}{(cz+d)^k}.$$

In the last sum setting $c' = -c$ and $d' = -d$ we see it is the same as the second, since k is even. So

$$G_k(z) = 2\zeta(k) + 2\sum_{c=1}^{\infty}\sum_{d=-\infty}^{\infty} \frac{1}{(cz+d)^k}.$$

Since $c > 0$, $\mathrm{Im}(cz) > 0$, and so Lemma 4.3 tells us

$$G_k(z) = 2\zeta(k) + 2\frac{(-2\pi i)^k}{(k-1)!} \sum_{c=1}^{\infty}\sum_{\nu=1}^{\infty} \nu^{k-1} e^{2\pi i \nu c z}.$$

Collecting terms for which $\nu c = \lambda$ gives the theorem. □

For even integers $2\ell \geq 2$, define the **Bernoulli numbers** $b_{2\ell} \in \mathbb{Q}$ by

$$z \cot z = 1 - \sum_{\ell=1}^{\infty} b_{2\ell} \frac{2^{2\ell} z^{2\ell}}{(2\ell)!}. \tag{10}$$

For example.

$$b_2 = \frac{1}{6},\ b_4 = \frac{1}{30},\ b_6 = \frac{1}{42},\ b_8 = \frac{1}{30}, b_{10} = \frac{5}{66},\ \cdots .$$

Then for $k \geq 2$ even

$$\zeta(k) := \sum_{n=1}^{\infty} \frac{1}{n^k} = \frac{b_k 2^{k-1} \pi^k}{k!} > 0$$

and so in particular $b_k > 0$. (One can see this by comparing

$$z \cot z = 1 + 2 \sum_{n=1}^{\infty} \frac{z^2}{z^2 - n^2\pi^2} = 1 - 2 \sum_{n=1}^{\infty} \sum_{\ell=1}^{\infty} \frac{z^{2\ell}}{n^{2\ell} \pi^{2\ell}}$$

with (10). The first equality here is obtained from (9) by replacing z by z/π and multiplying through by z/π.)

By Theorem 4.4 we then have

$$E_k(z) = \frac{1}{2\zeta(k)} G_k(z) = 1 + \frac{(-1)^{k/2} 2k}{b_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) e^{2\pi i n z}$$

is a modular form of weight k for $\mathrm{SL}_2(\mathbb{Z})$ with rational coefficients which takes the value one at the cusp.

4.2 Ramanujan's $\Delta(q)$

We saw that the dimension of the space of modular forms of even weight $k > 0$ for $\mathrm{SL}_2(\mathbb{Z})$ was $\left[\frac{k}{12}\right]$ for $k \equiv 2 \bmod 12$, and $\left[\frac{k}{12}\right] + 1$ for $k \not\equiv 2 \bmod 12$. Thus the dimensions for $k = 2, 4, 6, 8, 10, 12$ are $0, 1, 1, 1, 1, 2$. The Eisenstein series account for everything in weight < 12, but in dimension 12 we get our first cusp form, which normalised to have leading term $q := e^{2\pi i z}$ must be

$$\Delta(z) := \frac{E_4(z)^3 - E_6^2(z)}{1728} = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 + \cdots \neq 0.$$

We write $\Delta(q) = \sum_{n=1}^{\infty} \tau(n) q^n$. There is a surprising product expansion

$$\Delta(z) = q \prod_{n=1}^{\infty} (1 - q^n)^{24}.$$

To prove this one must verify that $F(-1/z) = z^{12} F(z)$, where $F(z)$ is the RHS in the above. This is a subtle exercise in conditional convergence (see for example [3, Chapter III.2] or [5, Page 95-96]).

On the exercise sheets we shall give explicit generators for the *ring* of modular forms for $\mathrm{SL}_2(\mathbb{Z})$, and bases for each space $M_k(\mathrm{SL}_2(\mathbb{Z}))$.

4.3 Some arithmetic applications

The non-existence of cusp forms in dimensions $4, 6, 8, 10$ and 14 leads to some curious identities involving the divisor function $\sigma_k(n) := \sum_{d|n} d^k$. For example, $(E_4(z))^2 = E_8(z)$ and so comparing coefficients

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_3(n-m)$$

and $E_4(z)E_6(z) = E_{10}(z)$ and so

$$11\sigma_9(n) = 21\sigma_5(n) - 10\sigma_3(n) + 5040 \sum_{m=1}^{n-1} \sigma_3(m)\sigma_5(n-m).$$

Also, the equalities $E_6(z)E_8(z) = E_4(z)E_{10}(z) = E_{14}(z)$ give similar identities.

4.4 The Eisenstein subspace in level 1

Assume $k \geq 4$ is even and let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ with $f(\infty) = a_0(f)$. Then we may write

$$f = (f - a_0(f) \cdot E_k(z)) + a_0(f) \cdot E_k(z).$$

Note that $f - a_0(f) \cdot E_k(z)$ takes the value $f(\infty) - a_0(f) \cdot 1 = 0$ at the cusp ∞, and hence lies in $S_k(\mathrm{SL}_2(\mathbb{Z}))$. So we have shown that

$$\langle E_k(z) \rangle + S_k(\mathrm{SL}_2(\mathbb{Z})) = M_k(\mathrm{SL}_2(\mathbb{Z}))$$

and this sum is obviously direct, since a non-zero multiple of $E_k(z)$ cannot vanish at ∞. We call $E_k(\mathrm{SL}_2(\mathbb{Z})) := \langle E_k(z) \rangle$ the *Eisenstein subspace*, and thus

$$M_k(\mathrm{SL}_2(\mathbb{Z})) = E_k(\mathrm{SL}_2(\mathbb{Z})) \oplus S_k(\mathrm{SL}_2(\mathbb{Z})).$$

5 The Petersson inner product

Let f and g be two modular forms of weight $k > 0$ for a subgroup Γ of finite index in $\mathrm{SL}_2(\mathbb{Z})$. Notice from the proof of Lemma 2.5 that $\mathrm{GL}_2^+(\mathbb{R})$, the group of real matrices with positive determinant, acts on $\mathfrak{H}$ via LFTs.

Lemma 5.1. *The function $f(z)\overline{g(z)}y^k$ and differential $dxdy/y^2$ are invariant under the actions of Γ and $\mathrm{GL}_2^+(\mathbb{R})$, respectively.*

Proof. We have

$$\begin{aligned} f(\gamma z) &= (cz+d)^k f(z) \text{ (modularity of } f) \\ \overline{g(\gamma z)} &= \overline{(cz+d)}^k \overline{g(z)} \\ \mathrm{Im}(\gamma z) &= \det(\gamma) \tfrac{\mathrm{Im}(z)}{|cz+d|^2}. \end{aligned}$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma, \Gamma$ and $\mathrm{GL}_2^+(\mathbb{R})$, respectively.

The first result now follows by multiplying the first two equations by the kth power of the third. Next, observe we can write

$$dxdy = \frac{i}{2} dz d\bar{z}$$

and

$$\gamma(dz) := d(\gamma z) = \det(\gamma) \frac{dz}{(cz+d)^2}.$$

So

$$\gamma(dxdy) = \frac{i}{2} d(\gamma z) d(\gamma \bar{z}) = \frac{i}{2} \det(\gamma)^2 |cz+d|^{-4} dz d\bar{z} = \det(\gamma)^2 |cz+d|^{-4} dxdy$$

for $\gamma \in \mathrm{GL}_2^+(\mathbb{R})$. Hence

$$\gamma(\mathrm{Im}(z)^{-2} dxdy) = \mathrm{Im}(\gamma z)^{-2} \det(\gamma)^2 |cz+d|^{-4} dxdy = \mathrm{Im}(z)^{-2} dxdy$$

which completes the proof. □

With D the usual fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ we have

$$\int_D \frac{dxdy}{y^2} < \int_{-1/2}^{1/2} \int_{\sqrt{3}/2}^{\infty} y^{-2} dy dx = \frac{2}{\sqrt{3}}$$

and hence if $\phi : \mathfrak{H} \to \mathbb{C}$ is bounded and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ then

$$\int_D \phi(\gamma(z)) \frac{dxdy}{y^2}$$

is finite (since $\phi(\gamma(z))$ is bounded on D). Now suppose at least one of f or g is a cusp form (or even just that that fg vanishes at all the cusps). We shall assume our fundamental domain D_Γ for Γ is a finite disjoint union $\cup \gamma_i(D)$ over some $\gamma_i \in \mathrm{SL}_2(\mathbb{Z})$. We first show that the integral

$$\int_{D_\Gamma} f(z)\overline{g(z)} y^{k-2} dxdy = \sum_i \int_D f(\gamma_i z)\overline{g(\gamma_i z)} \mathrm{Im}(\gamma_i z)^k \gamma_i \left(\frac{dxdy}{y^2} \right)$$

$$= \sum_i \int_D f(\gamma_i z)\overline{g(\gamma_i z)} \mathrm{Im}(\gamma_i z)^k \frac{dxdy}{y^2} \text{ (by the invariance of } dxdy/y^2)$$

is finite. It is enough to show that the integrand

$$f(\gamma_i z)\overline{g(\gamma_i z)} \mathrm{Im}(\gamma_i z)^k = f(\gamma_i z)\overline{g(\gamma_i z)} \mathrm{Im}(z)^k |cz+d|^{-2k} \text{ (proof of Lemma 5.1)}$$

is bounded on D.

Equivalently we need to show that

$$f|_k \gamma_i \cdot \overline{g|_k \gamma_i} \cdot \mathrm{Im}(z)^k$$

is bounded on D. Note that $f|_k \gamma_i$ and $g|_k \gamma_i$ are modular forms for $\gamma_i^{-1} \Gamma \gamma_i$.

Since this integrand is continuous it is bounded on compact subsets of D. So we need only examine the neighbourhoods $S_\varepsilon = \{\mathrm{Im}(z) > 1/\varepsilon\} \cup \{\infty\}$ of the cusp ∞. We have Fourier expansions

$$f|_k\gamma_i = \sum_{n=0}^{\infty} a_n(f|_k\gamma_i)q_h^n,\ g|_k\gamma_i = \sum_{n=0}^{\infty} a_n(g|_k\gamma_i)q_h^n$$

where $a_0(f|_k\gamma_i)a_0(g|_k\gamma_i) = 0$ and $q_h = e^{2\pi iz/h}$ with h the width of the cusp at infinity for $\gamma_i^{-1}\Gamma\gamma_i$. The product $f|_k\gamma_i \cdot \overline{g|_k\gamma_i}$ is $\mathcal{O}(|q_h|)$ as $\mathrm{Im}(z) \to \infty$ (since the constant term vanishes) and so

$$f|_k\gamma_i \cdot \overline{g|_k\gamma_i} \cdot \mathrm{Im}(z)^k = \mathcal{O}(|q_h|)\mathrm{Im}(z)^k.$$

Since $|q_h| = e^{-2\pi\mathrm{Im}(z)/h}$ and exponential decay dominates polynomial growth, the integrand tends to zero as $\mathrm{Im}(z) \to \infty$, and so is indeed bounded on D.

Definition 5.2. *(Petersson inner product) For $f, g \in M_k(\Gamma)$ with at least one a cusp form, define*

$$\langle f, g\rangle_\Gamma := \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]} \int_{D_\Gamma} f(z)\overline{g(z)}y^{k-2}dxdy.$$

Exercise 5.3 Show that

1. the integral is independent of the choice of fundamental domain. (You may assume here for simplicity that we only consider fundamental domains which are unions of translates of D by coset representatives.)
2. the scaling factor ensures that for $\Gamma' \leq \Gamma$ we have $\langle f, g\rangle_{\Gamma'} = \langle f, g\rangle_\Gamma$ (so we may omit the subscript).
3. $\langle \cdot, \cdot \rangle$ is a (Hermitian) inner product on the complex vector space $S_k(\Gamma)$.

6 Modular forms as functions on lattices

6.1 Functions on lattices

A **lattice** L in $\mathbb{C}$ is a subgroup such that there exists an $\mathbb{R}$-basis w_1, w_2 for $\mathbb{C}$ which is a $\mathbb{Z}$-basis for L. Let $\mathcal{L}$ denote the set of lattices in $\mathbb{C}$.

Theorem 6.1. *Let $k \in \mathbb{Z}$. There is a bijection between functions $F : \mathcal{L} \to \mathbb{C}$ on lattices satisfying*

$$F(\lambda L) = \lambda^{-k}F(L)$$

for all $\lambda \in \mathbb{C}^*$ and functions $f : \mathfrak{H} \to \mathbb{C}$ satisfying

$$f(\gamma z) = (cz + d)^k f(z)$$

for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

Here and below as usual $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Proof. Given such a function F on lattices, we define $f : \mathfrak{H} \to \mathbb{C}$ by

$$f(z) := F(\langle z, 1\rangle_{\mathbb{Z}})$$

where $\langle z, 1\rangle_{\mathbb{Z}}$ denotes the lattice spanned by 1 and z. Now

$$f(\gamma z) = F\left(\left\langle \frac{az+b}{cz+d}, 1 \right\rangle\right) = (cz+d)^k F(\langle az+b, cz+d\rangle)$$

by the homogenity property of F. But since $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ we have that $\langle az+b, cz+d\rangle = \langle z, 1\rangle$, and so $f(\gamma z) = (cz+d)^k f(z)$.

Conversely, given such an $f : \mathfrak{H} \to \mathbb{C}$ we define $F : \mathcal{L} \to \mathbb{C}$ as follows. Let $L = \langle w_1, w_2\rangle$ be a lattice. We may assume that $\mathrm{Im}(w_1/w_2) > 0$ (by switching the sign of one generator if necessary). Now define

$$F(\langle w_1, w_2\rangle) = w_2^{-k} f(w_1/w_2).$$

Then certainly $F(\lambda L) = F(\langle \lambda w_1, \lambda w_2\rangle) = \lambda^{-k} F(L)$, but the key point is really to check that the function F is well-defined. Suppose then that $L = \langle w_1, w_2\rangle = \langle w_1', w_2'\rangle$ with also $\mathrm{Im}(w_1'/w_2') > 0$. Then the first basis can be transformed into the second by an invertible integer matrix, and this matrix must have determinant one by the assumption $w_1/w_2, w_1'/w_2' \in \mathfrak{H}$. That is, there exists $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ such that $w_1' = aw_1 + bw_2$ and $w_2' = cw_1 + dw_2$. Then we have

$$F(\langle w_1', w_2'\rangle) = (w_2')^{-k} f(w_1'/w_2') = (cw_1 + dw_2)^{-k} f\left(\frac{aw_1 + bw_2}{cw_1 + dw_2}\right)$$

$$= (cw_1 + dw_2)^{-k} f\left(\frac{a(w_1/w_2) + b}{c(w_1/w_2) + d}\right) = (cw_1 + dw_2)^{-k} (c(w_1/w_2) + d)^k f(w_1/w_2)$$

$$= w_2^{-k} f(w_1/w_2) = F(\langle w_1, w_2\rangle).$$

One see that these two maps ("$f \mapsto F$" and "$F \mapsto f$") are inverses of one another. □

6.2 Eisenstein series for $\mathrm{SL}_2(\mathbb{Z})$, revisited

For $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ there was one Eisenstein series

$$E_k(z) := \frac{1}{2} \sum_{\gcd(c,d)=1} \frac{1}{(cz+d)^k}$$

of weight k for each even $k \geq 4$. We also considered the multiple

$$G_k(z) = \sum_{c,d \in \mathbb{Z}}{}' \frac{1}{(cz+d)^k} = 2\zeta(k) E_k(z).$$

Now define the following function $\mathcal{G}_k$ *on lattices*. For $L \subset \mathbb{C}$ a lattice

$$\mathcal{G}_k(L) := \sum_{w \in L}{}' w^{-k},$$

and so $\mathcal{G}_k(\lambda L) = \lambda^{-k}\mathcal{G}_k(L)$ for all $\lambda \in \mathbb{C}^*$. Notice then

$$\mathcal{G}_k(L_z) = G_k(z)$$

where we write L_z for the lattice spanned by 1 and z. Thus the Eisenstein series $G_k(z)$ is "visibly" a function on lattices of weight k.

7 Hecke operators in level 1

We have seen that modular forms for $\mathrm{SL}_2(\mathbb{Z})$ can be viewed as functions on lattices. To define operators on them it suffices to define operators on lattices, which we do now.

7.1 Hecke operators on lattices

Let $\mathcal{D}$ be the free abelian group generated by the set $\mathcal{L}$ of lattices in $\mathbb{C}$. Thus an element in $\mathcal{D}$ is a finite sum $\sum n_i[L_i]$ over lattices L_i in $\mathbb{C}$ with coefficients $n_i \in \mathbb{Z}$. For each $n \in \mathbb{N}$ we define the Hecke operator $T(n) : \mathcal{D} \to \mathcal{D}$ by

$$T(n) : [L] \mapsto \sum_{(L:L')=n} [L']$$

for L a lattice in $\mathbb{C}$, and then extend to all of $\mathcal{D}$ by linearity. Thus $T(n)$ associates with L the sublattices L' of index n in L, each with multiplicity 1. We define another operator $R(n) : \mathcal{D} \to \mathcal{D}$ by

$$R(n) : [L] \mapsto [nL].$$

So $R(n)$ associates to L the lattice consisting of all n-multiples of elements in L. One can easily check the operators $R(n)$ and $T(m)$ commute.

Theorem 7.1. *1.* $T(m)T(n) = T(mn) = T(n)T(m)$ *provided* $\gcd(m, n) = 1$

2. For a prime p *and integer* $r \geq 1$,

$$T(p^r)T(p) = T(p^{r+1}) + pR(p)T(p^{r-1}).$$

3. The algebra generated by the $T(n)$ *and* $R(n)$ *(all* n*) is generated by the operators* $R(p)$ *and* $T(p)$ *(all primes* p*), and is thus commutative.*

Proof. For Part (1) it is enough to show that $T(m)T(n)([L]) = T(mn)([L])$ for any lattice L when $\gcd(m, n) = 1$. Any finite abelian group of order mn with $\gcd(m, n) = 1$ contains a unique

subgroup of index n. It follows that each sublattice L'' of index mn in the lattice L lies in a unique sublattice $L' \subset L$ of index n, which gives the result.

We now prove part (2).

By linearity, it is enough to consider the result of applying each side of the equation to $[L]$ for a lattice L. Both sides of the equation associate to L sublattices of index p^{r+1}, and we need to show the multiplicities are the same. Let L' be such a sublattice.

If $L' \subset pL$ then the RHS gives L' multiplicity $1+p$ (note $R(p)T(p^{r-1}) = T(p^{r-1})R(p)$). Moreover, L' is contained in all $p+1$ sublattices of L of index p. The LHS associates all sublattices of index p^r to the lattices of index p in L. Thus L' appears once for each of the $(p+1)$ sublattices of index p, and so also has multiplicity $p+1$ on the LHS. (See Example 7.4 for an explicit description of these sublattices.)

Suppose now L' is not contained in pL. Then it has multiplicity 1 on the RHS. If it had multiplicity > 1 on the LHS then it would be contained in at least two sublattices of index p, and therefore their intersection, which is precisely pL. So L' has multiplicity 1 on the LHS, as required.

Note that the formula in (2) shows (by induction) that $T(p^{r+1})$ is a polynomial in $R(p)$ and $T(p)$, and hence by (1) each $T(n)$ is a polynomial in $R(p)$ and $T(p)$ (for primes p dividing n). This proves part (3). □

Let $F : \mathcal{L} \to \mathbb{C}$ be a function on lattices. We can extend F by linearity to a function on $\mathcal{D}$. For any operator T on $\mathcal{D}$ we define $T \cdot F : \mathcal{D} \to \mathbb{C}$ to be the function $F \circ T$. In particular for $L \in \mathcal{L}$ we have $(T \cdot F)([L]) = F(T([L]))$.

For example,

$$(T(n) \cdot F)([L]) = \sum F([L'])$$

where the sum is over sublattices of index n in L. Moreover, if F is of weight k, i.e. $F(\lambda L) = \lambda^{-k} F(L)$ for all lattices L, then

$$R(n) \cdot F = n^{-k} \cdot F \tag{11}$$

since for $L \in \mathcal{L}$ we have $R(n) \cdot F([L]) := F(R(n)([L])) = F([nL]) = n^{-k} F([L])$, and also we have that $T(n)(F)$ is of weight k.

7.1.1 Hecke operators on modular forms

Recall that we have a one-to-one correspondence between functions F on $\mathcal{L}$ of weight k and functions $f : \mathfrak{H} \to \mathbb{C}$ such that $f(\gamma z) = (cz + d)^k f(z)$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, under which

$$\begin{aligned} F(\langle w_1, w_2 \rangle_{\mathbb{Z}}) &= w_2^{-k} f(w_1/w_2) \\ f(z) &= F(\langle z, 1 \rangle_{\mathbb{Z}}). \end{aligned}$$

Let $f : \mathfrak{H} \to \mathbb{C}$ transform under $\mathrm{SL}_2(\mathbb{Z})$ like a modular form of weight k and F be the associated function of weight k on $\mathcal{L}$. We define $T(n)(f(z))$ to be the function on $\mathfrak{H}$ associated to $n^{k-1} T(n) \cdot F$. Thus

$$T(n) : f(z) \mapsto n^{k-1} (T(n) \cdot F)(\langle z, 1 \rangle_{\mathbb{Z}}). \tag{12}$$

Note 7.2 There is an alternative definition of the Hecke operator $T(n)$ which is perhaps more natural and accounts for the factor n^{k-1}. Instead of summing over lattices L' of index n in L, one can "average" over lattices L' containing L with index n. That is, define a new

$$\tilde{T}(n) : [L] \mapsto \frac{1}{n} \sum [L']$$

where now $[L' : L] = n$. If one scales such an L' by a factor n one gets a lattice contained in L of index n, and vice-versa. Our modular forms F on lattices have a weight k homogeneity property, and this accounts for a factor n^k difference between the two approaches; that we in addition average in the second accounts for a further $1/n$ factor difference. This total factor n^{k-1} in our approach is then put in when we define the Hecke operators on modular forms, but not in the other, and the two definitions *on modular forms* amount to the same thing[3].

The following matrix lemma allows us to give a much more explicit description of the Hecke operator $T(n)$ on modular forms (or lattices).

Lemma 7.3. *Let $A \in M_2(\mathbb{Z})$ with $\det(A) = n$. Then there exists $U \in \mathrm{SL}_2(\mathbb{Z})$ such that $UA = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with $ad = n$, $a \geq 1$ and $0 \leq b < d$. Moreover, the integers a, b, d are uniquely determined.*

Proof. First, one clears the bottom left-hand entry. That is, writing $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, solve $xa + yc = 0$ with x and y coprime, then construct a matrix in $\mathrm{SL}_2(\mathbb{Z})$ with x and y as the bottom row (using the extended Euclidean algorithm). Multiplying by $-I$ if necessary we can assume $a > 0$.

Next, by multiplying on the left by matrices $\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$ we can add the bottom row of our matrix to the top, and thus reduce b modulo d. For uniqueness, note that a is the gcd of the elements in the first column of A, d is the unique positive integer such that $ad = n$, and b is unique modulo d. □

[3] I find the proofs with our approach easier to follow though.

Let $M(n)$ be the set of 2×2 integer matrices with determinant n. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on $M(n)$ by left multiplication, and Lemma 7.3 provides us with a canonical set of representatives for the orbits.

Now let L be a lattice in $\mathbb{C}$. Choose a basis w_1, w_2 for L, so $L = \langle w_1, w_2 \rangle_{\mathbb{Z}}$. For any $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(n)$[4] define αL to be the lattice spanned by $aw_1 + bw_2, cw_1 + dw_2$. Then αL is a sublattice of L of index n, and every such sublattice is of this form for some $\alpha \in M(n)$. Since

$$\alpha L = \beta L \Leftrightarrow \mathbb{Z}\text{-span of rows of } \alpha \text{ and } \beta \text{ the same} \Leftrightarrow \beta = u\alpha \text{ for some } u \in \mathrm{SL}_2(\mathbb{Z})$$

we see that the sublattices of L of index n are precisely the αL where $\alpha = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with

$$ad = n,\ a \geq 1 \text{ and } 0 \leq b < d. \tag{13}$$

Using this explicit description of the lattices of index n in $\langle z, 1 \rangle_{\mathbb{Z}}$, we can rewrite (12) as

$$T(n) : f(z) \mapsto n^{k-1} \sum d^{-k} f\left(\frac{az+b}{d}\right) \tag{14}$$

where the sum is over the triples a, b, d satisfying (13).

Example 7.4 For p prime we get matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & p \end{pmatrix}, \cdots, \begin{pmatrix} 1 & p-1 \\ 0 & p \end{pmatrix}, \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Given a lattice L and fixed choice of basis w_1, w_2, these $p+1$ matrices describe explicitly all $p+1$ sublattices of index p in L in terms of this choice of basis, e.g. the second matrix describes the lattice $\langle w_1 + w_2, pw_2 \rangle_{\mathbb{Z}}$. Put another way, given any sublattice L' with $|L/L'| = p$, if we start with any choice of basis for L' described by a matrix α with $\det(\alpha) = p$, then multiplying by a suitable $u \in \mathrm{SL}_2(\mathbb{Z})$ on the left will give a new matrix $u\alpha$ of the form above. This is called the *Hermite normal form* of α.

One can also change the basis for the *original* lattice L, which corresponds to multiplying on the *right* by an invertible matrix v over $\mathbb{Z}$. For example, if we take the new basis $w_1' := w_1 + w_2$, $w_2' = w_2$ then our lattice $\langle w_1 + w_2, pw_2 \rangle_{\mathbb{Z}}$ is just $\langle w_1', pw_2' \rangle_{\mathbb{Z}}$ and is now described by the matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & p \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}.$$

From this description of the lattice it becomes obvious that indeed the index is p. More generally, given any sublattice L' with $|L/L'| = p$, after a suitable choice of basis for L itself and then one for L' in terms of this basis, once ends up with the matrix $\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$. In matrix language, given α of

[4]We from now on use the Greek alphabet for linear fractional transformations (as matrices) rather than T, to avoid confusion with Hecke operators.

determinant p we can find $u, v \in \mathrm{SL}_2(\mathbb{Z})$ so that $u\alpha v = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$. This is called the *Smith normal form* of α.[5] Note that the Smith normal form only describes the structure of the quotient L/L' group (cyclic of order p here), whereas the Hermite normal form distinguishing between different sublattices L' and L'' for which $L/L' \cong L/L''$.

Let us call a meromorphic function $f : \mathfrak{H} \to \mathbb{C}$ **weakly modular** of weight k for Γ if it satisfies the invariance property (1) in Definition 1.6.

Theorem 7.5. *1. If f is a weakly modular form of weight k for $\mathrm{SL}_2(\mathbb{Z})$ then $T(n)(f)$ is also weakly modular, and*

(a) $T(m)T(n)(f) = T(mn)(f)$ if m and n are coprime.

(b) $T(p^r)T(p)(f) = T(p^{r+1})(f) + p^{k-1}T(p^{r-1})(f)$ if p is prime and $r \geq 1$

2. Let f be a modular form of weight k for $\mathrm{SL}_2(\mathbb{Z})$, with Fourier expansion $f = \sum_{m\geq 0} c(m)q^m$. Then $T(n)(f)$ is also a modular form, and

$$T(n)(f) = \sum_{m\geq 0} \gamma(m)q^m$$

where

$$\gamma(m) = \sum_{a \mid \gcd(m,n), a\geq 1} a^{k-1} c\left(\frac{mn}{a^2}\right)$$

Proof. For Part (1), we see that $T(n)f(z)$ is meromorphic on $\mathfrak{H}$ given that $f(z)$ is from (14), and it has the correct invariance properties because $T(n) \cdot F$ is a weight k function on lattice (symbols). The equations in part (1) follow from Theorem 7.1, (11) and (12) (check this).

For part (2), using (14) we note that $T(n)(f)$ is holomorphic on $\mathfrak{H}$ because f is. We have

$$T(n)(f(z)) = n^{k-1} \sum_{a,b,d} d^{-k} \sum_{m\geq 0} c(m) e^{2\pi i \frac{az+b}{d} m}.$$

But

$$\sum_{0\leq b<d} e^{2\pi i \frac{bm}{d}} = \begin{cases} d & \text{if } d|m \\ 0 & \text{otherwise.} \end{cases}$$

Setting $m/d = m'$, then

$$T(n)(f(z)) = n^{k-1} \sum_{a,d,m'} d^{-k+1} c(m'd) q^{am'}$$

where the sum is over the integers a, d, m' such that $ad = n$ and $a \geq 1$. The coefficient of q^m in this is

$$\sum_{a \mid \gcd(n,m), a\geq 1} a^{k-1} c\left(\frac{m}{a}\frac{n}{a}\right)$$

as required. Because $\gamma(m) = 0$ for $m < 0$, $T(n)(f(z))$ is holomorphic at ∞. □

[5] To get from $\begin{pmatrix} p & 0 \\ \textit{uniquely} 0 & 1 \end{pmatrix}$ to $\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$ using matrices in $\mathrm{SL}_2(\mathbb{Z})$ rather than $\mathrm{GL}_2(\mathbb{Z})$ first switch rows and columns, and then pre- and post-multiply by $\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$.

The most important case is $T(p)$ for p prime, and we have for $f = \sum a_n q^n$ that

$$T(p)(f) = \sum_{n \geq 0} b_n q^n$$

where

$$b_n = \begin{cases} a_{pn} & \text{if } p \not| n \\ a_{pn} + p^{k-1} a_{n/p} & \text{if } p | n. \end{cases}$$

Note that if f is a cusp form ($a_0 = 0$) then so is $T(p)(f)$.

7.1.2 Hecke operators are Hermitian (self-adjoint)

Let

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})^+$$

and let $f : \mathfrak{H} \to \mathbb{C}$. Recall

$$f|_k\alpha := (\det \alpha)^{k/2}(cz+d)^{-k} f\left(\frac{az+b}{cz+d}\right).$$

Note that scalar matrices act trivially, and if f is weakly modular of weight k for $\Gamma \leq \Gamma(1)$ then $f|_k\alpha = f$ for all $\alpha \in \Gamma$.

We can rewrite (14) as[6]

$$T(n) : f \mapsto \sum n^{(k/2)-1} f|_k\alpha$$

where the α's run over a particular set of representative for the orbits of $\Gamma(1)$ acting on $M(n)$. The righthand side is independent on the choice of set of representatives (since $f|_k u = f$ for $u \in \Gamma(1)$).

Recall that the Petersson inner product of two cusp forms f and g for $\Gamma(1)$ is

$$\langle f, g \rangle := \int_D f(z)\overline{g(z)} y^{k-2} dx dy$$

where $z = x + iy$ and D is any fundamental domain for $\Gamma(1)$.

Proposition 7.6. *For every* $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(n)$

$$\langle f|_k\alpha, g \rangle = \langle f, g|_k\alpha^{-1} \rangle.$$

The key to proving this proposition is in fact making sense of the Petersson inner products on both sides. Now $f|_k\alpha$ is a modular form "for $\alpha^{-1}\mathrm{SL}_2(\mathbb{Z})\alpha$", but this is not (necessarily) a subgroup of $\mathrm{SL}_2(\mathbb{Z})$. However, from Lemma 7.7 below we have $\Gamma(n) \leq \alpha^{-1}\mathrm{SL}_2(\mathbb{Z})\alpha$. Thus we may think of both $f|_k\alpha$ and g on the LHS as modular forms for $\Gamma(n)$, and compute the Petersson inner product

[6]This is not a typo: it is really $n^{(k/2)-1}$ rather than n^{k-1}.

with respect to this group. On the RHS we have $g|_k\alpha^{-1}$ is a modular form "for $\alpha\mathrm{SL}_2(\mathbb{Z})\alpha^{-1}$", and this group contains $\alpha\Gamma(n)\alpha^{-1}$ which is a subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Thus we may think of f and $g|_k\alpha^{-1}$ on the RHS as modular forms for $\alpha\Gamma(n)\alpha^{-1}$ and compute the inner product with respect to this group.

Lemma 7.7. *Let $\alpha \in M(n)$. Then $\alpha\Gamma(n)\alpha^{-1} \leq \mathrm{SL}_2(\mathbb{Z})$.*

Proof. Using elementary row and column operators, we can write

$$\alpha = \delta_1 \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \delta_2$$

where $\delta_1, \delta_2 \in \mathrm{SL}_2(\mathbb{Z})$ and $A, B \in \mathbb{Z}$ with $AB = n$. Then

$$\begin{aligned} &\alpha\Gamma(n)\alpha^{-1} \subseteq \mathrm{SL}_2(\mathbb{Z}) \\ \Leftrightarrow\ &\delta_1 \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \delta_2\Gamma(n)\delta_2^{-1} \begin{pmatrix} 1/A & 0 \\ 0 & 1/B \end{pmatrix} \delta_1^{-1} \subseteq \mathrm{SL}_2(\mathbb{Z}) \\ \Leftrightarrow\ &\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \Gamma(n) \begin{pmatrix} 1/A & 0 \\ 0 & 1/B \end{pmatrix} \subseteq \mathrm{SL}_2(\mathbb{Z}) \end{aligned}$$

since $\delta_2\Gamma(n)\delta_2^{-1} = \Gamma(n)$ (by normality) and $\delta_1 \in \mathrm{SL}_2(\mathbb{Z})$. Matrices on the set on the LHS are certainly invertible over $\mathbb{Q}$ with determinant 1, so we need only check they have integer coefficients. This is an easy explicit calculation, using the fact $A, B|n$ and for matrices $C \in \Gamma(n)$, "$C \equiv I \bmod n$". □

We shall also need the following result.

Lemma 7.8. *For $\alpha \in M(n)$ we have*

$$[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(n)}] = [\mathrm{PSL}_2(\mathbb{Z}) : \alpha\overline{\Gamma(n)}\alpha^{-1}].$$

Proof. Defining

$$\mathrm{Vol}(D_{\Gamma(n)}) := \int_{D_{\Gamma(n)}} \frac{dxdy}{y^2}$$

one sees that

$$[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(n)}] = \frac{\mathrm{Vol}(D_{\Gamma(n)})}{\mathrm{Vol}(D_{\mathrm{SL}_2(\mathbb{Z})})}$$

since a fundamental domain for $\Gamma(n)$ is just $\mu = [\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(n)}]$ transforms of one for $\mathrm{SL}_2(\mathbb{Z})$, and $dxdy/y^2$ is $\mathrm{SL}_2(\mathbb{Z})$-invariant (so the volume of each transform is the same). Thus we must show $\mathrm{Vol}(D_{\Gamma(n)}) = \mathrm{Vol}(D_{\alpha\Gamma(n)\alpha^{-1}})$.

We may take $D_{\alpha\Gamma(n)\alpha^{-1}} = \alpha(D_{\Gamma(n)})$. (For given $z \in \mathfrak{H}$ since also $\alpha^{-1}(z) \in \mathfrak{H}$ there exists $\gamma \in \Gamma(n)$ and $w \in \overline{D}_{\Gamma(n)}$ such that $\alpha^{-1}(z) = \gamma w$, and so $z = \alpha\gamma\alpha^{-1}(\alpha w)$, and so on.) So $\mathrm{Vol}(D_{\alpha\Gamma(n)\alpha^{-1}}) = \mathrm{Vol}(\alpha(D_{\Gamma(n)}))$. Now the $\mathrm{GL}_2(\mathbb{Q})^+$-invariance of $dxdy/y^2$ (Lemma 5.1) shows us the volume $\mathrm{Vol}(\alpha(D_{\Gamma(n)}))$ equals $\mathrm{Vol}(D_{\Gamma(n)})$. □

We now prove Proposition 7.6

Proof. Write "$\omega(F,G) = F(z)\overline{G(z)}y^{k-2}dxdy$". One first checks by an explicit computation that

$$\omega(f|_k\alpha, g) = \alpha(\omega(f, g|_k\alpha^{-1}))$$

where α acts on $\omega(f, g|_k\alpha^{-1})$ just by substitution $z \mapsto \alpha z$. The LHS is

$$(\det\alpha)^{k/2}(cz+d)^{-k}f(\alpha z)\cdot\overline{g(z)}\cdot\mathrm{Im}(z)^k\frac{dxdy}{y^2}.$$

Since

$$\alpha^{-1} = \frac{1}{\det(\alpha)}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

the RHS is

$$\alpha\left(f(z)\cdot\overline{\det(\alpha^{-1})^{k/2}(\det(\alpha)^k(-cz+a)^{-k})g(\alpha^{-1}z)}\cdot\mathrm{Im}(z)^k\frac{dxdy}{y^2}\right).$$

Now

$$\alpha\left(\frac{dxdy}{y^2}\right) = \frac{dxdy}{y^2},\ \mathrm{Im}(\alpha z)^k = \det(\alpha)^k\frac{\mathrm{Im}(z)^k}{|cz+d|^k}.$$

Also one computes

$$(-c\cdot\alpha z + a) = \frac{\det(\alpha)}{(cz+d)}.$$

Thus the RHS is

$$f(\alpha z)\cdot\det(\alpha)^{-k/2}\det(\alpha)^k\det(\alpha)^{-k}(c\overline{z}+d)^k\overline{g(\alpha\alpha^{-1}z)}\cdot\det(\alpha)^k\frac{\mathrm{Im}(z)^k}{|cz+d|^k}\frac{dxdy}{y^2}$$

$$= \det(\alpha)^{-k/2+k-k+k}f(\alpha z)(c\overline{z}+d)^k\frac{1}{|cz+d|^k}\overline{g(\alpha\alpha^{-1}z)}\mathrm{Im}(z)^k\frac{dxdy}{y^2} = \mathrm{LHS}$$

as required.

Let $D_{\Gamma(n)}$ denote a fundamental domain for $\Gamma(n)$. Then $\alpha(D_{\Gamma(n)})$ is a fundamental domain for $\alpha\Gamma(n)\alpha^{-1}$. Also

$$\int_{D_{\Gamma(n)}}\omega(f|_k\alpha, g) = \int_{D_{\Gamma(n)}}\alpha(\omega(f, g|_k\alpha^{-1})) = \int_{\alpha(D_{\Gamma(n)})}\omega(f, g|_k\alpha^{-1}).$$

By Lemma 7.8 we have $[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(n)}] = [\mathrm{PSL}_2(\mathbb{Z}) : \alpha\overline{\Gamma(n)}\alpha^{-1}]$ and so

$$\langle f|_k\alpha, g\rangle_{\Gamma(n)} := \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma(n)}]}\int_{D_{\Gamma(n)}}\omega(f|_k\alpha, g)$$

$$= \frac{1}{[\mathrm{PSL}_2(\mathbb{Z}) : \alpha\overline{\Gamma(n)}\alpha^{-1}]}\int_{\alpha(D_{\Gamma(n)})}\omega(f, g|_k\alpha^{-1}) = \langle f, g|_k\alpha^{-1}\rangle_{\alpha\Gamma(n)\alpha^{-1}}$$

where we use the same notation for $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ and its image in $\mathrm{PSL}_2(\mathbb{Z})$. □

We now state our main theorem.

Theorem 7.9. *For cusp forms $f, g \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ we have*

$$\langle T(n)f, g\rangle = \langle f, T(n)g\rangle$$

for all n.

By Theorem 7.5 part (1), it is enough to prove this for n a prime p. Recall that $M(p)$ denote the set of 2×2 integer matrices with determinant p.

Proposition 7.10. *There exists a common set of representatives $\{\alpha_i\}$ for the set of left orbits $\mathrm{SL}_2(\mathbb{Z})\backslash M(p)$ and for the set of right orbits $M(p)/\mathrm{SL}_2(\mathbb{Z})$.*

Proof. Let $\alpha, \beta \in M(p)$. Using row and column operations there exists $U_\alpha, V_\alpha \in \mathrm{SL}_2(\mathbb{Z})$ such that $U_\alpha \alpha V_\alpha = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}$ and likewise for β. Thus we have $U_\alpha \alpha V_\alpha = U_\beta \beta V_\beta$ so $U_\beta^{-1} U_\alpha \alpha = \beta V_\beta V_\alpha^{-1} = \gamma$, say. Then $\mathrm{SL}_2(\mathbb{Z})\alpha = \mathrm{SL}_2(\mathbb{Z})\gamma$ and $\beta\mathrm{SL}_2(\mathbb{Z}) = \gamma\mathrm{SL}_2(\mathbb{Z})$. The result now follows by taking (α, β) to run through pairs (α_i, β_i) where we have partitions $M(p) = \cup_i \mathrm{SL}_2(\mathbb{Z})\alpha_i = \cup_i \beta_i \mathrm{SL}_2(\mathbb{Z})$. □

Example 7.11 Continuing the calculations in Example 7.4 we take our left coset representatives α to be

$$\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 0 & p \end{pmatrix}, \cdots, \begin{pmatrix} 1 & p-1 \\ 0 & p \end{pmatrix}, \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$$

and corresponding right coset representatives β to be

$$\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & p \end{pmatrix}, \cdots, \begin{pmatrix} 1 & 0 \\ p-1 & p \end{pmatrix}, \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Now

$$\begin{pmatrix} 1 & i \\ 0 & p \end{pmatrix}\begin{pmatrix} 1 & -i \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ -i & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ i & p \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}.$$

So for the first p coset representatives we find

$$\gamma = U_\beta^{-1} U_\alpha \alpha = \begin{pmatrix} 1 & 0 \\ -i & 1 \end{pmatrix}^{-1} \cdot I \cdot \begin{pmatrix} 1 & i \\ 0 & p \end{pmatrix} = \begin{pmatrix} 1 & i \\ i & p+i^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ i & p \end{pmatrix}\begin{pmatrix} 1 & i \\ 0 & 1 \end{pmatrix}.$$

For the final coset representative observe that indeed

$$U_\alpha = U_\beta = V_\alpha = V_\beta = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

and so

$$\gamma = \alpha = \beta = \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

For $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(p)$ write $\alpha' := \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = p\alpha^{-1} \in M(p)$.

Corollary 7.12. *Let $\{\alpha_i\}$ be a common set of coset representatives for the left orbits $\mathrm{SL}_2(\mathbb{Z})\backslash M(p)$ and for the right orbits $M(p)/\mathrm{SL}_2(\mathbb{Z})$. Then $\{\alpha_i'\}$ is also such a set.*

Proof. We have that

$$M(p) = \bigcup_i \mathrm{SL}_2(\mathbb{Z})\alpha_i = \bigcup_i \alpha_i \mathrm{SL}_2(\mathbb{Z}) \text{ (disjoint unions).}$$

Write $M(p)^{-1}$ for the set of inverses of elements in $M(p)$. Note that matrices in $pM(p)^{-1}$ are of the form $\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ for $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M(p)$ and thus $M(p) = pM(p)^{-1}$. So we find that

$$M(p) = pM(p)^{-1} = \bigcup_i p(\alpha_i \mathrm{SL}_2(\mathbb{Z}))^{-1} = \bigcup_i p\mathrm{SL}_2(\mathbb{Z})\alpha_i^{-1} = \bigcup_i \mathrm{SL}_2(\mathbb{Z})\alpha_i'$$

and likewise

$$M(p) = pM(p)^{-1} = \bigcup_i p(\mathrm{SL}_2(\mathbb{Z})\alpha_i)^{-1} = \bigcup_i p\alpha_i^{-1}\mathrm{SL}_2(\mathbb{Z}) = \bigcup_i \alpha_i'\mathrm{SL}_2(\mathbb{Z})$$

as required. □

We now prove the theorem.

Proof. Let $\{\alpha_i\}$ be a set of coset representatives for the left orbits $\mathrm{SL}_2(\mathbb{Z})\backslash M(p)$ such that $\{\alpha_i'\}$ is also a set of coset representatives for the left orbits. Then

$$\begin{aligned} \langle T(p)f, g\rangle &= p^{\frac{k}{2}-1}\textstyle\sum_i \langle f|_k\alpha_i, g\rangle \\ &= p^{\frac{k}{2}-1}\textstyle\sum_i \langle f, g|_k\alpha_i^{-1}\rangle \\ &= p^{\frac{k}{2}-1}\textstyle\sum_i \langle f, g|_k p\alpha_i^{-1}\rangle \\ &= p^{\frac{k}{2}-1}\textstyle\sum_i \langle f, g|_k\alpha_i'\rangle \\ &= \langle f, T(p)g\rangle \end{aligned}$$

The first equality is from the definition (using our particular choice of coset representatives) and linearity of the Petersson product, the second is from Proposition 7.6, the third since the scalar p acts trivially under our action, and the final one since the α_i' are also a set of representatives for the action on the left of $\mathrm{SL}_2(\mathbb{Z})$ on $M(p)$. □

Thus on the space of cusps forms the Hecke operators are Hermitian and the we have the following.

Corollary 7.13. *The eigenvalues of $T(n)$ on $S_k(\mathrm{SL}_2(\mathbb{Z}))$ (k even) are (totally) real algebraic integers.*

Proof. Using a basis for the space of cusps forms with coefficients in $\mathbb{Z}$ (Miller basis: Exercises), we see that the characteristic polynomial of $T(n)$ has integer coefficients, hence its roots are algebraic integers. Moreover, since $T(n)$ is Hermitian w.r.t. the Petersson inner product, by Part A algebra its eigenvalues are real numbers. □

Recall from Part A Linear Algebra that two commuting and diagonalisable linear maps on a finite dimensional vector space may be "simultaneously diagonalised". More general, given any set $\mathcal{T}$ of commuting and diagonalisable linear maps on a finite dimensional vector space V, one may decompose $V = E_1 \oplus \cdots \oplus E_r$ into subspaces E_i such that for every $T \in \mathcal{T}$ and $1 \leq i \leq r$, we have $T(E_i) \subseteq E_i$ and moreover $T|_{E_i}$ is multiplication by a scalar (one of the eigenvalues of T). (To see this, let us call a decomposition $V = V_1 \oplus \cdots \oplus V_s$, $\mathcal{T}$-stable if $T(V_i) \subseteq V_i$ for every $T \in \mathcal{T}$ and every $1 \leq i \leq s$. Start with any $\mathcal{T}$-stable decomposition $V = V_1 \oplus \cdots \oplus V_s$, e.g. just $V = V_1$. If there is some $T \in \mathcal{T}$ which does not act as a scalar on some V_i in the decomposition, then diagonalise that T on V_i; that is, decompose $V_i = V_i^{(1)} \oplus \cdots \oplus V_i^{(s_i)}$ with T acting as a scalar on each $V_i^{(j)}$. Since the linear maps in $\mathcal{T}$ commute, the new decomposition $V = V_1 \oplus \cdots \oplus V_{i-1} \oplus V_i^{(1)} \oplus \cdots \oplus V_i^{(s_i)} \oplus V_{i+1} \oplus \cdots \oplus V_s$ is also $\mathcal{T}$-stable, and has more summands. Continuing in this way, since V is finite-dimensional and every summand has dimension ≥ 1, the process eventually terminates and we have the decomposition we require.)

Corollary 7.14. *The space $S_k(\mathrm{SL}_2(\mathbb{Z}))$ has a basis consisting of simultaneous eigenvectors for the commuting Hecke operators.*

Proof. We use two facts from linear algebra. First, self-adjoint (Hermitian) operators on a finite-dimensional complex vector space can be diagonalised (and have real eigenvalues and orthogonal eigenspaces). Second, as just explained, any set of commuting and diagonalisable linear operators on such a space can be simultaneously diagonalised. Putting these together one sees that there is a basis $f_1, \cdots, f_d$ for $S_k(\mathrm{SL}_2(\mathbb{Z}))$ with each f_j an eigenvector for every Hecke operator $T(n)$. That is, we have the decomposition $S_k(\mathrm{SL}_2(\mathbb{Z})) = E_1 \oplus \cdots \oplus E_r$ into simultaneous eigenspaces and take a union of bases for these spaces. □

Let $E \subseteq S_k(\mathrm{SL}_2(\mathbb{Z}))$ denote an eigenspace for all of the Hecke operators $T(n)$ with $n \in \mathbb{N}$. Thus there exist (real algebraic) numbers λ_n for all $n \in \mathbb{N}$ with $T(n)f = \lambda_n f$ for all $f \in E$. Choose $f = \sum_{n \geq 1} c(n)q^n \in E$ and normalise it so that $c(1) = 1$. (Note that we may assume $c(1) \neq 0$. For if $n \geq 2$ is such that $c(n)q^n$ is the leading term of $f \in E$, then writing $n = pn'$ for some prime p we may apply the formula for the Hecke operator T_p to discover that $T(p)f$ has leading term $c(n)q^{n'}$—note $c(n'/p) = 0$ if $p|n'$—but we also know $T(p)f = \lambda_p f$ and comparing coefficients of $q^{n'}$ in both we find $c(n) = 0$, a contradiction.) By Theorem 7.5 Part (2), we also know that the coefficient of q in $T(n)f$ is $c(n)$. Thus we find $c(n) = \lambda_n$. This was true for all $n \in \mathbb{N}$ and so we must have $f = \sum_{n \geq 1} \lambda_n q^n$. We have proved:

Proposition 7.15 (Multiplicity one)**.** *Each common eigenspace for all Hecke operators on $S_k(\mathrm{SL}_2(\mathbb{Z}))$ ($k \geq 12$ even) is one dimensional.*

So the space $S_k(\mathrm{SL}_2(\mathbb{Z}))$ has a basis consisting of cusp forms f, called *eigenforms*, whose coefficients are totally real algebraic integers each of which is a simultaneous eigenvector for all of the Hecke operators $T(n)$ ($n \in \mathbb{N}$), and which are mutually orthogonal with respect to the Petersson inner product. (Any pair of eigenforms are orthogonal because they lie in different eigenspaces for some Hecke operator and the different eigenspaces of a given Hecke operator are orthogonal.)

Example 7.16

From our dimension formula $S_{12}(\mathrm{SL}_2(\mathbb{Z}))$ has dimension 1 and is spanned by the form

$$\Delta(q) = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 - 6048q^6 - 16744q^7 + 84480q^8 - 113643q^9 - 115920q^{10} + \cdots$$

Writing $\tau(n)$ for the nth Fourier coefficient, since Δ is an eigenform we find that

$$\tau(m)\tau(n) = \tau(mn) \text{ when } \gcd(m,n) = 1$$

$$\tau(p^r)\tau(p) = \tau(p^{r+1}) + p^{11}\tau(p^{r-1}) \text{ for } p \text{ prime and } r \geq 1.$$

Note that although $\Delta(q)$ has nice integral Fourier coefficients, the inner product $\langle \Delta, \Delta \rangle$ is a (conjecturally) transcendental number. Numerical one finds[7]

$$\langle \Delta, \Delta \rangle = 0.00000103536205680432092234781681222516459322491\cdots$$

Similarly $S_k(\mathrm{SL}_2(\mathbb{Z}))$ for $k = 16, 18, 20, 22, 26$ is spanned by a single eigenform with integer Fourier coefficients. The space $S_{24}(\mathrm{SL}_2(\mathbb{Z}))$ has dimension 2. From Sheet 4 the characteristic polynomial of $T(2)$ is $P_2(x) := x^2 - 1080x - 20468736$. Thus using $T(2)$ and the Miller basis we may compute a pair of eigenforms $f_a = q + aq^2 + \cdots$ and $f_b = q + bq^2 + \cdots$ defined over $\mathbb{Q}(a)$, where $P_2(a) = 0$ and $b = -a + 1080$ is the other root of $P_2(x)$. Note that $P_2(x)$ has positive discriminant, and thus two real roots. Thus we find two eigenforms with real Fourier coefficients.

One expects that in general $S_k(\mathrm{SL}_2(\mathbb{Z}))$ is spanned by the Galois conjugates of a single eigenform f_a defined over an extension of degree $\dim(S_k(\mathrm{SL}_2(\mathbb{Z})))$, and thus in weight > 26 there are not expected to be any nice cuspidal eigenforms for the full modular group with integral Fourier coefficients.

The methods we have used generalise (with some extra work and new features) to the spaces $S_k(\Gamma_0(N))$ and $S_k(\Gamma_1(N))$. Of particular arithmetic interest are the spaces $S_1(\Gamma_1(N))$ and $S_2(\Gamma_0(N))$. For $S_2(\Gamma_0(N))$ the cuspidal eigenforms with integer Fourier coefficients may be related to elliptic curves over $\mathbb{Q}$ of "conductor" N. For $S_1(\Gamma_1(N))$ the cuspidal eigenforms may be related to irreducible representations into $\mathrm{GL}_2(\mathbb{C})$ of Galois groups of finite extensions of $\mathbb{Q}$. See Examples 1.2 and 1.3: note that the recurrences amongst the Fourier coefficients is a consequence of these modular forms being eigenforms for suitably defined Hecke operators.

References

[1] [DS] F. Diamond and J. Shurman, A first course in modular forms, Springer GTM.

[2] [RG] R.C. Gunning, Lectures on modular forms, Princeton University Press, 1962.

[3] [NK] N. Koblitz, Introduction to elliptic curves and modular forms, Springer GTM

[4] [JSM] J.S. Milne, Modular functions and modular forms. Available at http://www.jmilne.org/math/CourseNotes/mf.html

[7] Page 2 of Henri Cohen's paper "Haberland's formula and numerical computation of Petersson scalar products".

[5] [JPS] J-P. Serre, A course in arithmetic, Springer GTM

[6] [WS] W. Stein, Modular forms, AMS (and available on-line).