Modular Forms *

Zhiyuan Bai

Compiled on February 24, 2022

This document serves as a set of revision materials for the Cambridge Mathematical Tripos Part III course *Modular Forms* in Lent 2021. However, despite its primary focus, readers should note that it is NOT a verbatim recall of the lectures, since the author might have made further amendments in the content. Therefore, there should always be provisions for errors and typos while this material is being used.

Contents

*Based on the lectures under the same name taught by Prof. J. A. Thorne in Lent 2021.

0 Introduction and Motivations

Denote the complex upper half-plane by $\mathfrak{h} = \{\tau \in \mathbb{C} : \operatorname{Im}\tau > 0\}$. Naturally, certain matrix groups act on it by Möbius transformations.

Lemma 0.1. $\mathrm{GL}_2(\mathbb{R})^+ = \{M \in \mathrm{GL}_2(\mathbb{R}) : \det M > 0\}$ *acts transitively on* $\mathfrak{h}$ *via Möbius transformations.*

Consequently, subgroups (in particular $\mathrm{SL}_2(\mathbb{Z})$) of $\mathrm{GL}_2(\mathbb{R})^+$ also acts on $\mathfrak{h}$ by Möbius transformations.

Proof. Let $\tau \in \mathfrak{h}$ and $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})^+$, then by brute-force computation,

$$\operatorname{Im}(g\tau) = \frac{\det(g)\operatorname{Im}(\tau)}{|c\tau + d|^2}$$

For transitivity, just observe that $x + iy = \left(\begin{smallmatrix} y & x \\ 0 & 1 \end{smallmatrix}\right) \cdot i$. □

Definition 0.1. Let $k \in \mathbb{Z}$, $f : \mathfrak{h} \to \mathbb{C}_\infty$ and $g \in \mathrm{GL}_2(\mathbb{R})^+$. Define $f|_k[g] : \mathfrak{h} \to \mathbb{C}_\infty$ by the formula

$$f|_k[g](\tau) = f(g\tau)\det(g)^{k-1}j(g,\tau)^{-k}$$

where $j(g,\tau) = c\tau + d$ if $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$.

Lemma 0.2. *The above definition defined a right action of* $\mathrm{GL}_2(\mathbb{R})^+$ *on the set of functions* $f : \mathfrak{h} \to \mathbb{C}_\infty$.

Proof. This reduces to the cocycle identity $j(gh,\tau) = j(h,\tau)j(g,h\tau)$, which can be verified by simple computation. □

Definition 0.2. Let $k \in \mathbb{Z}$ and let $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ be a subgroup of finite index. A function $f : \mathfrak{h} \to \mathbb{C}_\infty$ is called weakly modular of weight k and level Γ if it is meromorphic and $\forall \gamma \in \Gamma, f|_k[\gamma] = f$.

For it to be a modular form, we actually need a little bit more. But before jumping to actual definition and further theory, let's look at some motivating examples of it first.
Modular forms were first studied in the context of elliptic functions. Suppose E is an elliptic curve over $\mathbb{C}$ and let ω be a non-vanishing holomorphic differential on E. Then there is a unique holomorphic isomorphism $\psi : \mathbb{C}/\Lambda \to E(\mathbb{C})$ such that $\psi^*(\omega) = \mathrm{d}z$, where Λ is a lattice. Furthermore, E can be defined by the equation $y^2 = x^3 - 60G_4(\Lambda)x - 140G_6(\Lambda)$ where $G_k(\Lambda) = \sum_{\omega \in \Lambda\setminus\{0\}} \omega^{-k}$. But in fact, if $\tau \in \mathfrak{h}$, then $\Lambda_\tau = \mathbb{Z}\tau \oplus \mathbb{Z}$ is a lattice and the functions $G_k(\tau) = G_k(\Lambda_\tau)$ (when $k \geq 4$) are modular forms.
If $f : \mathfrak{h} \to \mathbb{C}$ is a modular form, then f has a Fourier expansion $f(\tau) = \sum_{n \geq 0} a_n e^{2\pi i n\tau/k}$ for some $a_n \in \mathbb{C}, k \in \mathbb{N}$. The coefficients a_n often carries very very useful arithmetic information. For example, consider the theta function $\theta(\tau) = \sum_{n \in \mathbb{Z}} e^{\pi i n^2 \tau}$, then θ^{2k} is (we will see later) a modular form (of weight k) for any integer $k \geq 2$. Its Fourier expansion is $\theta^{2k}(\tau) = \sum_{n]ge1} r_{2k}(n)e^{\pi i n\tau}$ where $r_{2k}(n)$ is actually the number of ways of writing n as a sum of $2k$ perfect squares. So by relating θ^{2k} to other modular forms with known Fourier series, we can get information about $r_{2k}(n)$ which is a very interesting subject in number theory. For example, we can show that $r_4(n) = 8\sum_{d|n, 4\nmid d} d$.

Consider the Riemann zeta function $\zeta(s) = \sum_{n\geq 1} n^{-s}$ which is well-defined for $\operatorname{Re} s > 1$ and can be continued meromorphically to $\mathbb{C}$. Studies in zeta function also shows that it satisfies a functional equation relating $\zeta(s)$ and $\zeta(1-s)$, and we have the Euler product

$$\zeta(s) = \prod_{p \text{ prime}} \frac{1}{1-p^{-s}}$$

There is actually a big family of functions that shares analogous properties to these, e.g. the Dirichlet L-functions

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

associated with a character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$. Just like ζ, $L(\chi, s)$ also possess significant arithmetic importance, namely it can be used to prove Dirichlet's theorem on arithmetic progressions. In general, L-functions are functions of the form $L(s) = \sum_n a_n n^{-s}$ which satisfies analogous properties to ζ. And as one can expect, most of them are related to number theory. However, the tricky part is that it is not usually easy to show that they have these properties, hence even harder to utilise them in problems in number theory. This is where modular forms come in. Turns out, modular forms can be used to construct L-functions satisfying these properties. The details involve the theory of Hecke operators, which will be discussed later in the course.
One last motivation to talk about is the Langlands programme, which predicts relations between objects in number theory and (certain generalisations of) modular forms. This includes a special case of the Shimura-Taniyama-Weil Conjecture (aka Modularity Theorem) which asserts a bijection between elliptic curves over $\mathbb{Q}$ up to isogeny and certain modular forms. This bijection may be characterised by saying the L-functions of elliptic curves corresponds to L-functions of associated modular form.

1 Modular Forms on $\mathrm{SL}_2(\mathbb{Z})$

1.1 Modular Functions and the Valence Formula

Let f be a weakly modular function of weight k and level $\mathrm{SL}_2(\mathbb{Z})$. Then $f(\tau) = f|_k \left[\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)\right](\tau) = f(\tau+1)$. This means that 1 is a period of f, hence there exists a meromorphic function $\tilde{f} : \{q \in \mathbb{C} : 0 < |q| < 1\} \to \mathbb{C}$ such that $\tilde{f}(e^{2\pi i\tau}) = f(\tau)$.

Definition 1.1. We say a weakly modular function f of weight k and level $\mathrm{SL}_2(\mathbb{Z})$ is meromorphic (resp. holomorphic) at ∞ if $\tilde{f}$ is meromorphic (resp. holomorphic) at 0.
We say f vanishes at ∞ if $\tilde{f}$ is holomorphic and vanishes at 0.

Remark. If f is meromorphic at ∞, then $\tilde{f}$ has a Laurent expansion

$$\tilde{f}(q) = \sum_{n\in\mathbb{Z}} a_n q^n$$

near 0 and $a_n = 0$ for sufficiently small n. Consequently we get a q-expansion (or Fourier expansion)

$$f(\tau) = \sum_{n\in\mathbb{Z}} a_n q^n, q = e^{2\pi i\tau}$$

valid for sufficiently large $\operatorname{Im}\tau$.
If f is actually holomorphic at ∞, then $a_n = 0$ when $n < 0$, so $f(\infty) = a_0$.

Definition 1.2. Let $f : \mathfrak{h} \to \mathbb{C}_\infty$ be a weakly modular function of weight k and level $\mathrm{SL}_2(\mathbb{Z})$. It is a modular function if f is meromorphic at ∞. It is a modular form if f is holomorphic in $\mathfrak{h}$ and at ∞. It is a cuspidal modular form if it is a modular form and vanishes at ∞.
We write $M_k(\mathrm{SL}_2(\mathbb{Z}))$ for the $\mathbb{C}$-vector space of modular forms of weight k and level $\mathrm{SL}_2(\mathbb{Z})$. Similarly we write $S_k(\mathrm{SL}_2(\mathbb{Z}))$ for the $\mathbb{C}$-vector space of cuspidal modular forms of weight k and level $\mathrm{SL}_2(\mathbb{Z})$.

Example 1.1. If $\tau \in \mathfrak{h}$, write $\Lambda_\tau = \mathbb{Z}\tau \oplus \mathbb{Z}$. For $k \in \mathbb{Z}$, we define

$$G_k(\tau) = \sum_{\omega\in\Lambda_\tau\setminus\{0\}} \omega^{-k}$$

If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, then $\Lambda_{\gamma\tau} = j(\gamma,\tau)^{-1}\Lambda_\tau$, so, if G_k converges absolutely,

$$\begin{aligned} G_k|_k[\gamma](\tau) = G_k(\gamma\tau)j(\gamma,\tau)^{-k} &= \sum_{\omega\in\Lambda_{\gamma\tau}\setminus\{0\}} (\omega j(\gamma,\tau))^{-k} \\ &= \sum_{\omega\in\Lambda_\tau\setminus\{0\}} \omega^{-k} = G_k(\tau) \end{aligned}$$

Proposition 1.1. *Suppose $k \geq 4$ and k is even. Then $G_k(\tau)$ converges absolutely and uniformly in compact subsets of $\mathfrak{h}$. Moreover, $G_k(\tau)$ is holomorphic at ∞ and $G_k(\infty) = 2\zeta(k)$. In particular, $G_k \in M_k(\mathrm{SL}_2(\mathbb{Z}))$.*

Remark. We are not really disappointed at k being required to be even. In fact, there is no nonzero weakly modular functions of odd weight k and level $\mathrm{SL}_2(\mathbb{Z})$ since we can take $T = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ which gives $f|_k[T] = (-1)^k f = -f$.

Proof. Fix $A \geq 1$ and define $\Omega_A = \{\tau \in \mathfrak{h} : |\operatorname{Re}\tau| \leq A, \operatorname{Im}\tau \geq 1/A\}$. We shall show absolute and uniform convergence of G_k in Ω_A. Note that if $\tau \in \Omega_A$, then $|\tau + x| \geq 1/A$ for any $x \in \mathbb{R}$ and $|\tau + x| \geq |x|/2$ if in addition that $|x| \geq 2A$. So $|\tau + x| \geq (2A^2)^{-1}\max\{1, |x|\}$. Hence for any $\tau \in \Omega_A$,

$$\begin{aligned} \sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} |m\tau + n|^{-k} &= \sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} |m|^{-k}\left|\tau + \frac{n}{m}\right|^{-k} \\ &\leq \sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \frac{|m|^{-k}}{(2A^2)^{-k}}(\max\{1, |n/m|\})^{-k} \\ &\leq (2A^2)^k \sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \min\{|m|^{-k}, |n|^{-k}\} \\ &= (2A^2)^k \sum_{r\in\mathbb{N}} r^{-k} 8r \\ &= (2A^2)^k 8\zeta(k-1) \end{aligned}$$

which shows absolute and uniform convergence.
To see it is holomorphic at ∞ and $G_k(\infty) = 2\zeta(k)$, it suffices to show that $G_k(x+iy) \to 2\zeta(k)$ as $y \to \infty$, which we can just calculate

$$\sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} \lim_{\operatorname{Im}\tau\to\infty} (m\tau+n)^{-k} = \sum_{n\in\mathbb{Z}\setminus\{0\}} n^{-k} = 2\zeta(k)$$

as desired. □

Definition 1.3. The normalised Eisenstein series is defined as

$$E_k(\tau) = \frac{1}{2\zeta(k)} G_k(\tau) = 1 + \sum_{n\geq 1} a_n q^n$$

We shall see later that these a_n are actually rational numbers of bounded (minimal) denominators.

Remark. if $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ and $g \in M_l(\mathrm{SL}_2(\mathbb{Z}))$, then $fg \in M_{k+l}(\mathrm{SL}_2(\mathbb{Z}))$.

Consequently E_4^3 and $E_6^2 \in M_{12}(\mathrm{SL}_2(\mathbb{Z}))$ and $E_4^3 = E_6^2$ at ∞. So $\Delta = (E_4^3 - E_6^2)/1728 \in S_{12}(\mathrm{SL}_2(\mathbb{Z}))$. We'll see later that $\Delta = \sum_{n\geq 1} b_n q^n$ is nonzero with $b_1 = 1$ and $\forall n, b_n \in \mathbb{Z}$.
We now turn to study a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathfrak{h}$. Write $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z}), \overline{\Gamma(1)} = \mathrm{SL}_2(\mathbb{Z})/\{\pm 1\}$. We are interested in the domains $\mathcal{F} = \{\tau \in \mathfrak{h} : -1/2 \leq \operatorname{Re}\tau \leq 1/2, |\tau| \geq 1\}$ (draw it!) and $\mathcal{F}' = \{\tau \in \mathcal{F} : \operatorname{Re}\tau < 1/2, |\tau| = 1 \implies \operatorname{Re}\tau \leq 0\}$ obtained by removing part of the boundary of $\mathcal{F}$.

Proposition 1.2. *1. $\mathcal{F}'$ is a fundamental domain for the action of $\overline{\Gamma(1)}$ on $\mathfrak{h}$. More precisely, $\forall \tau \in \mathfrak{h}, \exists! \tau' \in \overline{\Gamma(1)}\cdot\tau$ such that $\tau' \in \mathcal{F}'$.*
2. $\mathrm{Stab}_{\overline{\Gamma(1)}}(\tau)$ is trivial for any $\tau \in \mathcal{F}' - \{i, \rho\}$, while $\mathrm{Stab}_{\overline{\Gamma(1)}}(i) = \{1, S\}$ and $\mathrm{Stab}_{\overline{\Gamma(1)}}(\rho) = \{1, ST, (ST)^2\}$ where

$$\rho = -\frac{1}{2} + \frac{\sqrt{3}}{2}i, T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

3. $\overline{\Gamma(1)}$ is generated by S and T.

Proof. We shall first prove that any $\tau \in \mathfrak{h}$ is $\overline{\Gamma(1)}$-conjugate to an element of $\mathcal{F}$. We already know that if $\tau \in \mathfrak{h}$ and $\gamma \in \overline{\Gamma(1)}$, then $\operatorname{Im}(\gamma\tau) = \operatorname{Im}\tau/|j(\gamma,\tau)|^2$ For $\tau \in \mathfrak{h}$, as $(c,d) \in \mathbb{Z}^2 \setminus \{0\}$, the numbers $|c\tau + d|$ achieve a positive minimum (since $\Lambda_\tau = \mathbb{Z}\tau \oplus \mathbb{Z}$ is a lattice). Consequently, the numbers $\operatorname{Im}(\gamma\tau)$ for $\gamma \in \Gamma(1)$ achieve a maximum. So by applying certain elements of $\Gamma(1)$ we can arrive at $\forall \gamma \in \Gamma(1), \operatorname{Im}\tau \geq \operatorname{Im}(\gamma\tau)$ and $\operatorname{Re}(\tau) \in [-1/2, 1/2]$. One can check that these properties means $\tau \in \mathcal{F}$ by observing that $\operatorname{Im}(S\tau) = \operatorname{Im}(\tau)/|\tau|^2 \leq \operatorname{Im}\tau \implies |\tau| \geq 1$.
This implies that any $\tau \in \mathfrak{h}$ is $\overline{\Gamma(1)}$-conjugate to an element of $\mathcal{F}'$ since we can obviously get from $\mathcal{F}$ to $\mathcal{F}'$ by S and T.
Now suppose $\gamma\tau = \tau'$ for $\tau, \tau' \in \mathcal{F}', \gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \overline{\Gamma(1)}$. WLOG (up to possible swap of τ and τ') we have $\operatorname{Im}\tau' = \operatorname{Im}(\gamma\tau) \geq \operatorname{Im}\tau \implies |j(\gamma,\tau)| \leq 1$. Then since $\operatorname{Im}\tau \geq \sqrt{3}/2$, $|j(\gamma,\tau)| = |c\tau + d| \geq c\sqrt{3}/2$, which means that $|c| \leq 1$.
If $c = 0$, then γ is linear, which then forces $\gamma = 1, \tau = \tau'$. If $c = \pm 1$, WLOG $c = 1$. As $|\tau + d| \leq 1$, either $d = 0$ which means $|\tau| = 1$ or $d = 1$ which means

$\tau = \rho$. If $d = 0, |\tau| = 1$, then $\gamma\tau = a - \tau^{-1}$. We have $\operatorname{Re}\tau, \operatorname{Re}(\gamma\tau) = a - \operatorname{Re}\tau$ both residing in $[-1/2, 0]$, so the only possibilities are

$$\begin{cases} \operatorname{Re}\tau = -1/2, a = -1, \tau = \rho, \gamma = (ST)^2 \\ \operatorname{Re}\tau = 0, a = 0, \tau = i, \gamma = S. \end{cases}$$

If $d = 1, \tau = \rho$, then $\gamma\rho = (a\rho + b)/(\rho + 1)$. But $\operatorname{Im}(\gamma\rho) = \operatorname{Im}(\rho)/(\rho + 1) = \operatorname{Im}\rho$ which actually means $\gamma\rho = \rho$ since ρ has minimal imaginary part in $\mathcal{F}'$, hence $a = 0, b = -1$ and $\gamma = ST$.
It remains to show that S, T generate $\overline{\Gamma(1)}$. Take $G = \langle S, T\rangle$ and $\tau \in \mathfrak{h}$. WLOG $\forall \gamma \in G, \operatorname{Im}(\gamma\tau) \leq \operatorname{Im}\tau$ and $\operatorname{Re}\tau \in [-1/2, 1/2]$ as usual, so $\tau \in \mathcal{F}$ as $\operatorname{Im}(S\tau) = \operatorname{Im}(\tau)/|\tau|^2 \leq \operatorname{Im}\tau \implies |\tau| \geq 1$. Choose any $\tau \in \operatorname{Int}(\mathcal{F})$ and $\gamma \in \overline{\Gamma(1)}$. As $\gamma\tau \in \mathfrak{h}, \exists \delta \in G$ such that $\delta\gamma\tau \in \operatorname{Int}(\mathcal{F})$ and hence $\delta\gamma = 1 \implies \gamma = \delta^{-1} \in G$. □

For $p \in \overline{\Gamma(1)}\backslash\mathfrak{h}$, we write $e_p = |\operatorname{Stab}_{\overline{\Gamma(1)}}(\tau)|$ for any representative $\tau \in p$. Then $e_p = 1$ except when $e_{\overline{\Gamma(1)}\cdot\rho} = 3$ and $e_{\overline{\Gamma(1)}\cdot i} = 2$.

Definition 1.4. Suppose that f is a modular function of weight k and level $\mathrm{SL}_2(\mathbb{Z})$, we define $v_p(f)$ to be the order of f at $\tau \in p \in \overline{\Gamma(1)}\backslash\mathfrak{h}$.

Note that this is independent of the choice of τ since for any $\gamma \in \Gamma(1)$ we have $f(\gamma\tau)j(\gamma, \tau)^{-k} = f(\tau)$ and j is holomorphic and non-vanishing.

Definition 1.5. We define $v_\infty(f) = \inf_{n\in\mathbb{Z}}\{n \in \mathbb{Z} : a_n \neq 0\}$ where $f(\tau) = \sum_{n\in\mathbb{Z}} a_n q^n$ is the q-expansion of f. (equivalently, this is the order of f at $q = 0$).

Theorem 1.3 (Valence Formula)**.** *Let f be a nonzero modular function of weight k and level* $\mathrm{SL}_2(\mathbb{Z})$*, then*

$$v_\infty(f) + \sum_{p\in\overline{\Gamma(1)}\backslash\mathfrak{h}} \frac{1}{e_p} v_p(f) = \frac{k}{12}$$

The sum is, of course, always finite.

Proof. We shall first prove the theorem assuming that f has no zeros or poles on $\partial\mathcal{F}$. Since f is meromorphic at ∞, there exists $R > 0$ such that f has no zeros or poles in $\{\tau \in \mathfrak{h} : \operatorname{Im}\tau \geq R\}$. Consider the region $\{\tau \in \mathcal{F} : \operatorname{Im}\tau \leq R\}$ and let γ be its (anticlockwise) boundary. Let's label some points on γ:

$$A = -\frac{1}{2} + iR < B = \rho, C = i, D = \rho + 1, E = \frac{1}{2} + iR$$

The argument principle then gives

$$\frac{1}{2\pi i}\oint_\gamma \frac{f'(z)}{f(z)}\,\mathrm{d}z = \sum_{\tau\in\operatorname{Int}\mathcal{F}} v_\tau(f)$$

What is this countor integral? Well, $f(\tau) = f(\tau + 1)$, so

$$\int_A^B \frac{f'(z)}{f(z)}\,\mathrm{d}z = -\int_D^E \frac{f'(z)}{f(z)}\,\mathrm{d}z$$

Also, the image of the path from E to A under $\tau \mapsto e^{2\pi i\tau}$ is a clockwise circle α around $q = 0$, so

$$\frac{1}{2\pi i}\int_E^A \frac{f'(z)}{f(z)}\,\mathrm{d}z = \frac{1}{2\pi i}\oint_\alpha \frac{\tilde{f}'(q)}{\tilde{f}(q)}\,\mathrm{d}q = -v_0(\tilde{f}) = -v_\infty(f)$$

Next, the path from C to D is the image of the path from C to B under S, so

$$\frac{1}{2\pi i}\int_D^C \frac{f'(z)}{f(z)}\,\mathrm{d}z = \frac{1}{2\pi i}\int_B^C \frac{(f\circ S)'(z)}{(f\circ S)(z)}\,\mathrm{d}z = \frac{1}{2\pi i}\int_B^C \frac{k}{\tau}\,\mathrm{d}\tau + \frac{1}{2\pi i}\int_B^C \frac{f'(z)}{f(z)}\,\mathrm{d}z$$

Therefore,

$$\frac{1}{2\pi i}\left(\int_B^C \frac{f'(z)}{f(z)}\,\mathrm{d}z + \int_C^D \frac{f'(z)}{f(z)}\,\mathrm{d}z\right) = -\frac{1}{2\pi i}\int_B^C \frac{k}{\tau}\,\mathrm{d}\tau = \frac{1}{2\pi i}\int_C^B \frac{k}{\tau}\,\mathrm{d}\tau = \frac{k}{12}$$

Putting everything together gives the formula (since every zeros/poles p in the interior of $\mathcal{F}$ has $e_p = 1$).
What if there are zeros and poles of f on the contour? Well, if they are not $i, \rho, \rho+1$, then it is easy to deal with – just modify the contours by a properly oriented perturbation and use a limit argument. For $i, \rho, \rho+1$, we make the following observation: If g is a meromorhic function define in an open neighbourhood of $z = 0$ and γ_ϵ (for sufficiently small ϵ) be the path $\gamma_\epsilon(t) = \epsilon e^{\theta_0 + t\theta}$, then

$$\lim_{\epsilon\to 0}\frac{1}{2\pi i}\int_{\gamma_\epsilon}\frac{g'(z)}{g(z)}\,\mathrm{d}z = \frac{\theta}{2\pi}v_0(g)$$

(This can be easily shown by writing $g(z) = z^{v_0(g)}h(z)$ where h is nonzero and holomorphic at 0). So in the case where f has zeros or poles at i or $\rho, \rho+1$, we can modify the contour by a small ϵ-arc around them that are oriented inside $\mathcal{F}$. Sending $\epsilon \to 0$ then works everything out since the angles at $\rho, \rho+1$ are $\pi/3$ and $e_\rho = 3$ while the angle at i is π and $e_i = 2$. □

Modular forms are holomorphic in $\mathfrak{h} \cup \{\infty\}$, so it only have nonnegative orders. This fact allows us to do tonnes of stuff with this formula.

Example 1.2. Take $k = 4$ and $f = E_4 \in M_4(\mathrm{SL}_2(\mathbb{Z}))$. Then we get

$$v_\infty(E_4) + \sum_{p\in\overline{\Gamma(1)}\backslash\mathfrak{h}}\frac{1}{e_p}v_p(E_4) = \frac{1}{3}$$

which implies that E_4 has a simple zero at ρ and none else.
Take $k = 6$ and $f = E_6$ gives

$$v_\infty(E_6) + \sum_{p\in\overline{\Gamma(1)}\backslash\mathfrak{h}}\frac{1}{e_p}v_p(E_6) = \frac{1}{2}$$

So E_6 has a simple zero at i and none else.
Putting these two together shows that $\Delta = (E_4^3 - E_6^2)/1728 \in S_{12}(\mathrm{SL}_2(\mathbb{Z}))$ is nonzero at i. In particular $\Delta \neq 0$.
Also, applying the theorem to Δ gives

$$v_\infty(\Delta) + \sum_{p\in\overline{\Gamma(1)}\backslash\mathfrak{h}}\frac{1}{e_p}v_p(\Delta) = 1$$

But we already have $v_\infty(\Delta) \geq 1$ since Δ is cuspidal, so Δ has a simple zero at ∞ and is non-vanishing on $\mathfrak{h}$.

1.2 Structures of $M_k(\mathrm{SL}_2(\mathbb{Z}))$ and $S_k(\mathrm{SL}_2(\mathbb{Z}))$

Theorem 1.4. *Let $k \in 2\mathbb{Z}$.*
1. If $k < 0$ or $k = 2$, then $M_k(\mathrm{SL}_2(\mathbb{Z})) = 0$. Moreover, $M_0(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}$.
2. If $4 \leq k \leq 10$ or $k = 14$, then $M_k(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}E_k$.
3. If $k \geq 0$, then multiplication by Δ induces an isomorphism

$$M_k(\mathrm{SL}_2(\mathbb{Z})) \to S_{k+12}(\mathrm{SL}_2(\mathbb{Z}))$$

Proof. If $f \neq 0$, then

$$v_\infty(f) + \sum_{p \in \overline{\Gamma(1)}\backslash\mathfrak{h}} \frac{1}{e_p} v_p(f) = \frac{k}{12}$$

It follows straight away that $k \geq 0$ and $k \neq 2$. Suppose $f \in M_0(\mathrm{SL}_2(\mathbb{Z}))$, then there exists $\lambda \in \mathbb{C}$ such that $f - \lambda$ is cuspidal. The valence formula then yields a contradiction unless $f - \lambda = 0$, so f has to be a scalar. Therefore $M_0(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}$.
Suppose $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ and $4 \leq k \leq 10$ or $k = 14$, then $\exists \lambda \in \mathbb{C}, f - \lambda E_k \in S_k(\mathrm{SL}_2(\mathbb{Z}))$. If $f - \lambda E_k \neq 0$, then plugging it into the valence formula yields immediate contradiction, so the result follows.
For the last part of the theorem, we know the map is injective right away since $\Delta \neq 0$. It is surjective since Δ is nonvanishing in all of $\mathfrak{h}$ and has a simple zero at ∞ (which means that $f/\Delta \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ if $f \in S_{k+12}(\mathrm{SL}_2(\mathbb{Z}))$). □

Corollary 1.5. *For any $k \in \mathbb{Z}, k \geq 0$, we have*

$$\dim_\mathbb{C} M_k(\mathrm{SL}_2(\mathbb{Z})) = \begin{cases} \lfloor k/12 \rfloor, \text{ if } k \equiv 2 \pmod{12} \\ \lfloor k/12 \rfloor + 1, \text{ otherwise} \end{cases}$$

Proof. When $0 \leq k \leq 14$ the corollary follows immediately from the theorem. Also, $M_k(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}E_k \oplus S_k(\mathrm{SL}_2(\mathbb{Z}))$, so by part 3 of the theorem,

$$\dim_\mathbb{C} M_{k+12}(\mathrm{SL}_2(\mathbb{Z})) = 1 + \dim_\mathbb{C} M_k(\mathrm{SL}_2(\mathbb{Z}))$$

which is what we want. □

Corollary 1.6. *Let $k \geq 0$ be even. Then $M_k(\mathrm{SL}_2(\mathbb{Z}))$ is spanned as a vector space by the elements $\{E_4^a E_6^b : a, b \in \mathbb{Z}_{\geq 0}, 4a + 6b = k\}$.*

Proof. This certainly holds when $k \leq 10$. Also, if this holds for k, then choose $a, b \in \mathbb{Z}_{\geq 0}$ such that $4a + 6b = k + 12$. We know that $E_4^a E_6^b \in M_{k+12}$ and $M_{k+12}(\mathrm{SL}_2(\mathbb{Z})) = S_{k+12}(\mathrm{SL}_2(\mathbb{Z})) \oplus \mathbb{C}E_4^a E_6^b = \Delta M_k(\mathrm{SL}_2(\mathbb{Z})) \oplus \mathbb{C}E_4^a E_6^b$. The result for $k + 12$ follows. We conclude the corollary by induction. □

Definition 1.6. We define $j : \mathfrak{h} \to \mathbb{C}$ by the formula $j = E_4^3/\Delta$ which is a modular function of weight 0 and level $\mathrm{SL}_2(\mathbb{Z})$.

Remark. If $\tau \in \mathfrak{h}$, then $j(\tau)$ is the j-invariant of the elliptic curve $E_\tau = \mathbb{C}/\Lambda_\tau$.

Theorem 1.7. *j is holomorphic in $\mathfrak{h}$ and $v_\infty(j) = -1$, and it induces a bijection $\overline{\Gamma(1)}\backslash\mathfrak{h} \to \mathbb{C}$. Also, every modular function of weight 0 and level $\mathrm{SL}_2(\mathbb{Z})$ is a rational function of j.*

Remark. Later in the course, we will give $\overline{\Gamma(1)}\backslash\mathfrak{h} \cup \{\infty\}$ the structure of a compact Riemann surface. The theorem then means that j gives an isomorphism from this Riemann surface to the usual Riemann sphere.

Proof. We immediately have j being holomorphic on $\mathfrak{h}$ and having a simple zero at ∞ since E_4^3 is a non-cuspidal modular form and Δ is holomorphic in $\mathfrak{h}$ with a simple zero at ∞. For any $z \in \mathbb{C}$, we have

$$v_\infty(E_4^3 - z\Delta) + \sum_{p\in\overline{\Gamma(1)}\backslash\mathfrak{h}} \frac{1}{e_p} v_p(E_4^3 - z\Delta) = 1$$

Consequently $E_4^3 - z\Delta$ has exactly one zero $\tau \in \overline{\Gamma(1)}\backslash\mathfrak{h}$, which would be the unique solution to $j(\tau) = z$.
Suppose f is a nonzero modular function of weight 0 and level $\mathrm{SL}_2(\mathbb{Z})$, then WLOG (by multiplying f by terms of the form $j - j(\tau_0)$) f is holomorphic in $\mathfrak{h}$. Then $\exists n \geq 0$ such that $\Delta^n f \in M_{12n}(\mathrm{SL}_2(\mathbb{Z}))$. We already know that $M_{12n}(\mathrm{SL}_2(\mathbb{Z}))$ is generated by $\{E_4^a E_6^b : a, b \in \mathbb{Z}_{\geq 0}, 4a + 6b = 12n\}$, so it suffices to show that everything in this generating set is a rational function in j. If $4a + 6b = 12n$, then $a = 3p, b = 2q$ for some $p, q \in \mathbb{Z}_{\geq 0}$ with $p + q = n$, hence

$$\frac{E_4^a E_6^b}{\Delta^n} = \left(\frac{E_4^3}{\Delta}\right)^p \left(\frac{E_6^2}{\Delta}\right)^q = j^p (j - 1728)^q$$

as desired. □

Proposition 1.8. *Let $k \geq 4$ be an even integer, then the q-expansion of G_k is*

$$G_k(\tau) = 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!} \sum_{n=1}^\infty \sigma_{k-1}(n) q^n, \sigma_{k-1}(n) = \sum_{d|n} d^{k-1}$$

Proof. We have the identity

$$\pi\cot(\pi\tau) = \frac{1}{\tau} + \sum_{n=1}^\infty \left(\frac{1}{\tau - n} + \frac{1}{\tau + n}\right)$$

Note that $\pi\cot(\pi\tau) = \pi i(q+1)/(q-1) = -\pi i(1+q)(1+q+q^2+\cdots)$, so

$$-\pi i - 2\pi i \sum_{n\geq 1} q^n = \frac{1}{\tau} + \sum_{n=1}^\infty \left(\frac{1}{\tau - n} + \frac{1}{\tau + n}\right)$$

Differentiating $(k-1)$ times gives

$$\begin{aligned} -2\pi i \sum_{n\geq 1} (2\pi i n)^{k-1} q^n &= (-1)^{k-1}(k-1)! \left(\frac{1}{\tau^k} + \sum_{n=1}^\infty \left(\frac{1}{(\tau-n)^k} + \frac{1}{(\tau+n)^k}\right)\right) \\ &= (-1)^{k-1}(k-1)! \sum_{n\in\mathbb{Z}} (\tau+n)^{-k} \end{aligned}$$

(We can do all these sort of stuff since the series we deal with all converges absolutely and uniformly on compact sets.) So,

$$\begin{aligned} G_k(\tau) &= \sum_{(m,n)\in\mathbb{Z}^2\setminus\{0\}} (m\tau+n)^{-k} \\ &= 2\zeta(k) + 2\sum_{m\geq 1}\sum_{n\in\mathbb{Z}}(m\tau+n)^{-k} \\ &= 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!}\sum_{m,n\geq 1} n^{k-1}q^{nm} \\ &= 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!}\sum_{n\geq 1}\sigma_{k-1}(n)q^n \end{aligned}$$

which is what we wanted. $\square$

Corollary 1.9. *$E_k(\tau)$ has q-expansion*

$$E_k(\tau) = 1 + \frac{(2\pi i)^k}{(k-1)!\zeta(k)}\sum_{n\geq 1}\sigma_{k-1}(n)q^n$$

Consequently, the coefficients are all rationals. In particular, they are integers when $k = 4, 6$, in which case

$$E_4(\tau) = 1 + 240\sum_{n\geq 1}\sigma_3(n)q^n, E_6(\tau) = 1 - 504\sum_{n\geq 1}\sigma_5(n)q^n$$

Proof. It is well known that $\pi^k/\zeta(k)$ is rational for any $k \in 2\mathbb{Z}_{>0}$. In particular, $\zeta(4) = \pi^4/90$ and $\zeta(6) = \pi^6/945$. $\square$

Proposition 1.10. *The q-expansions of Δ is $q+\sum_{n\geq 2} a_n q^n$ where $\forall n \geq 2, a_n \in \mathbb{Z}$. The q-expansion of j is $q^{-1}\sum_{n\geq 0} b_n q^n$ where $\forall n \geq 0, b_n \in \mathbb{Z}$.*

Proof. As $j = E_4^3/\Delta$, it suffices to show that Δ has the said properties. Write $E_4 = 1 + 240U, E_6 = 1 - 504V$ where $U = \sum_{n\geq 1}\sigma_3(n)q^n, V = \sum_{n\geq 1}\sigma_5(n)q^n$, then $\Delta = (E_4^3 - E_6^2)/1728 = 5(U-V)/12 + P(U,V)$ where $P(U,V) \in \mathbb{Z}[U,V]$. Consequently, the q-expansion of Δ has integer coefficients since $\sigma_3(n) \equiv \sigma_5(n) \pmod{12}$ for all $n \in \mathbb{N}$. The leading term is q as $(3\times 240+2\times 504)/1728 = 1$. $\square$

Proposition 1.11. *Let $k \geq 0$ be an even integer. Then there exists a basis $f_1, \ldots, f_N$ for $S_k(\mathrm{SL}_2(\mathbb{Z}))$ such that:*
(a) If f_i has q-expansion $\sum_{n\geq 1} a_{i,n}q^n$, then $a_{i,n} \in \mathbb{Z}$.
(b) $a_{i,n} = \delta_{in}$ if $i, n \in \{1, \ldots, N\}$.

Proof. Assume that $S_k(\mathrm{SL}_2(\mathbb{Z})) \neq 0$. Write $k = 12a + d$ where $d = 14$ if $k \equiv 2 \pmod{12}$ and $0 \leq d < 12$ otherwise, then $N = \dim_{\mathbb{C}} S_k(\mathrm{SL}_2(\mathbb{Z})) = a$. Write $d = 4A + 6B$ for some $A, B \in \mathbb{Z}_{\geq 0}$ and introduce $g_i = \Delta^i E_4^A E_6^B E_6^{2(N-i)}$ for $i = 1, \ldots, N$. The weight of g_i is $12i + 4A + 6B + 12(N-i) = 12N + d = k$. Also the leading term in the q-expansion of g_i is q^i and the coefficients in the q-expansions of g_i are integers. Performing row reduction on the first N coefficients in the q-expansions of g_i shall produce the required f_i. $\square$

2 Hecke Operators and L-functions

Hecke operators are just automorphisms of $S_k(\mathrm{SL}_2(\mathbb{Z}))$ and $M_k(\mathrm{SL}_2(\mathbb{Z}))$. They exists, in fact, for rather abstract reasons. So naturally, we want to start from an abstract point of view.

2.1 Abstract Hecke Algebras

Definition 2.1. Let G be a group and $H \leq G$. We say (G, H) is a Hecke pair if for all $g \in G$, the set of orbits $H\backslash(HgH)$ (where H acts on the double coset HgH by left multiplication) is finite.

Lemma 2.1. *Let $H \leq G$. For any $g \in G$, there is a bijection $(H \cap g^{-1}Hg)\backslash H \to H\backslash(HgH)$ where $(H \cap g^{-1}Hg)\backslash H$ is the set of right cosets of $H \cap g^{-1}Hg$ in H. In particular, (G, H) is a Hecke pair iff $\forall g \in G, [H : H \cap g^{-1}Hg] < \infty$*

Proof. Consider the surjective map $H \to H\backslash(HgH)$ via $h \mapsto Hgh$. We have $Hgh_1 = Hgh_2 \iff gh_1h_2^{-1}g^{-1} \in H \iff h_1h_2^{-1} \in H \cap g^{-1}Hg \iff (H \cap g^{-1}Hg)h_1 = (H \cap g^{-1}Hg)h_2$. So we can induce a bijection $(H \cap g^{-1}Hg)\backslash H \to H\backslash(HgH)$ which is what we want. $\square$

Definition 2.2. Suppose (G, H) is a Hecke pair. The Hecke algebra $\mathcal{H}(G, H)$ is defined to be the set of all functions $f : G \to \mathbb{C}$ satisfying the following condition:
(a) $\forall g \in G, h_1, h_2 \in H, f(h_1gh_2) = f(g)$.
(b) f is nonzero on only finitely many H-double cosets.
If $f_1, f_2 \in \mathcal{H}(G, H)$, then we define their product as

$$(f_1 \cdot f_2)(g) = \sum_i f_1(gg_i^{-1})f_2(g_i)$$

where $\{g_i\}$ is a set of representatives for the disjoint decomposition $G = \coprod_i Hg_i$.

Remark. 1. We can, of course, switch the actions to the other side instead.
2. A continuous analogue of this notion can be defined in the context where G is a locally compact topological group. In this case, the product on the Hecke algebra is the convolution that comes with it.

Lemma 2.2. *The multiplication in $\mathcal{H}(G, H)$ is well-defined and associative. Moreover, its unit element is the indicator $[H] = 1_H$.*

Proof. The sum is finite since we can write $f_1(g) = \sum_i \lambda_i[Hx_iH]$ where $\lambda_i \in \mathbb{C}$ and $\{x_i\}$ is a finite set of representatives of the double cosets of H where f_1 is supported. As (G, H) is a Hecke pair, we can write $[Hx_iH] = \sum_j [Hy_{ij}]$ where Hy_{ij} are the (finitely many) orbits in $H\backslash(Hx_iH)$. Hence f_1 can be written as a finite sum $f_1 = \sum_{i,j} \lambda_i[Hy_{ij}]$. Similarly, we can write $f_2 = \sum_{k,l} \mu_k[Hz_{kl}]$, so

$$(f_1 \cdot f_2)(g) = \sum_{i,j,k,l,r} \lambda_j\mu_k[Hy_{ij}](gg_r^{-1})[Hz_{kl}](g_r)$$

which has to be a finite sum since i, j, k, l are indices of finite range.
To see it is independent of the choice of $\{g_i\}$, just observe that for any $\{h_i\} \subset H$,

$$\sum_i f_1(g(h_ig_i)^{-1})f_2(h_ig_i) = \sum_i f_i(gg_i^{-1}h_i^{-1})f_2(h_ig_i) = \sum_i f_i(gg_i^{-1})f_2(g_i)$$

To see $f_1 \cdot f_2$ is in $\mathcal{H}(G,H)$, take any $g \in G, h_1, h_2 \in H$ and

$$\begin{aligned}(f_1 \cdot f_2)(h_1 g h_2) &= \sum_i f_1(h_1 g h_2 g_i^{-1}) f_2(g_i) \\ &= \sum_i f_1(g(g_i h_2^{-1})^{-1}) f_2(g_i h_2^{-1}) \\ &= (f_1 \cdot f_2)(g)\end{aligned}$$

as $\cdot$ is well-defined. It is easy to see $f_1 \cdot f_2$ is supported on finitely many H-double cosets (by e.g. the formula we obtained earlier in terms of indicator functions), so indeed $\cdot : \mathcal{H}(G,H) \times \mathcal{H}(G,H) \to \mathcal{H}(G,H)$ is well-defined.
To see associativity, we simply compute

$$\begin{aligned}(f_1 \cdot (f_2 \cdot f_3))(g) &= \sum_i f_1(g g_i^{-1})(f_2 \cdot f_3)(g_i) = \sum_{i,j} f_1(g g_i^{-1}) f_2(g_i g_j^{-1}) f_3(g_j) \\ ((f_1 \cdot f_2) \cdot f_3)(g) &= \sum_j (f_1 \cdot f_2)(g g_j^{-1}) f_3(g_j) \\ &= \sum_{i,j} f_1(g g_j^{-1}(g_i g_j^{-1})^{-1}) f_2(g_i g_j^{-1}) f_3(g_j) \\ &= (f_1 \cdot (f_2 \cdot f_3))(g)\end{aligned}$$

$f \cdot [H] = f = [H] \cdot f$ is easy to see. □

Definition 2.3. Let G be a group and V be a $\mathbb{C}$-vector space on which G acts on the right by linear maps. For a subgroup $H \leq G$, the subspace of H-invariants is denoted by $V^H = \{v \in V : \forall h \in H, v \cdot h = v\}$.

Proposition 2.3. *Suppose (G,H) is a Hecke pair. Let V be a $\mathbb{C}$-vector space on which G acts on the right by linear maps, then V^H is a right $\mathcal{H}(G,H)$-module under the action ($v \in V^H, f \in \mathcal{H}(G,H)$) $v \cdot f = \sum_i f(g_i)(v \cdot g_i)$ where $\{g_i\}$ is a family of representatives for the right cosets of H in G.*

Proof. To see $v \cdot f$ is well-defined, note that the sum is finite as f is only supported on finitely many H-double cosets (which are in particular unions of finitely many right cosets). As both the action and f are invariant under H, the sum is independent of the choice of representatives.
We next check that $v \cdot f$ is in V^H. Indeed, whenever $h \in H$,

$$(v \cdot f) \cdot h = \sum_i f(g_i)((v \cdot g_i) \cdot h) = \sum_i f(g_i h)(v \cdot (g_i h)) = v \cdot f$$

The last thing to check is that $v \cdot (f_1 \cdot f_2) = (v \cdot f_1) \cdot f_2$. Indeed, we can just calculate

$$\begin{aligned}v \cdot (f_1 \cdot f_2) &= \sum_i (f_1 \cdot f_2)(g_i)(v \cdot g_i) = \sum_{i,j} f_1(g_i g_j^{-1}) f_2(g_j)(v \cdot g_i) \\ (v \cdot f_1) \cdot f_2 &= \sum_j f_2(g_j)((v \cdot f_1) \cdot g_j) \\ &= \sum_{i,j} f_2(g_j) f_1(g_i g_j^{-1})(v \cdot (g_i g_j^{-1} g_j)) \\ &= v \cdot (f_1 \cdot f_2)\end{aligned}$$

as desired. □

These are all quite dry since what we have done is basically just verifying everything is well-defined. Thankfully, as we've just done all these, it's the time for concrete applications.

2.2 Hecke Operators on Modular Functions

Fix $k \in \mathbb{Z}$ and let $V_k = \{f : \mathfrak{h} \to \mathbb{C} \text{ meromorphic}\}$ equipped with the weight k modular action of $\mathrm{GL}_2(\mathbb{Q})^+ = \mathrm{GL}_2(\mathbb{Q}) \cap \mathrm{GL}_2(\mathbb{R})^+$.

Definition 2.4. For $N \in \mathbb{N}$, we define $\Gamma(N) = \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$ where the map is induced by the canonical quotient map $\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$.

So $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$ and each $\Gamma(N)$ is a finite index subgroup of $\Gamma(1)$.

Lemma 2.4. $(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$ *is a Hecke pair.*

Proof. We need to check that $\forall g \in \mathrm{GL}_2(\mathbb{Q})^+, [\Gamma(1) : \Gamma(1) \cap g^{-1}\Gamma(1)g] < \infty$. We will prove this by showing that $\Gamma(1) \cap g^{-1}\Gamma(1)g$ contains $\Gamma(N^2)$ for some N. Fix $g \in \mathrm{GL}_2(\mathbb{Q})^+$. We can find $N \in \mathbb{N}$ such that Ng and Ng^{-1} both have integer entries. Then

$$g\Gamma(N^2)g^{-1} \subset g(1 + N^2 M_2(\mathbb{Z}))g^{-1} = 1 + (Ng)M_2(\mathbb{Z})(Ng^{-1}) \subset M_2(\mathbb{Z})$$

So $g\Gamma(N^2)g^{-1} \le \Gamma(1) \implies \Gamma(N^2) \le \Gamma(1) \cap g^{-1}\Gamma(1)g$. $\square$

Hence $V_k^{\Gamma(1)}$, which is the space of weakly modular functions of weight k and level $\Gamma(1)$, is a $\mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$-module.

Definition 2.5. For $n \in \mathbb{N}$, we write T_n for the endomorphism of $V_k^{\Gamma(1)}$ induced by $[X_n] \in \mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$ where $X_n = \{g \in M_2(\mathbb{Z}) : \det(g) = n\}$.

It is not quite immediate that $[X_n]$ is indeed an element of the Hecke algebra $\mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$. This is justified with the following lemma:

Lemma 2.5. *The set* $\Gamma(1)\backslash X_n$ *(the set of left* $\Gamma(1)$*-orbits in* X_n*) is finite, and a set of representatives is*

$$\left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : a, d \in \mathbb{N}, ad = n, b \in \mathbb{Z}, 0 \le b \le d \right\}$$

Proof. If $\alpha \in X_n$, then $\mathbb{Z}^2\alpha \le \mathbb{Z}^2$ has index $\det \alpha = n$. There is then a map $\Gamma(1)\backslash X_n \to L_n = \{\Lambda \le \mathbb{Z}^2 : [\mathbb{Z}^2 : \Lambda] = n\}$ via $\alpha \mapsto \mathbb{Z}^2\alpha$ (one can check this is well-defined). Our claim is that this map is bijective.
It is surjective since if $\Lambda \le \mathbb{Z}^2$ has index n then we can find $u, v \in \Lambda$ such that $\Lambda = \mathbb{Z}u \oplus \mathbb{Z}v$, so $\Lambda = \mathbb{Z}^2\alpha$ where $\alpha = \left(\begin{smallmatrix} u \\ v \end{smallmatrix}\right)$ where $\det \alpha = n$ after possibly swapping u, v.
To see it is injective, suppose $\alpha, \beta \in X_n$ and $\mathbb{Z}^2\alpha = \mathbb{Z}^2\beta$, then $\mathbb{Z}^2 = \mathbb{Z}^2\beta\alpha^{-1}$. But this means that $\gamma = \beta\alpha^{-1} \in \mathrm{SL}_2(\mathbb{Z})$, so $\beta = \gamma\alpha \implies \Gamma(1)\beta = \Gamma(1)\alpha$.
It remains to find a suitable representative α for each choice of $\Lambda \le \mathbb{Z}^2$ of index n. Let $e_1 = (1, 0), e_2 = (0, 1)$, then $\mathbb{Z}e_2 \cap \Lambda \le \mathbb{Z}e_2$ has finite index, say d. We have a short exact sequence

$$0 \longrightarrow \mathbb{Z}e_2/(\mathbb{Z}e_2 \cap \Lambda) \longrightarrow \mathbb{Z}^2/\Lambda \longrightarrow \mathbb{Z}^2/(\Lambda + \mathbb{Z}e_2) \longrightarrow 0$$

Note that $\mathbb{Z}^2/(\Lambda+\mathbb{Z}e_2) \cong \mathbb{Z}e_1/(\mathbb{Z}e_1 \cap (\Lambda+\mathbb{Z}e_2))$. Let $a = |\mathbb{Z}^2/(\Lambda+\mathbb{Z}e_2)|$, then $ad = n$. Observe that d is the least natural number such that $de_2 \in \Lambda$. Similarly, a is the least natural number such that $ae_1 \in \Lambda + \mathbb{Z}e_2$. Equivalently, it is the least natural number such that there exists some $b \in \mathbb{Z}$ with $ae_1+be_2 \in \Lambda$. Take the unique choice of b in $[0, d)$ gives the required representative $\alpha = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$. $\square$

What do these T_n's do, precisely?

Proposition 2.6. *Let $n \in \mathbb{N}$ and f be a modular function of weight k and level $\Gamma(1)$. Suppose it has q-expansion $f(\tau) = \sum_{m\in\mathbb{Z}} a_m q^m$. Then $T_n f$ is also a modular function with q-expansion $T_n f(\tau) = \sum_{m\in\mathbb{Z}} b_m q^m$ where*

$$b_m = \sum_{a|\gcd(m,n), a\geq 1} a^{k-1} a_{mn/a^2}$$

Proof. By definition,

$$\begin{aligned}
T_n f(\tau) = (f \cdot [X_n])(\tau) &= \sum_{g\in\Gamma(1)\backslash X_n} f|_k[g](\tau) \\
&= \sum_{ad=n, 0\leq b<d} f\left(\frac{a\tau+b}{d}\right) d^{-k} n^{k-1} \\
&= n^{k-1} \sum_{m\in\mathbb{Z}} \sum_{ad=n} d^{-k} a_m \sum_{0\leq b<d} e^{2\pi i m(a\tau+b)/d} \\
&= n^{k-1} \sum_{m\in\mathbb{Z}} \sum_{ad=n} d^{-k} a_m e^{2\pi i a m\tau/d} \sum_{0\leq b<d} e^{2\pi i m b/d}
\end{aligned}$$

Note that $\sum_{0\leq b<d} e^{2\pi i m b/d} = d 1_{d|m}$, so

$$\begin{aligned}
T_n f(\tau) &= n^{k-1} \sum_{m\in\mathbb{Z}} \sum_{ad=n, d|m} d^{1-k} a_m e^{2\pi i a m\tau/d} \\
&= \sum_{ad=n} a^{k-1} \sum_{m\in\mathbb{Z}} a_{dm} e^{2\pi i a m\tau} \\
&= \sum_{m\in\mathbb{Z}} q^m \sum_{a|\gcd(m,n), a\geq 1} a^{k-1} a_{mn/a^2}
\end{aligned}$$

which is the desired q-expansion. The expression also implies that $b_m = 0$ if m is negative and $|m|$ is sufficiently large, therefore $T_n f$ is indeed a modular function (we already know that $T_n f$ is weakly modular by our established theory on Hecke algebra). $\square$

Corollary 2.7. *T_n also preserves $S_k(\mathrm{SL}_2(\mathbb{Z}))$ and $M_k(\mathrm{SL}_2(\mathbb{Z}))$.*

Write $a_m(g)$ to denote the coefficients in the q-expansion of g, i.e. $g = \sum_{m\in\mathbb{Z}} a_m(g) q^m$.

Corollary 2.8. *If f is a modular function, then $a_0(T_n f) = \sigma_{k-1}(n) a_0(f)$ and $a_1(T_n f) = a_n(f)$.*

Recall from our study of abstract Hecke algebra that each $f \in \mathcal{H}(G, H)$ is a linear combination of indicator functions for right cosets of H. We can say somthing more about this.

Lemma 2.9. *Let (G,H) be a Hecke pair and $f_1 = \sum_i \lambda_i[Hx_i], f_2 = \sum_j \mu_j[H\mu_j]$ be elements of $\mathcal{H}(G,H)$. Then $f_1 \cdot f_2 = \sum_{i,j} \lambda_i\mu_j[Hx_iy_j]$.*

Proof. Simple calculation.

$$(f_1 \cdot f_2)(g) = \sum_k f_1(gg_k^{-1})f_2(g_k) = \sum_{i,j,k} \lambda_i\mu_j[Hx_i](gg_k^{-1})[Hy_j](g_k)$$

For each j, there is a unqiue $k(j)$ such that $g_{k(j)} \in Hy_j$. Write $g_{k(j)} = h_jy_j$, then

$$\begin{aligned}(f_1 \cdot f_2)(g) &= \sum_{i,j} \lambda_i\mu_j[Hx_i](gy_j^{-1}h_j^{-1}) = \sum_{i,j} \lambda_i\mu_j[Hx_i](gy_j^{-1}h_j^{-1}) \\ &= \sum_{i,j} \lambda_i\mu_j[Hx_i](gy_j^{-1}) = \sum_{i,j} \lambda_i\mu_j[Hx_iy_j](g)\end{aligned}$$

as desired, ☐

This allows us to study the composition of Hecke operators.

Proposition 2.10. *1. Suppose $n, m \in \mathbb{N}$ are coprime, then $T_n \circ T_m = T_{nm}$. In particular, $T_n \circ T_m = T_m \circ T_n$*
2. Suppose $p, n \in \mathbb{N}$ with p prime, then $T_{p^n} \circ T_p = T_{p^{n+1}} + p^{k-1}T_{p^{n-1}}$.

Proof. Let $X_+ = \{g \in M_2(\mathbb{Z}) : \det g > 0\} = \bigcup_n X_n$ and $L_+ = \{\Lambda \leq \mathbb{Z}^2 : [\mathbb{Z}^2 : \Lambda] < \infty\} = \bigcup_n L_n$. We have shown in the proof of Lemma 2.5 that $\alpha \mapsto \mathbb{Z}^2\alpha$ is a bijection $\mathrm{SL}_2(\mathbb{Z})\backslash X_n \to L_n$. This can then be extended to a bijection $\mathrm{SL}_2(\mathbb{Z})\backslash X_+ \to L_+$.
If $f \in \mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \mathrm{SL}_2(\mathbb{Z}))$ is supported in X_+, then we can identify it as a function $\phi_f : L_+ \to \mathbb{C}$ given by $\phi_f(\mathbb{Z}^2\alpha) = f(\alpha)$ (well-defined since f is, by definition, $\mathrm{SL}_2(\mathbb{Z})$-invariant). For $n, m \in \mathbb{N}$, we can find coset representatives x_i, y_j to write $[X_n] = \sum_i [\mathrm{SL}_2(\mathbb{Z})x_i], [X_m] = \sum_j [\mathrm{SL}_2(\mathbb{Z})y_j]$ which gives $[X_n] \cdot [X_m] = \sum_{i,j}[\mathrm{SL}_2(\mathbb{Z})x_iy_j]$ by the preceding lemma.
For fixed j, we have $\mathbb{Z}^2 \geq \mathbb{Z}^2y_j \geq \mathbb{Z}^2x_iy_j$. For varying i,j, $\mathbb{Z}^2x_iy_j$'s are exactly the subgroups of $\mathbb{Z}^2y_j$ of index n, each appearing exactly $|\{\mathbb{Z}^2 \geq \Lambda^1 \geq \Lambda : [\mathbb{Z}^2 : \Lambda^1] = m, [\Lambda^1 : \Lambda] = n\}|$ times.
If n, m are coprime, then such Λ_1 is unique for each Λ, therefore $[X_n] \cdot [X_m] = [X_{mn}]$, in other words $T_m \circ T_n = T_{nm}$.
Suppose $n = p, m = p^n$, then $f = [X_p] \cdot [X_{p^n}]$ corresponds to $\phi_f : L_{p^{n+1}} \to \mathbb{C}$ given by $\phi_f(\Lambda) = |\{\mathbb{Z}^2 \geq \Lambda^1 \geq \Lambda : [\mathbb{Z}^2 : \Lambda^1] = p^n, [\Lambda^1 : \Lambda] = p\}|$. If $\mathbb{Z}^2/\Lambda$ is cyclic, then $\phi_f(\Lambda) = 1$ since $\mathbb{Z}^2/\Lambda$ has a unique subgroup of order p.
Otherwise, $\mathbb{Z}^2/\Lambda \cong \mathbb{Z}/p^a\mathbb{Z} \oplus \mathbb{Z}/p^b\mathbb{Z}$ for some $a, b \geq 1, a + b = n + 1$. Notably, this is equivalent to $\Lambda \leq p\mathbb{Z}^2$. In this case, $\mathbb{Z}^2/\Lambda$ has $p + 1$ subgroups of order p, so $\phi_f(\Lambda) = p + 1$.
In conclusion $\phi_f(\Lambda) = 1 + p1_{\Lambda \leq p\mathbb{Z}^2}$, so $[X_p] \cdot [X_{p^n}] = [X_{p^{n+1}}] + p[pX_{p^{n-1}}] \implies T_{p^n} \circ T_p = T_{p^{n+1}} + p \cdot p^{2k-2} \cdot p^{-k}T_{p^{n-1}} = T_{p^{n+1}} + p^{k-1}T_{p^{n-1}}$. ☐

Corollary 2.11. *1. For any prime p, T_{p^n} is a polynomial in T_p.*
2. T_n, T_m commute for arbitrary m, n.

Proof. 1 follows from induction. To see 2, just decompose $n = \prod_i p_i^{\alpha_i}$ and observe that $T_n = \prod_i T_{p_i^{\alpha_i}}$. ☐

2.3 Diagonalising the Hecke Operators

Proposition 2.12. *Let $k \geq 4$ be even. Then for all $n \in \mathbb{N}$, $T_n E_k = \sigma_{k-1}(n) E_k$.*

So, in fact, the Eisenstein series are all eigenvectors of the Hecke operators T_n. This allows us to diagonalise T_n in $M_k(\mathrm{SL}_2(\mathbb{Z}))$ by diagonalising T_n in $S_k(\mathrm{SL}_2(\mathbb{Z}))$.

Proof. It suffices to show that E_k is an eigenvector. We shall show that it is an eigenvector of T_p for prime p, which implies the general claim by Corollary 2.11. It is of course easier to work with G_k instead.

$$\begin{aligned}(T_p G_k)(\tau) &= \sum_{ad=p, 0 \leq b < d} G_k|_k \left[\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)\right] \\ &= \sum_{ad=p, 0 \leq b < d} p^{k-1} \sum_{\omega \in \Lambda_{(a\tau+b)/d} \setminus \{0\}} (d\omega)^{-k} \\ &= \sum_{\Lambda' \leq \Lambda_\tau, [\Lambda:\Lambda']=p} p^{k-1} \sum_{\omega \in \Lambda' \setminus \{0\}} \omega^{-k}\end{aligned}$$

For $\omega \in \Lambda_\tau \setminus \{0\}$, how many $\Lambda' \leq \Lambda_\tau$ with index p contains ω? if $\omega \notin p\Lambda_\tau$, then $\Lambda' = \mathbb{Z}\omega + p\Lambda_\tau$ is the unique such subgroup. Otherwise, every such $\Lambda' \leq \Lambda_\tau$ with index p contains ω. Hence

$$\begin{aligned}(T_p G_k)(\tau) &= p^{k-1} \left(\sum_{\omega \in \Lambda_\tau \setminus \{0\}} \omega^{-k} + \sum_{\omega \in p\Lambda_\tau \setminus \{0\}} p\omega^{-k} \right) \\ &= p^{k-1} \left(G_k(\tau) + \sum_{\omega \in \Lambda_\tau \setminus \{0\}} p(p\omega)^{-k} \right) \\ &= (p^{k-1} + 1) G_k(\tau) = \sigma_{k-1}(p) G_k(\tau)\end{aligned}$$

as desired. □

Inevitably, our focus is now on $S_k(\mathrm{SL}_2(\mathbb{Z}))$. Finding explicit eigenvectors does seem hard, so our strategy would be to harness the power of spectral theory by finding an inner product on $S_k(\mathrm{SL}_2(\mathbb{Z}))$ such that the Hecke operators are self-adjoint.

Lemma 2.13. *1. $y^{-2}\,\mathrm{d}x\,\mathrm{d}y$ (where $\tau = x + iy$) is a $\mathrm{GL}_2(\mathbb{R})^+$-invariant area form.*
2. If $f, g : \mathfrak{h} \to \mathbb{C}$ are smooth and invariant under the weight k modular action of a subgroup $G \leq \mathrm{GL}_2(\mathbb{R})^+$, then $\omega(f,g) = f(\tau)\overline{g(\tau)} y^k\,\mathrm{d}x\,\mathrm{d}y/y^2$ satisfies $\alpha^*\omega(f,g) = (\det \alpha)^{2-k}\omega(f,g)$ for all $\alpha \in G$. In particular, if $G \leq \mathrm{SL}_2(\mathbb{R}) \cap \mathrm{GL}_2(\mathbb{R})^+$, then ω is G-invariant.

Proof. 1. Note that $\mathrm{d}\tau\,\mathrm{d}\bar{\tau} = -2i\,\mathrm{d}x\,\mathrm{d}y$. For $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{GL}_2(\mathbb{R})^+$,

$$g^*(\mathrm{d}\tau) = \mathrm{d}\left(\frac{a\tau + b}{c\tau + d}\right) = \frac{\det g}{j(g,\tau)^2}\,\mathrm{d}\tau$$

So

$$g^*\left(\frac{\mathrm{d}\tau\,\mathrm{d}\bar{\tau}}{y^2}\right) = \frac{(\det g)^2\,\mathrm{d}\tau\,\mathrm{d}\bar{\tau}}{|j(g,\tau)|^4}\left(\frac{y \det g}{|j(g,\tau)|^2}\right)^{-2} = \frac{\mathrm{d}\tau\,\mathrm{d}\bar{\tau}}{y^2}$$

2. Take $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$, then

$$\begin{aligned}
\alpha^*\omega(f,g) &= f(\alpha\tau)\overline{g(\alpha\tau)}\,\mathrm{Im}(\alpha\tau)^k \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= f(\tau)j(\alpha,\tau)^k(\det\alpha)^{1-k}\overline{g(\tau)j(\alpha,\tau)^k}(\det\alpha)^{1-k}\frac{y^k(\det\alpha)^k}{|j(\alpha,\tau)|^{2k}}\frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= f(\tau)\overline{g(\tau)}y^k\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}(\det\alpha)^{2-k} \\
&= (\det\alpha)^{2-k}\omega(f,g)
\end{aligned}$$

which is what we wanted. □

So if f, g are invariant under the weight k action of $\Gamma(1)$, we have a $\Gamma(1)$-invariant area form $\omega(f,g)$.

Definition 2.6. For a $\Gamma(1)$-invariant function $\phi : \mathfrak{h} \to \mathbb{C}$, we define

$$\int_{\Gamma(1)\backslash\mathfrak{h}} \phi\frac{\mathrm{d}x\,\mathrm{d}y}{y^2} = \int_{\mathcal{F}} \phi\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$$

when this integral is absolutely convergent.

Definition 2.7. Let $f, g \in S_k(\Gamma(1))$, we define

$$\langle f, g\rangle = \int_{\Gamma(1)\backslash\mathfrak{h}} \omega(f,g)$$

We certainly want this to be a well-defined inner product.

Lemma 2.14. *For any $f, g \in S_k(\Gamma(1))$, the integral defining $\langle f, g\rangle$ is absolutely convergent.*

Proof. It is easy to see that

$$\int_{\Gamma(1)\backslash\mathfrak{h}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} = \int_{\mathcal{F}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \leq \int_{x=-1/2}^{1/2}\int_{y=1/2}^{\infty} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} < \infty$$

Also, for any $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$, the expression $|f(\tau)|y^{k/2}$ is bounded above by $\sum_{n\geq 1}|a_n(f)|e^{-2\pi ny}y^{k/2}$, which tends to 0 uniformly as $y \to \infty$. In particular, $|f(\tau)|y^{k/2}$ is bounded for $\tau = x+iy \in \mathcal{F}$. Combining both gives the lemma. □

It is then easy to see that it does satisfy what we want from an inner product.

Definition 2.8. Suppose $\Gamma \leq \Gamma(1)$ has finite index. Let $\phi : \mathfrak{h} \to \mathbb{C}$ be a smooth function invariant under translation by Γ. Then we define

$$\int_{\Gamma\backslash\mathfrak{h}} \phi\frac{\mathrm{d}x\,\mathrm{d}y}{y^2} = \int_{\Gamma(1)\backslash\mathfrak{h}} \sum_{\gamma\in\Gamma\backslash\Gamma(1)} (\phi\circ\gamma)\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$$

where $\Gamma\backslash\Gamma(1)$ denotes the set of right cosets of Γ in $\Gamma(1)$.

Definition 2.9. Suppose $\Gamma \leq \Gamma(1)$ has finite index and $f, g : \mathfrak{h} \to \mathbb{C}$ are invariant under the weight k modular action of Γ. The Petersson inner product of f and g is defined by

$$\langle f, g\rangle = [\Gamma(1) : \Gamma]^{-1} \int_{\Gamma\backslash\mathfrak{h}} \omega(f, g)$$

whenever the integral converges absolutely.

Remark. 1. If $\Gamma' \leq \Gamma \leq \Gamma(1)$ and everything has finite index, then $\langle f, g\rangle$ defined using Γ is the same as $\langle f, g\rangle$ defined using Γ'. So we don't have to specify the subgroup Γ.
2. If $f : \mathfrak{h} \to \mathbb{C}$ is invariant under the weight k action of $\Gamma(1)$ and $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$, then $f|_k[\alpha]$ is invariant under the weight k modular action of $\alpha^{-1}\Gamma(1)\alpha \cap \Gamma(1)$ which has finite index in $\Gamma(1)$ (since $(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$ is a Hecke pair).

Proposition 2.15. *Let $f, g : \mathfrak{h} \to \mathbb{C}$ be smooth and invariant under the weight k modular action of $\Gamma(1)$, then we have $\langle f|_k[\alpha], g|_k[\alpha]\rangle = (\det \alpha)^{k-2}\langle f, g\rangle$ provided that both sides of this equality are defined.*

Proof. Computation gives

$$\begin{aligned}
&\langle f|_k[\alpha], g|_k[\alpha]\rangle[\Gamma(1) : \alpha^{-1}\Gamma(1)\alpha \cap \Gamma(1)] \\
&= \int_{(\alpha^{-1}\Gamma(1)\alpha\cap\Gamma(1))\backslash\mathfrak{h}} f(\alpha\tau)\overline{g(\alpha\tau)}|j(\alpha, \tau)|^{-2k}(\det \alpha)^{2k-2}y^k \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= \int_{(\alpha\Gamma(1)\alpha^{-1}\cap\Gamma(1))\backslash\mathfrak{h}} f(\tau)\overline{g(\tau)}|j(\alpha, \alpha^{-1}\tau)|^{-2k}(\det \alpha)^{2k-2}\,\mathrm{Im}(\alpha^{-1}\tau)^k \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= \int_{(\alpha\Gamma(1)\alpha^{-1}\cap\Gamma(1))\backslash\mathfrak{h}} f(\tau)\overline{g(\tau)}|j(\alpha, \alpha^{-1}\tau)|^{-2k}(\det \alpha)^{2k-2} \frac{(\det \alpha)^{-k}\,\mathrm{Im}(\tau)^k}{|j(\alpha^{-1}, \tau)|^{2k}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= \int_{(\alpha\Gamma(1)\alpha^{-1}\cap\Gamma(1))\backslash\mathfrak{h}} f(\tau)\overline{g(\tau)}(\det \alpha)^{k-2}\,\mathrm{Im}(\tau)^k \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= (\det \alpha)^{k-2}\langle f, g\rangle[\Gamma(1) : \alpha\Gamma(1)\alpha^{-1} \cap \Gamma(1)]
\end{aligned}$$

So it remains to check that $[\Gamma(1) : \alpha^{-1}\Gamma(1)\alpha \cap \Gamma(1)] = [\Gamma(1) : \alpha\Gamma(1)\alpha^{-1} \cap \Gamma(1)]$ which follows from

$$\begin{aligned}
\int_{\Gamma(1)\backslash\mathfrak{h}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} &= \frac{1}{[\Gamma(1) : \Gamma(1) \cap \alpha^{-1}\Gamma(1)\alpha]} \int_{(\Gamma(1)\cap\alpha^{-1}\Gamma(1)\alpha)\backslash\mathfrak{h}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= \frac{1}{[\Gamma(1) : \Gamma(1) \cap \alpha^{-1}\Gamma(1)\alpha]} \int_{(\alpha\Gamma(1)\alpha^{-1}\cap\Gamma(1))\backslash\mathfrak{h}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\
&= \frac{[\Gamma(1) : \alpha\Gamma(1)\alpha^{-1} \cap \Gamma(1)]}{[\Gamma(1) : \Gamma(1) \cap \alpha^{-1}\Gamma(1)\alpha]} \int_{\Gamma(1)\backslash\mathfrak{h}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}
\end{aligned}$$

This completes the proof. □

Proposition 2.16. *For any $n \in \mathbb{N}$, the Hecke operator T_n is self-adjoint on $S_k(\mathrm{SL}_2(\mathbb{Z}))$ with respect to $\langle\cdot, \cdot\rangle$.*

Proof. We need to show $\langle T_n f, g\rangle = \langle f, T_n g\rangle$. It suffices to show the case where $n = p$ is prime.
We first claim that we can find matrices $x_i \in X_p$ such that $X_p = \coprod_i \Gamma(1)x_i = \coprod_i \Gamma(1)px_i^{-1}$. To see this, observe that $\mathrm{Adj} : X_p \to X_p$ is a bijection. Choose a_i, b_i such that $X_p = \coprod_i \Gamma(1)a_i = \coprod_i b_i\Gamma(1)$. Observe that X_p consists of a single $\Gamma(1)$-double coset, i.e. $\Gamma(1)\backslash X_p/\Gamma(1)$ has a single element, because every matrix in X_p can be written in Smith normal form. So we can find elements $u_i, v_i \in \Gamma(1)$ such that $u_i a_i = b_i v_i = x_i$ which concludes the claim.
Then, we can compute

$$\begin{aligned}
\langle T_p f, g\rangle &= \left\langle \sum_i f|_k[x_i], g \right\rangle = \sum_i \langle f|_k[x_i], g\rangle \\
&= \sum_i \langle f|_k[x_i]|_k[x_i^{-1}], g|_k[x_i^{-1}]\rangle \det(x_i^{-1})^{2-k} \\
&= \sum_i \langle f, g|_k[x_i^{-1}]\rangle p^{k-2} = \sum_i \langle f, g|_k[px_i^{-1}]\rangle \\
&= \left\langle f, \sum_i g|_k[px_i^{-1}] \right\rangle \\
&= \langle f, T_p g\rangle
\end{aligned}$$

as desired. □

Corollary 2.17. *The operators T_n on $S_k(\mathrm{SL}_2(\mathbb{Z}))$ can be simultaneously diagonalised with real eigenvalues.*

Proof. General spectral theory on self-adjoint operators. □

Definition 2.10. We say an element $f \in M_k(\Gamma(1))$ is a normalised eigenform if f is an eigenvector for all Hecke operators T_n and $a_1(f) = 1$.

Lemma 2.18. *If $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ is an eigenvector for all T_n and $k > 0$, then there is a unique scalar multiple of f which is a normalised eigenform.*

Proof. If $a_1(f) = 0$ and $T_n f = \alpha_n f$, then $\forall n, a_n(f) = a_1(T_n f) = \alpha_n a_1(f) = 0 \implies f = 0$, contradiction. So $a_1(f) = 0$ from where the lemma is obvious. □

From the idea in the proof, we also know that if f is a normalised eigenform then $a_n(f)$ is the eigenvalue of f wrt T_n. Consequently, the simultaneous eigenspaces all have dimension 1.

Example 2.1. We have seen that

$$G_k(\tau) = 2\zeta(k) + \frac{2(2\pi i)^k}{(k-1)!} \sum_{n\geq 1} \sigma_{k-1}(n)q^n$$

is an eigenvector for all Hecke operators T_n. The corresponding normalised eigenform is

$$\frac{(k-1)!\zeta(k)}{(2\pi i)^k} + \sum_{n\geq 1} \sigma_{k-1}(n)q^n$$

Proposition 2.19. *The eigenvalues of T_n on $M_k(\mathrm{SL}_2(\mathbb{Z}))$ are all (real) algebraic integers which lie in a number field independent of n.*

Proof. Let $f_1, \ldots, f_N$ be the basis for $S_k(\mathrm{SL}_2(\mathbb{Z}))$ as described in Proposition 1.11. Consequently, for any $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$, we have $f = \sum_{i=1}^N a_i(f) f_i$. The matrix of T_n with respect to this basis then has integer entries, so its eigenvalues are algebraic integers.
The eigenvalues lie in the number field generated by $a_1(f), \ldots, a_N(f)$ where f is a normalised eigenform. $\square$

2.4 Ramanujan's Conjectures

For any weight k, we can associate with it a sequence $(a_1(f), a_2(f), \ldots)$ of Hecke eigenvalues (where f is a normalised eigenform). Since the operators T_n are polynomials in T_p for p prime, this sequence is determined by $(a_p(f_i))$ where p is prime. When $S_k(\mathrm{SL}_2(\mathbb{Z}))$ is one-dimensional (when $k = 12, 16, 18, 20, 22, 26$), then the Hecke eigenvalues are integers. When $k = 12$, Δ is a normalised eigenform which was first considered by Ramanujan (1916) who defined it as

$$\Delta = q \prod_{n \geq 1} (1 - q^n)^{24} = \sum_{n \geq 1} \tau(n) q^n$$

He then conjectured:

Proposition 2.20. *1. $\tau(mn) = \tau(m)\tau(n)$ for m, n coprime.*
2. If p is prime and $n \in \mathbb{N}$, then $\tau(p)\tau(p^n) = \tau(p^{n+1}) + p^{11}\tau(p^{n-1})$.

which follows from the general properties of Hecke operators we have established!
What about when $\dim S_k(\mathrm{SL}_2(\mathbb{Z})) > 1$? The first example occurs when $k = 24$. We know we can find a basis g_1, g_2 for $S_{24}(\mathrm{SL}_2(\mathbb{Z}))$ such that $g_1 = q_1 + O(q^2), g_2 = q^2 + O(q^3)$. Let's compute the matrix of T_n in this basis. We know

$$T_2 \left(\sum_{n \geq 1} a_n q^n \right) = \sum_{n \geq 1} \sum_{a | \gcd(2,n)} a^{k-1} a_{2n/a^2} q^n$$

so $a_1(T_2 f) = a_2(f), a_2(T_2 f) = a_4(f) + 2^{23} a_1(f)$, or $T_2 f = a_2(f) g_1 + (a_4(f) + 2^{23} a_1(f)) g_2$. With some messy algebra (that can be outsourced to stuff like Wolfram Mathematica), we obtain the matrix

$$\begin{pmatrix} 0 & 1 \\ 20468736 & 1080 \end{pmatrix}$$

for T_2. So T_2 has eigenvalues $12(45 \pm \sqrt{144169})$ and all the eigenvalues of T_2 are in $\mathbb{Q}(\sqrt{144169})$. Nicely 144169 is prime.
There's another related conjecture from Ramanujan regarding $\tau(n)$. We first need some motivation.

Lemma 2.21. *Let p be a prime number, then we have an identity of formal power series*

$$1 + \sum_{n \geq 1} \tau(p^n) X^n = (1 - \tau(p) X + p^{11} X^2)^{-1}$$

Proof. Follows directly from preceding proposition. □

Let α_p, β_p be the roots of the polynomial $1 - \tau(p)X + p^{11}X^2$. If $-\tau(p)^2 - 4p^{11} \leq 0$, then α_p, β_p are complex conjugates with absolute value $p^{11/2}$; Otherwise, they are distinct real numbers. Regarding this, Ramanujan conjectured that only the first case may happen. Or, more generally,

Proposition 2.22 (Ramanujan-Petersson Conjecture)**.** *Let f be a cuspidal normalised eigenform of weight k and level $\mathrm{SL}_2(\mathbb{Z})$, then for any prime p, we have $|a_p(f)| \leq 2p^{(k-1)/2}$*

Why is this important? It can actually provide links between modular forms and number theory in various ways. Later, we will prove that

$$r_{24}(p) = \frac{16}{691}\sigma_{11}(p) + \frac{33152}{691}\tau(p)$$

for any odd prime p, where $r_{24}(n)$ is the number of ways to write $n = x_1^2 + \cdots + x_{24}^2$ for $x_1, \ldots, x_{24} \in \mathbb{Z}$. The Ramanujan-Petersson conjecture then gives the estimate $r_{24}(p) = (16/691)p^{11} + O(p^{11/2})$.
Deligne proved this conjecture in 1973 as part of his proof of the Weil conjectures. As part of this, he proved the existence of Galois representations attached to cuspidal normalised eigenforms. This is now seen as part of the Langlands programme. Specifically, Deligne proved that

Theorem 2.23 (Deligne)**.** *Let $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ be a normalised eigenform and $K_f/\mathbb{Q}$ be the number field generated by $\{a_p(f) : p \text{ prime}\}$ (in other words, the number field generated by the eigenvalues of the Hecke operators). Suppose $\mathfrak{p} \subset \mathcal{O}_{K_f}$ is a prime ideal lying above a prime number p, then there exists a continuous homomorphism $\rho_{f,\mathfrak{p}} : \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(K_{f,\mathfrak{p}})$ (where $K_{f,\mathfrak{p}}$ is the completion of K_f at $\mathfrak{p}$) such that:*
1. $\rho_{f,\mathfrak{p}}$ is unramified at all prime $\ell \neq p$.
2. For all primes $\ell \neq p$, $\det(X - \rho_{f,\mathfrak{p}}(\mathrm{Frob}_\ell)) = X^2 - a_\ell(f)X + \ell^{k-1}$.

So each cuspidal normalised eigenform gives rise to a family of Galois representations in the way stated in the theorem. There are many applications of this, for example,

Theorem 2.24 (Herbrand-Ribet)**.** *Let p be an odd prime and $4 \leq k \leq p-3$ be an integer, then the followings are equivalent:*
(i) p divides the numerator of $\zeta(1-k)$.
(ii) There exists a non-zero element $x \in \mathrm{Cl}(\mathbb{Z}[\exp(2\pi i/p)])$ annihilated by p such that for all elements $\sigma = \sigma_a \in \mathrm{Gal}(\mathbb{Q}(\exp(2\pi i/p))/\mathbb{Q}) \cong (\mathbb{Z}/p\mathbb{Z})^\times$ (where the isomorphism sends σ_a to a), we have $\sigma_a(x) = a^{1-k}x$.

Sketch of proof that (i) implies (ii). The normalised eigenform associated to G_k can be written as

$$(-1)^k\zeta(1-k) + \sum_{n\geq 1}\sigma_{k-1}(n)q^n$$

When $p \mid \zeta(1-k)$, we can reduce this to get a "mod p cuspidal modular form" which can be used to show the result. □

2.5 L-Functions of Modular Forms

A practical importance of Hecke operators is that we can construct L-functions from them. Recall that the Riemann zeta function $\zeta(s) = \sum_{n\geq 1} n^{-s}$ has very good properties: Firstly, we can write it as its Euler product (when $\operatorname{Re} s > 1$)

$$\zeta(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}$$

Secondly, we can continue it meromorphically to the whole complex plane with a simple pole at $s = 1$. Thirdly, it satisfies a functional equation $\xi(s) = \xi(1-s)$ where $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ is the completed ζ function. One also might hypothesize that the values ζ takes at integers should relate to arithmetics, but it is in general hard to prove results of this form.
Conventionally, we call a Dirichlet series $\sum_{n\geq 1} a_n n^{-s}$ an L-function if analogues of these three conditions hold. Examples of this includes the Dirichlet L-functions

$$L(\chi, s) = \sum_{n\geq 1, \gcd(n,N)=1} \chi(n \bmod N) n^{-s}$$

where $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ is a homomorphism. Through Hecke operators, we can construct another big class of L-functions from modular forms.

Definition 2.11. Let $f \in M_k(\mathrm{SL}_2(\mathbb{Z}))$ be a nonzero modular form and let its q-expansion be $f(\tau) = \sum_{n\geq 0} a_n q^n$. The associated L-function is defined as $L(f, s) = \sum_{n\geq 1} a_n n^{-s}$.

Example 2.2. Take f to be the normalised eigenform associated with G_k, then

$$\begin{aligned} L(f,s) &= \sum_{n\geq 1} \sigma_{k-1}(n) n^{-s} = \sum_{n\geq 1, d|n} d^{k-1} n^{-s} = \sum_{n,m\geq 1} m^{k-1}(mn)^{-s} \\ &= \zeta(s)\zeta(s+1-k) \end{aligned}$$

In particular, there is a functional equation relating $L(f, s)$ and $L(f, k-s)$ in this case.

Lemma 2.25. *Suppose $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ is nonzero, then $L(f, s)$ converges absolutely for $\operatorname{Re} s > 1 + k/2$.*

Proof. Followed form the fact that $|a_n(f)| = O(n^{k/2})$ (exercise). □

Theorem 2.26. *Let $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ be nonzero, then:*
1. $L(f, s)$ has an meromorphic continuation to $\mathbb{C}$.
2. Let $\Lambda(f, s) = (2\pi)^{-s}\Gamma(s)L(f, s)$ be the completed L-function, then $\Lambda(f, s) = i^k\Lambda(f, k-s)$.

Before we prove this, let's remind ourselves of some facts about the Γ function. By definition,

$$\Gamma(s) = \int_0^\infty e^{-y} y^s \frac{\mathrm{d}y}{y}$$

which converges absolutely and uniformly on compact sets of the right half-plane $\{\operatorname{Re} s > 0\}$. Via integration by parts, we also know that $\Gamma(s+1) = s\Gamma(s)$ which allows us to meromophically continue Γ to $\mathbb{C}$.

Proof. Define

$$F(s) = \int_0^\infty f(iy)y^s \frac{\mathrm{d}y}{y}$$

as the Mellin transform of $f(iy)$. The integral converges absolutely in $\mathbb{C}$ and hence F defines an entire function. This follows from the fact that $|f(iy)| = O(e^{-2\pi y})$ as $y \to \infty$ and that $f(-1/\tau)\tau^{-k} = f(\tau) \implies f(iy) = f(i/y)(iy)^{-k}$ so $|f(i/y)| = |f(iy)(iy)^k| = O(y^k e^{-2\pi y}) = O(e^{-\pi y})$ as $y \to \infty$.
Formally, we can compute

$$\begin{aligned} F(s) &= \int_0^\infty \sum_{n\geq 1} a_n e^{-2\pi n y} y^s \frac{\mathrm{d}y}{y} = \sum_{n\geq 1} a_n \int_0^\infty e^{2\pi n y} y^s \frac{\mathrm{d}y}{y} \\ &= \sum_{n\geq 1} a_n n^{-s}\Gamma(s)(2\pi)^{-s} = (2\pi)^{-s}\Gamma(s)L(f,s) = \Lambda(f,s) \end{aligned}$$

Does it actually work? Of course. This is because the absolute convergence of $L(f,s)$ in the region $\operatorname{Re} s > 1 + k/2$ implies the computation is valid in the same region.
Consequently, Λ can be meromorphically continued to $\mathbb{C}$. As we have $L(f,s) = (2\pi)^s\Lambda(f,s)/\Gamma(s)$ and $\Gamma \neq 0$ on $\mathbb{C}$, this also gives a meromorphic continuation of $L(f,s)$ to $\mathbb{C}$.
To get the functional equation, we compute

$$\begin{aligned} \Lambda(f,s) &= \int_0^1 f(iy)y^s \frac{\mathrm{d}y}{y} + \int_1^\infty f(iy)y^s \frac{\mathrm{d}y}{y} \\ &= \int_0^1 f(i/y)(iy)^{-k} y^s \frac{\mathrm{d}y}{y} + \int_1^\infty f(iy)y^s \frac{\mathrm{d}y}{y} \\ &= \int_1^\infty f(iy) i^k y^{k-s} \frac{\mathrm{d}y}{y} + \int_1^\infty f(iy)y^s \frac{\mathrm{d}y}{y} \end{aligned}$$

which implies the $\Lambda(f,s) = \Lambda(f,k-s)$. □

Theorem 2.27. *Let $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ be a normalised eigenform, then the Dirichlet series $L(f,s)$ admits the Euler product*

$$L(f,s) = \prod_{p \text{ prime}} (1 - a_p(f)p^{-s} + p^{k-1-2s})^{-1}$$

Proof. We know $a_{nm}(f) = a_n(f)a_m(f)$ when $\gcd(n,m) = 1$. Hence

$$\sum_{n\geq 1} a_n(f)n^{-s} = \prod_{p \text{ prime}} \left(\sum_{k\geq 1} a_{p^k}(f)p^{-ks} \right)$$

So we only need to show that

$$(1 - a_p(f)p^{-s} + p^{k-1-2s})^{-1} = \sum_{k\geq 1} a_{p^k}(f)p^{-ks}$$

which indeed follows from the relation $T_p \cdot T_{p^k} = T_{p^{k+1}} + p^{k-1}T_{p^{k-1}}$ for $k \in \mathbb{N}$. □

So the class of Dirichlet series $L(f,s)$ does deserve to be called L-functions, at least when f is a cuspidal normalised eigenform. For each prime p, we factor $1 - a_pX + p^{k-1}X^2 = (1-\alpha_pX)(1-\beta_pX)$.

Definition 2.12. For each $m \geq 1$, we define the associated symmetric power L-function as

$$L(\mathrm{Sym}^m, f, s) = \prod_{p \text{ prime}} \prod_{i=0}^{m} (1 - \alpha_p^i \beta_p^{m-i} p^{-s})^{-1}$$

Lemma 2.28. *$L(\mathrm{Sym}^m, f, s)$ converges in the right half-plane.*

Proof. Again $|a_p(f)| = O(p^{k-2})$ which implies that $|\alpha_p|, |\beta_p| = O(p^{k-2})$, therefore $|\alpha_p^i\beta_p^{m-i}| = O(p^{mk/2})$. ☐

These L-functions were first constructed in the 1960s by Langlands and Serre. The motivating idea was to associate to each prime p the (conjugacy class of) matrix $\left(\begin{smallmatrix} \alpha_p & 0 \\ 0 & \beta_p \end{smallmatrix}\right)$ and to associate to any algebraic representation $R : \mathrm{GL}_2(\mathbb{C}) \to \mathrm{GL}_N(\mathbb{C})$ the L-function

$$L(R, f, s) = \prod_{p \text{ prime}} \det(1 - R(t_p)p^{-s})^{-1}$$

We of course want to take R to be the irreducible (algebraic) representations of $\mathrm{GL}_2(\mathbb{C})$, which are those symmetric powers $\mathrm{Sym}^m : \mathrm{GL}_2(\mathbb{C}) \to \mathrm{GL}_{m+1}(\mathbb{C})$ of the identity representation (up to twist).
Naturally, Langlands and Serre then conjectured

Proposition 2.29 (Conjecture)**.** *For any $m \geq 1$, $L(\mathrm{Sym}^m, f, s)$ has an analytic continuation and satisfies a functional equation*

$$\Lambda(\mathrm{Sym}^m, f, s) = \Lambda(\mathrm{Sym}^m, f, m(k-1)+1-s)$$

where Λ is some explicit completed L-function.

Proposition 2.30 (Langlands)**.** *If $L(\mathrm{Sym}^m, f, s)$ has an analytic continuation to $\mathbb{C}$ for all $m \geq 1$, then the Ramanujan-Petersson conjecture holds for f.*

The idea behind Langlands' proof of this proposition made another appearance in Deligne's proof of the Weil conjectures in the early 1970s. Assuming the Ramanujan-Petersson conjecture, we can make a further conjecture about asymptotic behaviour of these Hecke eigenvalues.

Proposition 2.31 (Sato-Tate Conjecture)**.** *Suppose $f \in S_k(\mathrm{SL}_2(\mathbb{Z}))$ is a normalised eigenform, then the values $a_p(f)/(2p^{(k-1)/2})$ are equidistributed in the interval $[-1,1]$ as $p \to \infty$ with respect to the Sato-Tate measure $(2/\pi)\sqrt{1-t^2}\,\mathrm{d}t$. More precisely, for any $g \in C([-1,1])$, we have*

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \sum_{p \text{ prime}, p<x} g\left(\frac{a_p(f)}{2p^{(k-1)/2}}\right) \to \frac{2}{\pi}\int_{-1}^{1} g(t)\sqrt{1-t^2}\,\mathrm{d}t$$

Proposition 2.32 (Serre)**.** *Suppose for all $m \geq 1$, $L(\mathrm{Sym}^m, f, s)$ has an analytic continuation to $\mathbb{C}$ and is non-vanishing on the line $\mathrm{Re}(s) = 1 + m(k-1)/2$, then the Sato-Tate conjecture holds for f.*

So these symmetric power L-functions are indeed interesting.
Recall that for an odd prime p we stated

$$r_{24}(p) = \frac{16}{691}(1 + p^{11}) + \frac{33152}{691}\tau(p)$$

The Sato-Tate conjecture implies that the normalised error terms

$$\frac{691}{33152 \times 2 \times p^{11/2}} \left(r_{24}(p) - \frac{16}{691}(1 + p^{11}) \right) \in [-1, 1]$$

are equidistributed with respect to the Sato-Tate measure.
The Sato-Tate conjecture was proved in 2010 by Banet-Lamb, Geraghty, Harris, and Taylor. The complete analytic continuation of the symmetric power L-function was established in 2019 by J. Newton and J. Thorne (the lecturer).

3 Modular Forms on Congruence Subgroups

3.1 Congruence Subgroups and their Cusps

We now turn our focus to modular forms of level Γ where Γ is any congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. Recall that for $N \in \mathbb{N}$, we defined $\Gamma(N) = \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$.

Definition 3.1. A congruence subgroup $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ is any subgroup of $\mathrm{SL}_2(\mathbb{Z})$ containing $\Gamma(N)$ for some $N \in \mathbb{N}$.

Example 3.1. $\Gamma(N)$ are, of course, congruence subgroups themselves. Also,

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

and

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

are congruence subgroups as well with $\Gamma_0(N) \supset \Gamma_1(N)$. Observe that the homomorphism $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ is surjective, so $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times$ via the map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto d$. Another example of an important congruence subgroup which you met in example sheet is $\Gamma = \langle S, \Gamma(2) \rangle$.

Definition 3.2. A weakly modular function f of weight $k \in \mathbb{Z}$ and level Γ (where Γ is a congruence subgroup) is a meromorphic function on $\mathfrak{h}$ such that $\forall \gamma \in \Gamma, f|_k[\gamma] = f$.

Recall that the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathfrak{h}$ has a fundamental domain $\mathcal{F}$ which is the hyperbolic triangle with vertices ρ, ρ^2, ∞. In example sheet, you have shown that a fundamental domain of $\langle S, \Gamma(2) \rangle$ is the hyperbolic triangle with vertices $-1, 0, \infty$. Naturally, we want some sort of quantity that captures certain features of the fundamental domain.

Definition 3.3. A cusp of a congruence subgroup Γ is a Γ-orbit on $\mathbb{P}^1_{\mathbb{Q}} \subset \mathbb{P}^1_{\mathbb{C}} = \mathbb{C}_\infty$.

Note that $\mathrm{GL}_2(\mathbb{Q})^+$ acts on $\mathbb{P}^1_{\mathbb{Q}}$ by Möbius transformations.

Lemma 3.1. $\mathrm{SL}_2(\mathbb{Z})$ *acts transitively on* $\mathbb{P}^1_{\mathbb{Q}}$.

In particular, $\mathrm{SL}_2(\mathbb{Z})$ has a unique cusp.

Proof. For any $a, c \in \mathbb{Z}$ coprime, choose $b, d \in \mathbb{Z}$ such that $ad - bc = 1$, then we have $\gamma \cdot (0) = a/c$ where $\gamma = \left(\begin{smallmatrix} -b & a \\ -d & c \end{smallmatrix}\right)$. ☐

Corollary 3.2. *If* $\Gamma \leq \mathrm{SL}_2(\mathbb{Z})$ *is a congruence subgroup, then* Γ *has only finitely many cusps.*

Proof. Let

$$\Gamma_\infty = \mathrm{Stab}_{\Gamma(1)}(\infty) = \left\{ \pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{Z} \right\}$$

There is a bijection $\Gamma(1)/\Gamma_\infty \to \mathbb{P}^1_{\mathbb{Q}}$ sending $\gamma\Gamma_\infty \mapsto \gamma \cdot (\infty)$. The set of cusps of Γ can then be identified with the double quotient $\Gamma\backslash\Gamma(1)/\Gamma_\infty$ which is finite. ☐

Example 3.2. Let's compute the cusps of $\Gamma = \langle S, \Gamma(2) \rangle$. Note that we have $\Gamma\backslash\Gamma(1)/\Gamma_\infty = \langle S \rangle \backslash \mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})/\Gamma_\infty$. Now $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})/\Gamma_\infty = \mathbb{F}_2^2 - \{0\}$ as

$$\mathrm{Im}(\Gamma_\infty \to \mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})) = \mathrm{Stab}_{\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

So indeed

$$\Gamma\backslash\Gamma(1)/\Gamma_\infty = \langle S \rangle \Big\backslash \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} = \left\{ \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

More precisely, $\Gamma\backslash\Gamma(1)/\Gamma_\infty$ has representatives $1, \left(\begin{smallmatrix} -1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ in $\Gamma\backslash\Gamma(1)/\Gamma_\infty$. So Γ has two cusps corresponding to $\Gamma \cdot (-1)$ and $\Gamma \cdot (\infty)$.

Remark. We will see later that if Γ is a congruence subgroup, then the set $(\Gamma\backslash\mathfrak{h}) \sqcup (\Gamma\backslash\mathbb{P}^1_{\mathbb{Q}})$ can be given the natural structure of a compact Riemann surface. From this point of view, we can say that the set of cusps compactifies the Riemann surface $\Gamma\backslash\mathfrak{h}$ (with its complex structure inherited from the quotient).

Definition 3.4. If Γ is a congruence subgroup, then the width of the cusp ∞ is

$$\min \left\{ h \in \mathbb{N} : \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \in \Gamma \right\} = [\Gamma_\infty : \Gamma_\infty \cap \{\pm 1\}\Gamma]$$

Let h be the width of the cusp ∞, we define $q_h = e^{2\pi i \tau/h}$. So if f is a weakly modular function of weight k and level Γ, then by modularity $f(\tau + h) = f(\tau)$, so $f(\tau) = \tilde{f}(q_h)$ for some meromorphic $\tilde{f} : \{0 < |q_h| < 1\} \to \mathbb{C}_\infty$.

Definition 3.5. We say f is meromorphic at ∞ if $\tilde{f}$ is meromorphic at 0. It is holomorphic at ∞ if $\tilde{f}$ is holomorphic at 0.

If f is meromorphic at ∞, then by passing to $\tilde{f}$ we have a q-expansion of the form $f(\tau) = \sum_{n \in \mathbb{Z}} a_n q_h^n$ for suffciently large $\mathrm{Im}\, \tau$.

Definition 3.6. Let f be a weakly modular function of weight k and level Γ. We say f is meromorphic (resp. holomorphic/vanishing) at the cusp $\Gamma \cdot z, z \in \mathbb{P}^1_{\mathbb{Q}}$ if, choosing $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ such that $\alpha \cdot (\infty) = z$, $f|_k[\alpha]$ is meromorphic (resp. holomorphic/vanishing) at ∞ when viewed as a weakly modular function of level $\alpha^{-1}\Gamma\alpha$.
The width of the cusp $\Gamma \cdot z$ is the width of the cusp ∞ for the group $\alpha^{-1}\Gamma\alpha$.

Lemma 3.3. *The width and the meromorphy (resp. holomorphy/vanishing) of a weakly modular function at a cusp is well-defined.*

Proof. To see the width is well-defined, observe that for any $\gamma \in \Gamma, \delta \in \Gamma_\infty$,

$$\begin{aligned}[\Gamma_\infty : \Gamma_\infty \cap \{\pm 1\}(\gamma\alpha\delta)^{-1}\Gamma(\gamma\alpha\delta)] &= [\Gamma_\infty : \Gamma_\infty \cap \{\pm 1\}\delta^{-1}\alpha^{-1}\gamma^{-1}\Gamma\gamma\alpha\delta] \\ &= [\Gamma_\infty : \Gamma_\infty \cap \{\pm 1\}\alpha^{-1}\Gamma\alpha]\end{aligned}$$

To see the meromorphy of a weakly modular function at a cusp is well-defined, note that for any $\gamma \in \Gamma, \delta = \pm \left(\begin{smallmatrix} 1 & a \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_\infty$, we have

$$f|_k[\gamma\alpha\delta](\tau) = f|_k[\alpha\delta](\tau) = (f|_k[\alpha])|_k[\delta](\tau) = (\pm 1)^k f|_k[\alpha](\tau + a)$$

Suppose $f|_k[\alpha] = \tau = \sum_{n\in\mathbb{Z}} q_h^n$, then

$$f|_k[\gamma\alpha\delta](\tau) = \sum_{n\in\mathbb{Z}} a_n e^{2\pi i\tau/h} e^{2\pi i a/h}(-1)^k$$

which shows the result. □

Remark. The q-expansion of a weakly modular function f at a particular cusp (via the weight k action of the α as above) is, however, not independent of choice of α.
By "the" q-expansion of a weakly modular function which is meromorphic at ∞, we mean the q-expansion at ∞.

Definition 3.7. Let Γ be a congruence subgroup and f a weakly modular function of weight k and level Γ. We say f is a modular function (of weight k and level Γ) if f is meromorphic at every cusp of Γ. It is a modular form (of weight k and level Γ) if f is holomorphic on $\mathfrak{h}$ and at every cusp of Γ. It is a cuspidal modular form (of weight k and level Γ) if in addition that it vanishes at every cusp.
We write $M_k(\Gamma)$ as the $\mathbb{C}$-vector space of all modular forms of weight k and level Γ and $S_k(\Gamma)$ be its subspace containing cuspidal modular forms of weight k and level Γ.

Remark. An equivalent definition of what it means for f to be a modular form (of weight k and level Γ) is that f is holomorphic in $\mathfrak{h}$ and $f|_k[\alpha]$ is holomorphic at ∞ for any $\alpha \in \mathrm{SL}_2(\mathbb{Z})$.

Lemma 3.4. *1. If $f \in M_k(\Gamma)$ and $g \in M_l(\Gamma)$, then $fg \in M_{k+l}(\Gamma)$.*
2. If $\Gamma' \leq \Gamma$ is another congruence subgroup, then $M_k(\Gamma) \leq M_k(\Gamma')$.
3. If $\Gamma, \Gamma' \leq \mathrm{SL}_2(\mathbb{Z})$ are congruence subgroups and $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$ has $\Gamma' \leq \alpha^{-1}\Gamma\alpha$, then for any $f \in M_k(\Gamma)$, $f|_k[\alpha] \in M_k(\Gamma')$.

Remark. These also hold when one replace M_k by S_k.

Proof. 1 is obvious and 2 is a special case of 3.
To prove 3, note that $f|_k[\alpha]$ is holomorphic in $\mathfrak{h}$ and is invariant under the weight k action of $\alpha^{-1}\Gamma\alpha$, so we just need to check that $f|_k[\alpha]$ is holomorphic at the cusps of Γ'. Equivalently, we need to check that for all $\beta \in \Gamma(1)$, $f|_k[\alpha\beta]$ is holomorphic at ∞. Choose $\gamma \in \Gamma(1)$ such that $(\alpha\beta)\cdot(\infty) = \gamma\cdot(\infty)$. We can write $\alpha\beta = \gamma\delta$ for some $\delta \in \mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Q})^+}(\infty)$. But then δ has the form $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ for some $a, b, d \in \mathbb{Q}$ with $ad > 0$. Then

$$f|_k[\alpha\beta](\tau) = f|_k[\gamma\delta](\tau) = (ad)^{k-1}d^{-k}f|_k[\gamma]\left(\frac{a\tau + b}{d}\right)$$

which implies the claim. □

Example 3.3. Recall that we defined $\Gamma_0(N) = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \pmod N\}$. If $f \in M_k(\Gamma_0(M))$ and $L \in \mathbb{N}$, then the function $g(\tau) = f(L\tau)$ belongs to $M_k(\Gamma_0(ML))$. Indeed, we have $L\tau = \left(\begin{smallmatrix} L & 0 \\ 0 & 1 \end{smallmatrix}\right)\cdot\tau$. By the lemma, we need to check the inclusion

$$\Gamma_0(ML) \leq \begin{pmatrix} L & 0 \\ 0 & 1 \end{pmatrix}^{-1} \Gamma_0(M) \begin{pmatrix} L & 0 \\ 0 & 1 \end{pmatrix}$$

which follows from direct computation.
So if $k \in 2\mathbb{N}, k \geq 4$ and p is prime, then $G_k(\tau), G_k(p\tau)$ are both in $M_k(\Gamma_0(p))$.

3.2 An Instance of θ Function

Consider the q-expansion $\theta(\tau) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2\tau}$. Easily, the series converges absolutely and uniformly in compact subsets of $\mathfrak{h}$, hence defines a holomorphic function on $\mathfrak{h}$. We also have $\theta(\tau + 2) = \theta(\tau)$.

Proposition 3.5 (Poisson Summation Formula)**.** *Suppose $f : \mathbb{R} \to \mathbb{C}$ is a continuous function such that $\sum_{n\in\mathbb{Z}} |f(n + x)|$ converges uniformly in $\mathbb{R}$ and $\sum_{n\in\mathbb{Z}} |\hat{f}(n)|$ converges where*

$$\hat{f}(n) = \int_{-\infty}^{\infty} f(t)e^{-2\pi i n t}\,\mathrm{d}t$$

Then,

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \hat{f}(n)$$

Proof. Let $\phi(x) = \sum_{n\in\mathbb{Z}} f(n + x)$ which is continuous by assumption. We want to expand it as a Fourier series $\phi(x) = \sum_{n\in\mathbb{Z}} \hat{\phi}(n)e^{2\pi i n x}$. If we could do that, then the coefficients have the form

$$\begin{aligned}\hat{\phi}(n) &= \int_0^1 \sum_{n\in\mathbb{Z}} f(n + t)e^{-2\pi i n t}\,\mathrm{d}t = \sum_{n\in\mathbb{Z}} \int_0^1 f(n + t)e^{-2\pi i n t}\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty} f(t)e^{-2\pi i n t}\,\mathrm{d}t = \hat{f}(n)\end{aligned}$$

It is a fact from Fourier analysis that if $\psi : \mathbb{R} \to \mathbb{C}$ is a conitnuous function with period 1 and $\sum_{n\in\mathbb{Z}} |\hat{\psi}(n)| < \infty$, then the series $\sum_{n\in\mathbb{Z}} \hat{\psi}(n)e^{2\pi i n x}$ converges uniformly to $\psi(x)$. Applying this to ϕ then gives

$$\sum_{n\in\mathbb{Z}} f(n) = \phi(0) = \sum_{n\in\mathbb{Z}} \hat{f}(n)$$

which is the summation formula we were after. □

Applying this formula to $f_y(x) = e^{-\pi x^2 y}$ gives

$$\theta(iy) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 y} = \sum_{n\in\mathbb{Z}} f_y(n) = \sum_{n\in\mathbb{Z}} \hat{f}_y(n) = \sum_{n\in\mathbb{Z}} \frac{e^{-\pi s^2/y}}{\sqrt{y}} = \frac{\theta(i/y)}{\sqrt{y}}$$

for $y \in \mathbb{R}_+$. By identity principle, we can extend this to (taking principal branch of $\sqrt{\cdot}$).

$$\theta(\tau) = \frac{\theta(-1/\tau)}{\sqrt{\tau/i}}$$

for $\tau \in \mathfrak{h}$. It almost looks as if θ is a modular form of "weight 1/2", which is however not permitted in our definition. Nonetheless, we can construct some modular forms of certain weights and levels by using θ as a building block. Observe that $\theta^8|_4[\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right)] = \theta^8$ and $\theta^8|_4[\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)] = \theta^8$.

Proposition 3.6. *$\theta^8 \in M_4(\Gamma)$ where $\Gamma \leq \Gamma(1)$ is the subgroup of matrices γ such that either $\gamma \equiv 1 \pmod 2$ or $\gamma \equiv S \pmod 2$.*

Proof. Easy to see that θ^8 is holomorphic in $\mathfrak{h}$ and that Γ is generated by $\pm\left(\begin{smallmatrix} 1 & 2 \\ 0 & 1 \end{smallmatrix}\right), \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. It remains to show that θ^8 is holomorphic at the cusps $\Gamma\cdot\infty, \Gamma\cdot 1$ of Γ. The cusp $\Gamma\cdot\infty$ is easy to deal with since we know that θ^8 has the form

$$\theta^8 = \left(1 + \sum_{n\geq 1} q_2^{n^2}\right)^8$$

As for $\Gamma\cdot 1$, we need to show that $\theta^8|_2[\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)]$ is holomorphic at ∞. In fact, we shall show that it is not only holomorphic, but also vanishing at ∞. We have $\theta(\tau+1) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 \tau} e^{\pi i n^2} = \sum_{n\in\mathbb{Z}} (-1)^n q_2^n$, so

$$\theta(\tau) + \theta(\tau+1) = \sum_{n\in\mathbb{Z}} (1 + (-1)^n) q_2^{n^2} = \sum_{n\in 2\mathbb{Z}} 2q_2^{n^2} = 2\theta(4\tau)$$

The substitution $\tau \mapsto -1/\tau$ gives

$$\theta(-1/\tau) + \theta((\tau-1)/\tau) = 2\theta(-4/\tau) \implies \frac{\theta((\tau-1)/\tau)}{\sqrt{\tau/i}} = \theta(\tau/4) - \theta(\tau)$$

It follows that $\theta^8|_4[\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)] = (\theta(\tau/4) - \theta(\tau))^8$ which gives the result. □

What actually happens if we raise θ to a power? We have

$$\theta^k(\tau) = \left(\sum_{n\in\mathbb{Z}} q_2^{n^2}\right)^k = \sum_{n\in\mathbb{Z}_{\geq 0}} r_k(n) q_2^n$$

where $r_k(n)$ is the number of ways to write $n = m_1^2 + \cdots + m_k^2$ for $m_1, \ldots, m_k \in \mathbb{Z}$. Let $k = 8$. What are the elements of $M_4(\Gamma)$? Since $M_4(\Gamma(1)) \le M_4(\Gamma), E_4(\tau) \in M_4(\Gamma)$. Also we have $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right) \Gamma \left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right)^{-1} \le \Gamma(1)$ by direct calculation. Consequently, $E_4((\tau+1)/2) = (1/2) E_4|_4[\left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right)] \in M_4(\Gamma)$. We will later see some techniques that allows us to prove $\dim M_4(\Gamma) = 2$. Combining them shows that $E_4(\tau)$ and $E_4((\tau+1)/2)$ spans $M_4(\Gamma)$. In particular, θ^8 is a linear combination of $E_4(\tau)$ and $E_4((\tau+1)/2)$.
There are several ways to calculate the coefficients in this q-expansion. One of them is as follows: We know that $\theta^8(\infty) = 1$ and $\theta^8|_4[\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)](\infty) = 0$. At the same time, $E_4(\infty) = 1, E_4|_4[\left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)](\infty) = E_4(\infty) = 1$. As for $E_4((\tau+1)/2)$, we have

$$E_4\left(\frac{\tau+1}{2}\right) = 1 + \sum_{n \ge 1} \sigma_3(n) e^{2\pi i(\tau+1)/2} = 1 + \sum_{n \ge 1} \sigma_3(n)(-1)^n q_2^n$$

Therefore $E_4((\tau+1)/2)$ attains 1 at $\tau = \infty$. To find its value at the cusp $\Gamma \cdot 1$, we need to compute

$$\begin{aligned} 2E_4|_4\left[\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}\right](\tau) &= 2E_4|_4\left[\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right](\tau) \\ &= 2E_4|_4\left[\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\right](\tau) = 16E_4(2\tau) \end{aligned}$$

Putting $\tau = \infty$ shows that $E_4((\tau+1)/2)$ has value 16 at ∞. Incorporating these pieces of information gives

$$\theta^8(\tau) = \frac{16}{15} E_4(\tau) - \frac{1}{15} E_4\left(\frac{\tau+1}{2}\right)$$

And thus $r_8(n) = 2^8 \sigma_3(n/2) - 16(-1)^n \sigma_3(n)$ where, by convention, σ_3 vanishes at non-integers. In particular $r_8(p) = 16\sigma_3(p)$ for any odd prime p.

3.3 Hecke Operators

$M_k(\Gamma)$ and $S_k(\Gamma)$ are finite-dimensional vector spaces for any congruence subgroup Γ. We will prove this (with a more precise statement) later.

Lemma 3.7. *If* $\Gamma \le \Gamma(1)$ *is a congruence subgroup, then* $(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma)$ *is a Hecke pair.*

Proof. We shall show that $(g^{-1}\Gamma g \cap \Gamma)\backslash\Gamma$ is finite for any $g \in \mathrm{GL}_2(\mathbb{Q})^+$. Choose $N \in \mathbb{N}$ such that $\Gamma(N) \le \Gamma$ and $Ng, Ng^{-1} \in M_2(\mathbb{Z})$. We have $g\Gamma(N^3)g^{-1} \subset 1 + gN^3 M_2(\mathbb{Z}) g^{-1} \subset 1 + (Ng)NM_2(\mathbb{Z})(Ng^{-1}) \subset 1 + NM_2(\mathbb{Z})$. Hence $g\Gamma(N^3)g^{-1} \le \Gamma(N) \implies \Gamma(N^3) \le g^{-1}\Gamma(N)g \cap \Gamma(N) \le g^{-1}\Gamma g \cap \Gamma$. □

Definition 3.8. Let $k \in \mathbb{Z}$. We define

$$\mathcal{M}_k = \bigcup_{\Gamma \le \Gamma(1)} M_k(\Gamma), \mathcal{S}_k = \bigcup_{\Gamma \le \Gamma(1)} S_k(\Gamma)$$

Proposition 3.8. $\mathrm{GL}_2(\mathbb{Q})^+$ *acts on* $\mathcal{M}_k$ *and* $\mathcal{S}_k$ *via the weight* k *modular action.*

Proof. Suppose we have $f \in M_k(\Gamma)$ and $g \in \mathrm{GL}_2(\mathbb{Q})^+$, then $f|_k[g]$ is invariant under the weight k modular action of $g^{-1}\Gamma g$ which contains a congruence subgroup $\Gamma' \leq g^{-1}\Gamma g$. Then $f|_k[g] \in M_k(\Gamma')$. The proof for $\mathcal{S}_k$ is analogous. □

Proposition 3.9. *$\mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma)$ acts on $M_k(\Gamma)$ and $S_k(\Gamma)$.*

Proof. General theory on Hecke algebra. □

$\mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$ is commutative and $S_k(\Gamma(1))$ decomposes with multiplicity 1 under the action of $\mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma(1))$. Sadly, we cannot generalise these facts directly to other congruence subgroups: In general, $\mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma)$ is not commutative. How shall we proceed, then?
Let's look at a particular family of congruence subgroups we've come across before.

Definition 3.9. For $N \in \mathbb{N}$, we define

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : c \equiv 0 \pmod{N} \right\}$$

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : c \equiv 0 \pmod{N}, a \equiv 1 \pmod{N} \right\}$$

Recall the fact that we have the surjective map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ via reduction. In particular, $\Gamma_1(N) \lhd \Gamma_0(N)$ and $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^\times$ via the map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto d \bmod N$. Conversely, for $d \in \mathbb{Z}, (d, N) = 1$, we can make sense of the endomorphism $\langle d \rangle$ of $M_k(\Gamma_1(N))$ induced by

$$\left[\left\{ \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \Gamma_0(N) : D \equiv d \pmod{N} \right\} \right] \in \mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma_1(N))$$

For a prime p, we write T_p for the endomorphism of $M_k(\Gamma_1(N))$ induced by

$$\left[\Gamma_1(N) \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \Gamma_1(N) \right] \in \mathcal{H}(\mathrm{GL}_2(\mathbb{Q})^+, \Gamma_1(N))$$

Remark. When $N = 1$, this is exactly the same sets of T_p that we had for $\Gamma(1)$. In general, if $p \nmid N$, T_p behaves in a similar way to the $N = 1$ case. Things can, however, get trickier when $p \mid N$.

How does $\langle d \rangle$ act on $M_k(\Gamma_1(N))$?

Lemma 3.10. *If $f \in M_k(\Gamma_1(N))$ and $\gamma = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in \Gamma_0(N)$ and $D \equiv d \pmod{N}$, then $f|_k[\gamma] = \langle d \rangle(f)$*

Proof. We in fact have

$$\left\{ \begin{pmatrix} X & Y \\ Z & W \end{pmatrix} \in \Gamma_0(N) : W \equiv d \pmod{N} \right\} = \Gamma_1(N) \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

So $\langle d \rangle(f) = f|_k[\gamma]$. □

Lemma 3.11. *If A is a finite abelian group which acts by linear maps on a $\mathbb{C}$-vector space V, then*

$$V = \bigoplus_{\chi \in \mathrm{Hom}(A, \mathbb{C}^\times)} V_\chi, V_\chi = \{v \in V : \forall a \in A, a \cdot v = \chi(a) v\}$$

Proof. Representation theory. □

Corollary 3.12. *We have* $M_k(\Gamma_1(N)) = \bigoplus_\chi M_k(\Gamma_1(N),\chi)$ *and* $S_k(\Gamma_1(N)) = \bigoplus_\chi S_k(\Gamma_1(N),\chi)$ *where the direct sums run over the set of homomorphisms* $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ *and*

$$M_k(\Gamma_1(N),\chi) = \{f \in M_k(\Gamma_1(N)) : \forall d \in \mathbb{Z}, (d,N)=1, \langle d\rangle f = \chi(d)f\}$$

$$S_k(\Gamma_1(N),\chi) = \{f \in S_k(\Gamma_1(N)) : \forall d \in \mathbb{Z}, (d,N)=1, \langle d\rangle f = \chi(d)f\}$$

Proposition 3.13. *1. If* $p \mid N$*, then*

$$\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\Gamma_1(N) = \coprod_{j=0}^{p-1}\Gamma_1(N)\begin{pmatrix}1 & j\\ 0 & p\end{pmatrix}$$

2. If $p \nmid N$*, then*

$$\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\Gamma_1(N) = \coprod_{j=0}^{p-1}\Gamma_1(N)\begin{pmatrix}1 & j\\ 0 & p\end{pmatrix} \sqcup \Gamma_1(N)\begin{pmatrix}p^a & r\\ pN & ps\end{pmatrix}$$

for any integers r, s *such that* $p^a s - Nr = 1$.

In other words, $T_p f = \sum_{j=0}^{p-1} f|_k\left[\left(\begin{smallmatrix}1 & j\\ 0 & p\end{smallmatrix}\right)\right]$ if $p \mid N$ and $T_p f = f|_k\left[\left(\begin{smallmatrix}p^a & r\\ pN & ps\end{smallmatrix}\right)\right] + \sum_{j=0}^{p-1} f|_k\left[\left(\begin{smallmatrix}1 & j\\ 0 & p\end{smallmatrix}\right)\right]$ for some $p^a s - Nr = 1$ if $p \nmid N$.

Proof. Recall that for any $g \in \mathrm{GL}_2(\mathbb{Q})^+$, there is a bijection

$$(g^{-1}\Gamma_1(N)g \cap \Gamma_1(N))\backslash\Gamma_1(N) \to \Gamma_1(N)\backslash(\Gamma_1(N)g\Gamma_1(N)), \gamma \mapsto \Gamma_1(N)g\gamma$$

We want to compute

$$\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}^{-1}\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix} = \left\{\begin{pmatrix}a & pb\\ p^{-1}c & d\end{pmatrix} : \begin{pmatrix}a & b\\ c & d\end{pmatrix} \in \Gamma_1(N)\right\}$$

In particular,

$$\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}^{-1}\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix} \cap \Gamma_1(N) \le \left\{\begin{pmatrix}a & b\\ c & d\end{pmatrix} \in \Gamma_1(N) : p \mid b\right\}$$

Conversely, if $\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_1(N)$ and $p \mid b$, then

$$\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}^{-1} = \begin{pmatrix}a & b/p\\ pc & d\end{pmatrix} \in \Gamma_1(N)$$

So in fact

$$\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}^{-1}\Gamma_1(N)\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix} \cap \Gamma_1(N) = \left\{\begin{pmatrix}a & b\\ c & d\end{pmatrix} \in \Gamma_1(N) : p \mid b\right\}$$

Let Γ be this group.
For part 1 of the proposition, observe that Γ is the pre-image under the surjective homomorphism $\Gamma_1(N) \to \{\left(\begin{smallmatrix}1 & a\\ 0 & 1\end{smallmatrix}\right) : a \in \mathbb{Z}/N\mathbb{Z}\}$ via reduction of the subgroup

$\{\left(\begin{smallmatrix}1 & a\\ 0 & 1\end{smallmatrix}\right) : a \in p\mathbb{Z}/N\mathbb{Z}\}$. Any set $\gamma_0, \dots, \gamma_{p-1}$ in $\Gamma_1(N)$ projecting to a set of representatives for $\{\left(\begin{smallmatrix}1 & a\\ 0 & 1\end{smallmatrix}\right) : a \in p\mathbb{Z}/N\mathbb{Z}\} \setminus \{\left(\begin{smallmatrix}1 & a\\ 0 & 1\end{smallmatrix}\right) : a \in \mathbb{Z}/N\mathbb{Z}\}$ will give representation for $\Gamma\backslash\Gamma_1(N)$. We can take $\gamma_j = \left(\begin{smallmatrix}1 & j\\ 0 & 1\end{smallmatrix}\right)$ which then gives the identity.
For part 2 of the proposition, Γ is the preimage in $\Gamma_1(N)$ of the surjective homomorphism $\Gamma_1(N) \to \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ of the subgroup $\{\left(\begin{smallmatrix}a & 0\\ c & d\end{smallmatrix}\right) : ad = 1\}$. Now

$$\left\{\begin{pmatrix}a & 0\\ c & d\end{pmatrix} : ad = 1\right\} \setminus \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z}) \cong \{v \in \mathbb{F}_p^2, v \neq 0\}/\mathbb{F}_p^\times$$

via $g \mapsto \mathbb{F}_p^\times \cdot (1\ 0)g$. So for any matrices $\gamma_0, \dots, \gamma_p$ in $\Gamma_1(N)$ such that $\{\mathbb{F}_p^\times \cdot (1\ 0)\gamma_i\} = \{v \in \mathbb{F}_p^2, v \neq 0\}/\mathbb{F}_p^\times$, they are a transversal for $\Gamma\backslash\Gamma_1(N)$. We can take $\gamma_j = \left(\begin{smallmatrix}1 & j\\ 0 & 1\end{smallmatrix}\right)$ for $j = 0, \dots, p-1$ and $\gamma_p = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$ such that $N \mid c, p \mid a, a \equiv d \equiv 1 \pmod N$. Choose $a \geq 1$ such that $p^a \equiv 1 \pmod N$ (e.g. $a = \phi(N)$) and $r, s \in \mathbb{Z}$ such that $p^a s - Nr = 1$, then $\gamma_p = \left(\begin{smallmatrix}p^a & r\\ N & s\end{smallmatrix}\right)$ works. The corresponding representatives for $\Gamma_1(N)\backslash\Gamma_1(N)\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)\Gamma_1(N)$ are given by

$$\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\gamma_j = \begin{pmatrix}1 & j\\ 0 & p\end{pmatrix}, j = 0, \dots, p-1, \begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\gamma_p = \begin{pmatrix}p^a & r\\ pN & ps\end{pmatrix}$$

which gives what we wanted. □

The proof actually gives us something more general: If $p \mid N$, then $T_p f = \sum_{j=0}^{p-1} f|_k \left[\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)\gamma_j\right]$ for any $\gamma_0, \dots, \gamma_{p-1} \in \Gamma_1(N)$ projecting to representatives for the quotient

$$\left\{\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) : b \equiv 0 \pmod p\right\} \setminus \left\{\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})\right\}$$

if $p \nmid N$, then $T_p f = \sum_{j=0}^{p} f|_k \left[\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right)\gamma_j\right]$ for any $\gamma_0, \dots, \gamma_p \in \Gamma_1(N)$ projecting to representatives for the quotient

$$\left\{\begin{pmatrix}a & 0\\ c & d\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})\right\} \setminus \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$$

3.4 Diagonalisation; Newforms

Lemma 3.14. *$\langle d\rangle$ commutes with T_p.*

Proof. We shall prove the case where $p \nmid N$. Choose $\delta = \left(\begin{smallmatrix}A & B\\ C & D\end{smallmatrix}\right) \in \Gamma_0(N)$ with $\delta \equiv 1 \pmod{p^2}$.

$$\langle d\rangle T_p f = \sum_{j=0}^{p} f|_k \left[\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\gamma_j\delta\right], T_p\langle d\rangle f = \sum_{j=0}^{p} f|_k \left[\delta\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\gamma_j\right]$$

So the lemma shall follow if

$$\epsilon_j = \begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}^{-1}\delta^{-1}\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}\gamma_j\delta$$

lie in $\Gamma_1(N)$ and project to representatives for the quotient

$$\left\{\begin{pmatrix}a & 0\\ c & d\end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})\right\} \setminus \mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$$

Indeed,

$$\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix}^{-1}\delta^{-1}\begin{pmatrix}1 & 0\\ 0 & p\end{pmatrix} = \begin{pmatrix}D & -pB\\ -C/p & A\end{pmatrix} \in \Gamma_0(N)$$

and is congruent to 1 modulo p. So $\epsilon_j \in \Gamma_1(N)$ and $\epsilon_j \equiv \gamma_j \pmod p$. □

Corollary 3.15. *The operators T_p preserve the subspaces $M_k(\Gamma_1(N), \chi)$ and $S_k(\Gamma_1(N), \chi)$ for characters $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$.*

Proposition 3.16. *Let $f \in M_k(\Gamma_1(N))$ and write $f(\tau) = \sum_{n\geq 0} a_n(f)q^n$, then*

$$T_p(f) = \sum_{n\geq 0}\left(a_{np}(f) + p^{k-1}1_{p\nmid N}a_{n/p}(\langle p\rangle f)\right)q^n$$

where by convention $a_{n/p} = 1_{p|n}a_{n/p}$.

Proof.

$$\begin{aligned}
T_p f &= \sum_{j=1}^{p-1} f|_k\left[\begin{pmatrix}1 & j\\ 0 & p\end{pmatrix}\right] + 1_{p\nmid N} f|_k\left[\begin{pmatrix}p^a & r\\ pN & ps\end{pmatrix}\right]\\
&= p^{k-1}p^{-k}\sum_{j=0}^{p-1} f\left(\frac{\tau+j}{p}\right) + 1_{p\nmid N} f|_k\left[\begin{pmatrix}p^{a-1} & r\\ N & ps\end{pmatrix}\begin{pmatrix}p & 0\\ 0 & 1\end{pmatrix}\right]\\
&= \frac{1}{p}\sum_{n\geq 0}\sum_{j=0}^{p-1} a_n(f)e^{2\pi i n\tau/p}e^{2\pi i nj/p} + 1_{p\nmid N}(\langle p\rangle f)|_k\left[\begin{pmatrix}p & 0\\ 0 & 1\end{pmatrix}\right]\\
&= \sum_{n\geq 0} a_{np}(f)q^n + 1_{p\nmid N}\sum_{n\geq 0} a_n(\langle p\rangle f)q^{np}
\end{aligned}$$

which is equivalent to the claimed identity. □

Corollary 3.17. *If p, q are primes, then T_p and T_q commute.*

Proof. Computation reveals that $a_n(T_pT_qf)$ is symmetric in p, q. □

Recall that we can define an inner product on $S_k(\Gamma_1(N))$ by

$$\langle f, g\rangle = [\Gamma(1) : \Gamma_1(N)]^{-1}\int_{\Gamma_1(N)\backslash\mathfrak{h}} f(\tau)\overline{g(\tau)}y^k\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$$

Proposition 3.18. *If $d \in (\mathbb{Z}/N\mathbb{Z})^\times$, then its adjoint is $\langle d\rangle^* = \langle d^{-1} \bmod N\rangle$. if $p \nmid N$, then its adjoint is $T_p^* = \langle p^{-1} \bmod N\rangle T_p$.*

Corollary 3.19. *The operators $\langle d\rangle$ and T_p (for $p \nmid N$) are simultaneously diagonalisable on $S_k(\Gamma_1(N))$.*

Remark. If $p \mid N$, then T_p is not in general diagonalisable on $S_k(\Gamma_1(N))$. Meanwhile, the simultaneous eigenspaces of $\langle d\rangle$ and T_p ($p \nmid N$) are not one-dimensional in general either. Both of these defects are explained by the presence of so-called "old forms".

To see the simultaneous eigenspaces don't have to be one-dimensional, suppose $N > 1$ and $f \in S_k(\Gamma(1))$ is a normalised eigenform. Then $f(\tau), g(\tau) = f(N\tau) \in S_k(\Gamma_1(N))$ are linearly independent. If $p \nmid N$ is a prime, then f, g are eigenvectors of T_p with the same eigenvalue $a_p(f)$. Indeed, note that $T_p f$ is the same no matter whether we take the definition of T_p in $\Gamma(1)$ or $\Gamma_1(N)$, so $T_p f = a_p(f) f$. As for g,

$$\begin{aligned} a_n(T_p g) &= a_{np}(g) + p^{k-1} a_{n/p}(g) = a_{np/N}(f) + p^{k-1} a_{n/(pN)}(f) \\ &= a_{n/N}(T_p f) = a_p(f) a_{n/N}(f) = a_p(f) a_n(g) \end{aligned}$$

So the defects happened when we take some eigenforms from some strictly bigger congruence subgroups ("old forms").

Definition 3.10. Let $N \in \mathbb{N}$. The old subspace $S_k^{\text{old}}(\Gamma_1(N)) \leq S_k(\Gamma_1(N))$ consists of all forms of type $f(M\tau)$ where $M \mid N, M > 1$ and $f \in S_k(\Gamma_1(N/M))$. The new subspace $S_k^{\text{new}}(\Gamma_1(N))$ is the orthogonal complement of $S_k^{\text{old}}(\Gamma_1(N))$ in $S_k(\Gamma_1(N))$.

Theorem 3.20. *All the operators $\langle d \rangle, T_p$ preserve the direct sum decomposition $S_k(\Gamma_1(N)) = S_k^{\text{old}}(\Gamma_1(N)) \oplus S_k^{\text{new}}(\Gamma_1(N))$. Furthermore, the simultaneous eigenspaces of the operators $\langle d \rangle, T_p$ ($d \in (\mathbb{Z}/N\mathbb{Z})^\times, p \nmid N$) on $S_k^{\text{new}}(\Gamma_1(N))$ are one-dimensional. In addition, $S_k^{\text{new}}(\Gamma_1(N))$ has a unique basis of normalised eigenforms, i.e. forms $f \in S_k^{\text{new}}(\Gamma_1(N))$ which are eigenvectors for all these operators $\langle d \rangle, T_p$ and satisfy $a_1(f) = 1$.*

Definition 3.11. A newform of weight k and level $\Gamma_1(N)$ is a normalised eigenform in $S_k^{\text{new}}(\Gamma_1(N))$.

Example 3.4. $S_2(\Gamma(1)) = 0$, so if p is a prime, so $S_2(\Gamma_1(p)) = S_2^{\text{new}}(\Gamma_1(p))$. The first prime p for which this space is nonzero is $p = 11$, in which case it has dimension 1 (as will be shown later). So it's basis consist of a (new) eigenform. What is it? Recall from example sheet the expression for the Dedekind η function $\eta(\tau) = q_{24} \prod_{n \geq 1}(1 - q^n)$. Turns out $\eta(\tau)^2 \eta(11\tau)^2$ is the unique newform in $S_2(\Gamma_1(11))$.

The eigenforms in $\Gamma(1)$ has relations to L-functions, so there should certainly be some sort of results in this flavour in this new context.

Theorem 3.21. *Let $f \in S_k(\Gamma_1(N))$ be a newform, say $f(\tau) = \sum_{n \geq 1} a_n q^n$ and define $L(f, s) = \sum_{n \geq 1} a_n n^{-s}$. Then:*
1. $L(f, s)$ is absolutely convergent in $\operatorname{Re} s > 1 + k/2$.
2. There is a character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$ such that $f \in S_k(\Gamma_1(N), \chi)$.
3. There is an Euler product $L(f, s) = \prod_p (1 - a_p p^{-s} + \chi(p) p^{k-1-2s})^{-1}$ (where by convention $\chi(p) = 1_{p \nmid N} \chi(p)$).
4. $L(f, s)$ has an analytic continuation to $\mathbb{C}$. Furthermore, if we define $\Lambda(f, s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f, s)$, then there is a sign $\epsilon_f \in \{\pm 1\}$ such that $\Lambda(f, s) = \epsilon_f \Lambda(f, k - s)$.

Theorem 3.22 (Shimura-Taniyama-Weil Conjecture, aka Modularity Theorem)**.** *There is a bijection between the collection of elliptic curves $E/\mathbb{Q}$ up to isogeny and the newforms $f \in S_2(\Gamma_1(N), 1)$ with integer coefficients. Specifically, an elliptic curve E corresponds to f if for all but finitely many primes p, $a_p(f) = p + 1 - |\tilde{E}(\mathbb{F}_p)|$.*

4 Modular Curves

Given a congruence subgroup Γ, our goal is to endow $\Gamma\backslash\mathfrak{h}$ and $\Gamma\backslash\mathfrak{h} \sqcup \Gamma\backslash\mathbb{P}^1_{\mathbb{Q}}$ with the structure of Riemann surfaces. They are known as modular curves.

4.1 Conformal Structures of Modular Curves

Definition 4.1. A Riemann surface is a connected Hausdorff topological space X together with an atlas of charts $\mathcal{A} = \{(U_i, V_i, \phi_i : U_i \to V_i) : i \in I\}$ such that:
(a) $\forall i \in I, U_i \subset X, V_i \subset \mathbb{C}$ are open.
(b) $\forall i \in I, \phi_i$ is a homeomorphism.
(c) $\forall i, j \in I, \phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$ is holomorphic.
(d) $\bigcup_{i \in I} U_i = X$.

Definition 4.2. If X, Y are Riemann surfaces, then a function $f : X \to Y$ is a morphism if it's a continuous map such that for any chart $\phi : U \to V$ of X and $\psi : U' \to V'$ of Y, $\psi \circ f \circ \phi^{-1} : V \cap f^{-1}(U') \to V'$ is holomorphic.

Example 4.1. 1. Take $X = \mathbb{C}$ and $\mathcal{A} = \{(\mathbb{C}, \mathbb{C}, \mathrm{id}_{\mathbb{C}})\}$. Then X is a Riemann surface and morphisms $f : Y \to \mathbb{C}$ would be continuous functions such that for any chart $\phi : U \to V$ of Y, $f \circ \phi^{-1} : V \to \mathbb{C}$ is holomorphic.
2. Take $X = \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ and $\mathcal{A} = \{(\mathbb{C}, \mathbb{C}, \mathrm{id}_{\mathbb{C}}), (\mathbb{C}^\times \cup \{\infty\}, \mathbb{C}, \phi)\}$ with $\phi(z) = 1/z$. Then $\phi \circ \mathrm{id}_{\mathbb{C}}^{-1} : \mathbb{C}^\times \to \mathbb{C}^\times$ is $z \mapsto 1/z$ which is holomorphic, so X is a Riemann surface.

Our first goal is to make $Y(\Gamma) = \Gamma\backslash\mathfrak{h}$ a Riemann surface. The first step is to make it a connected Hausdorff topological space. Give $\mathfrak{h}$ its usual topology and $Y(\Gamma)$ the quotient topology. Let $\pi : \mathfrak{h} \to Y(\Gamma)$ be the quotient map, i.e. $U \subset Y(\Gamma)$ is open iff $\pi^{-1}(U)$ is open. $Y(\Gamma)$ has its connectednes inherited from π. As for Hausdorff-ness,

Lemma 4.1. *Suppose $A, B \subset \mathfrak{h}$ are compact subsets, then $\{\gamma \in \Gamma(1) : \gamma A \cap B \neq \varnothing\}$ is finite.*

Proof. Let $\Omega_X = \{\tau \in \mathfrak{h} : |\operatorname{Re}\tau| \leq X, \operatorname{Im}\tau \geq 1/X\}$. We showed that $\exists c_X > 0, \forall \tau \in \Omega_X, \forall t \in \mathbb{R}, |\tau + t| \geq c_X \max\{1, |t|\}$. If $c, d \in \mathbb{Z}$, then $|c\tau + d| = |\tau + dc^{-1}||c| \geq C_X|c|\sup\{1, |dc^{-1}|\} = C_X \sup(|c|, |d|)$. hence if $\tau \in \Omega_X$, then $\operatorname{Im}(\gamma\tau) \leq \operatorname{Im}(\tau)/(C_X^2 \sup\{|c|, |d|\}^2)$, hence $\sup(|c|, |d|)^2 \leq \operatorname{Im}(\tau)/(C_X^2 \operatorname{Im}(\gamma\tau))$. Now suppose that $A, B \in \Omega_X$. If $\tau \in A, \gamma\tau \in B$, then $\sup\{|c|, |d|\}$ is bounded above by a constant only depending on A, B. If $\gamma \in \Gamma(1)$ and $\gamma A \cap B \neq \varnothing$, then there are only finitely many possibilities for (c, d). if γ, δ have the same denominator, then $\delta = \epsilon\gamma$ for some $\epsilon \in \Gamma_\infty = \{\pm \left(\begin{smallmatrix} 1 & m \\ 0 & 1 \end{smallmatrix}\right) : m \in \mathbb{Z}\}$. Since A, B are compact hence bounded, there can only be finitely many ϵ such that $\epsilon\gamma A \cap B \neq \varnothing$. $\square$

Proposition 4.2. *$Y(\Gamma)$ is Hausdorff.*

Proof. Let $x, y \in \mathfrak{h}$ be such that x, y are not Γ-conjugate. Enough to find open sets $W_x \ni x$ and $W_y \ni y$ of $\mathfrak{h}$ such that if $\gamma \in \Gamma(1)$ then $\gamma W_x \cap W_y = \varnothing$. Choose open subsets $U_x \ni x$ and $U_y \ni y$ and compact subsets $K_x \supset U_x$ and $K_y \supset U_y$ of $\mathfrak{h}$. The lemma shows that the set $N = \{\gamma \in \Gamma : K_x \cap \gamma K_y \neq \varnothing\}$ is finite. For all $\gamma \in N$, choose open sets $A_\gamma \ni x$ and $B_\gamma \ni \gamma y$ such that $A_\gamma \cap B_\gamma = \varnothing$ as $\mathfrak{h}$ is Hausdorff. We now take $W_x = U_x \cap \bigcap_{\gamma \in N} A_\gamma \ni x$ and

$W_y = U_y \cap \bigcap_{\gamma \in N} \gamma^{-1} B \gamma \ni y$, which are both open and $W_x \cap \gamma W_y = \varnothing$ for any $\gamma \in \Gamma$. $\square$

How about $X(\Gamma) = \Gamma \backslash \mathfrak{h} \sqcup \Gamma \backslash \mathbb{P}^1_{\mathbb{Q}} = \Gamma \backslash (\mathfrak{h} \sqcup \mathbb{P}^1_{\mathbb{Q}})$? Let $\mathfrak{h}^* = \mathfrak{h} \sqcup \mathbb{P}^1_{\mathbb{Q}}$ on which Γ acts on. Give $\mathfrak{h}^*$ a topology by declaring a basis that consists of open sets on $\mathfrak{h}$ and sets of the form $\alpha V_R \cup \{z\} = \alpha(V_R \cup \{\infty\})$ where $z = \alpha\infty$ for $\alpha \in \Gamma(1)$ and $V_R = \{\tau \in \mathfrak{h} : \operatorname{Im} \tau > R\}$. Note that $\alpha(V_R \cup \{\infty\})$ only depends on z and R and not on the choice of α.
Any $\gamma \in \Gamma(1)$ is a homeomorphism $\mathfrak{h} \to \mathfrak{h}$, so $X(\Gamma)$ inherits a quotient topology from the quotient map $\pi : \mathfrak{h}^* \to X(\Gamma)$.
Note that if $R > 2$, then $\forall \gamma \in \Gamma(1), \gamma V_R \cap V_R \neq \varnothing \implies \gamma \in \Gamma_\infty = \operatorname{Stab}_{\Gamma(1)}(\infty)$. Indeed, if $\tau \in \gamma V_R \cap V_R$, say $\tau = \gamma\tau'$ for some $\tau' \in V_R$, then $\exists \delta, \epsilon \in \Gamma_\infty$ such that $\delta\tau, \epsilon\tau' \in \mathcal{F}'$. But $\delta\tau, \epsilon\tau'$ are $\Gamma(1)$-conjuagte, so $\delta\tau = \epsilon\tau'$, i.e. $\tau = \delta^{-1}\epsilon\gamma\tau \implies \delta^{-1}\epsilon\gamma = \pm 1 \implies \gamma \in \Gamma_\infty$, which is what we wanted.
We can draw a commutative diagram

$$\begin{array}{ccc} \mathfrak{h}^* & \longrightarrow & X(\Gamma) \\ \uparrow & & \uparrow \\ V_R & \longrightarrow & V_R/(\gamma \cap \gamma_\infty) \end{array}$$

where the map $V_R/(\Gamma \cap \Gamma_\infty) \to X(\Gamma)$ is injective, continuous, and open. This means that it is in fact an embedding. What does $V_R/(\Gamma \cap \Gamma)$ look like? Note that $\Gamma \cap \Gamma_\infty$ has the form of either $\{\left(\begin{smallmatrix} 1 & hn \\ 0 & 1 \end{smallmatrix}\right)\}$ or $\{\pm 1\}\{\left(\begin{smallmatrix} 1 & hn \\ 0 & 1 \end{smallmatrix}\right)\}$ (depending on whether $-1 \in \Gamma$), where h is the width of the cusp ∞. There is a continuous bijection $V_R/(\Gamma \cap \Gamma_\infty) \to D_R = \{q_h \in \mathbb{C} : |q_h| < e^{-2\pi R/h}\}$ via $\infty \mapsto 0, (\Gamma \cap \Gamma_\infty) \cdot \gamma \mapsto e^{2\pi i \tau/h}$ which is also open by open mapping theorem. So $\pi(V_R)$ is an open neighbourhood of $\pi(\infty)$ in $X(\Gamma)$ which is homeomorphic to the disc $D_R \subset \mathbb{C}$.
Similarly, if $\alpha \in \Gamma(1)$ and $\Gamma \cdot \alpha\infty$ is a cusp of Γ of width h, then $\pi(\alpha V_R)$ (for any $R > 0$) is an open neighbourhood of $\pi(\alpha\infty)$ which is homeomorphic to the disk D_R via $\alpha\infty \mapsto 0, \Gamma \cdot \tau \mapsto e^{2\pi i \alpha^{-1}\tau/h}$ for $\tau \in \alpha V_R$.

Theorem 4.3. *$X(\Gamma)$ is connected, Hausdorff, and compact.*

Proof. It is easy to see that $\mathfrak{h}^*$ is connected, so $X(\Gamma)$ inherits this connectedness from π.
To see $X(\Gamma)$ is Hausdorff, it is enough to find open sets that seperate the pairs $\Gamma \cdot \tau, \Gamma \cdot \alpha\infty$ and $\Gamma \cdot \alpha\infty, \Gamma \cdot \beta\infty$.
In the first case, we can find an open set $U_\tau \in \mathfrak{h}$ and a compact set $K_\tau \in \mathfrak{h}$ such that $\tau \in U_\tau \subset K_\tau$. We already know previously that $\exists C_\tau > 0$ such that $\operatorname{Im}(\gamma z) \leq C_\tau$ for any $\gamma \in \Gamma(1), z \in K_\tau$. If $R > C_\tau$, then $\gamma K_\tau \cap V_R = \varnothing$ for any $\gamma \in \Gamma(1)$, hence $\gamma U_\tau \cap \alpha V_R = \varnothing$ for any $\gamma \in \Gamma$. This means that $\pi(U_\tau)$ and $\pi(\alpha V_R)$ are disjoint.
In the second case, suppose $\Gamma \cdot \alpha\infty$ and $\Gamma \cdot \beta\infty$ are distinct cusps, then αV_R and βV_R are disjoint when $R > 2$ since if $\gamma \in \Gamma$ has $\gamma\alpha V_R \cap \beta V_R \neq \varnothing$, then $\exists \delta \in \Gamma_\infty$ such that $\gamma\alpha = \beta\delta$, therefore $\gamma\alpha\infty = \beta\infty$, contradiction. So $\pi(\alpha V_R)$ and $\pi(\beta V_R)$ are disjoint and seperate the two cusps.
How do we show compactness? Note that $\mathcal{F}^* = \mathcal{F} \cup \{\infty\}$ is a compact subset of $\mathfrak{h}^*$ since $\mathcal{F}^* - V_R$ is closed and bounded hence compact for any $R > 0$. Take a transversal $\gamma_1, \ldots, \gamma_n$ for $\Gamma \backslash \Gamma(1)$, then $\pi\left(\bigcup_{i=1}^n \gamma_i \mathcal{F}^*\right) = X(\Gamma)$ is compact as a continuous image of a compact set. $\square$

We can give $X(\Gamma)$ the structure of Riemann surface by giving explicit charts. Let I be a set of representatives for the Γ-orbits in $\mathfrak{h}^*$. For each $x \in I$, we need to construct a chart $\phi_x : U_x \to V_x$ where $U_x \subset X(\Gamma)$ is an open neighbourhood of $\pi(x)$ and $V_x \in \mathbb{C}$ is open.
If $x \in \mathbb{P}^1_{\mathbb{Q}}$, say $x = \alpha\infty$ for some $\alpha \in \Gamma(1)$, then we define $U_x = \pi(\alpha V_2), V_x = \{q_h \in \mathbb{C} : |q_h| < e^{-4\pi/h}\}$ and $\phi_x(\pi(\alpha\infty)) = 0, \phi_x(\pi(\alpha\tau)) = e^{2\pi i\tau/h}$ for $\tau \in V_2$. This is a homeomorphism as we've already seen.
If $x \in \mathfrak{h}$, we first show that we can find an open set $W_x \subset \mathfrak{h}$ such that $x \in W_x$, W_x is stable under $\mathrm{Stab}_\Gamma(x)$ and if $y \in W_x$ and $y \neq x$, then $\mathrm{Stab}_\Gamma(y) \leq \{\pm 1\}$. To do this, choose $M_x \subset \mathfrak{h}$ open and $K_x \subset \mathfrak{h}$ compact such that $x \in M_x \subset K_x$ and let $N = \{\gamma \in \Gamma : \gamma K_x \cap K_x \neq \varnothing\}$. N is finite and contains $\mathrm{Stab}_\Gamma(x)$. If $\gamma \in N - \mathrm{Stab}_\Gamma(x)$, then $\gamma x \neq x$, so we can find $A_\gamma, B_\gamma \subset \mathfrak{h}$ open such that $x \in A_\gamma, \gamma x \in B_\gamma$ and $A_\gamma \cap B_\gamma = \varnothing$. We define $W_x = M_x \cap \bigcap_{\gamma \in N - \mathrm{Stab}_\Gamma(x)} (A_\gamma \cap \gamma^{-1} B\gamma)$ which is open in $\mathfrak{h}$ and contains x. Also, if $\gamma \in \Gamma$ and $W_x \cap \gamma W_x \neq \varnothing$, then $\gamma \in N$, so unless $\gamma \notin \mathrm{Stab}_\Gamma(x)$ we must have $W_x \cap \gamma W_x \subset A_\gamma \cap B_\gamma = \varnothing$ which is a contradiction.
Consider the Möbius transformation $\psi_x(z) = (z - x)/(z - \bar{x})$, then $\psi_x(x) = 0$, $\psi_x(\infty) = \infty$, and $\psi_x \mathrm{Stab}_\Gamma(x)\psi_x^{-1}$ stablises both 0 and ∞ – but any Möbius transformation which stablises 0 and ∞ and has finite order must be given by $z \mapsto \zeta z$ for some root of unity ζ.
By further shrinking W_x, we can assume WLOG that $\psi_x(W_x)$ is an open disk D_ϵ of radius ϵ around 0 for some $\epsilon \in (0, 1)$. Consequently, W_x is stable under $\mathrm{Stab}_\Gamma(x)$.
If $y \in W_x$ and $y \neq x$ and $\delta \in \Gamma$ such that $\delta y = y$, then $\delta W_x \cap W_x \neq \varnothing$, therefore $\delta \in \mathrm{Stab}_\Gamma(x)$. $\psi_x \delta \psi_x^{-1}$ stabilises $\psi_x(y) \neq 0, \infty$, so $\psi_x \delta \psi_x^{-1}$ is trivial, therefore $\delta \in \{\pm 1\}$.
We conclude that $\pi(W_x)$, which is an open subset of $X(\Gamma)$ that contains $\pi(x)$, is homeomorphic to $W_x/\mathrm{Stab}_\Gamma(x) \cong D_\epsilon/\psi_x \mathrm{Stab}_\Gamma(x)\psi_x^{-1}$. We know that group $\psi_x \mathrm{Stab}_\Gamma(x)\psi_x^{-1}/(\Gamma \cap \{\pm 1\})$ is finite and consists of rotations by roots of unity, therefore is cyclic of some order n_x and is generated by $z \mapsto e^{2\pi i/n_x} z$. It follows that $D_\epsilon/\psi_x \mathrm{Stab}_\Gamma(x)\psi_x^{-1}$ is homeomorphic to $D_{\epsilon^{n_x}}$ via the map $z \mapsto z^{n_x}$. Our desired charts around $\pi(x)$ can then be defined by taking $U_x = \pi(W_x), V_x = D_{\epsilon^{n_x}}, \phi_x : U_x \to V_x$ defined by $\phi_x(\pi(z)) = \psi_x(z)^{n_x}$ where $z \in W_x$.
Checking the holomorphy of transition functions is left as exercise. So we have made $X(\Gamma)$ a Riemann surface which also gives $Y(\Gamma)$ the structure of a Riemann surface by considering it as a subsurface.
Naturally, we should be able to link all these to certain modular functions.

Proposition 4.4. *The map $f \mapsto \pi \circ f/\mathfrak{h}$ defines a bijection between the meromorphic functions on $X(\Gamma)$ and modular functions of weight 0 and level Γ.*

Proof. The map $\pi|_{\mathfrak{h}} : \mathfrak{h} \to Y(\Gamma)$ is a morphism of Riemann surfaces. So if $f : X(\Gamma) \to \hat{\mathbb{C}}$ is a morphism, then $F = f \circ \pi|_{\mathfrak{h}} : \mathfrak{h} \to \hat{\mathbb{C}}$ is meromorphic on $\mathfrak{h}$. F is certainly weakly modular as it factors through $\pi|_{\mathfrak{h}}$, so it remains to check that F is meromorphic at the cusps. Let $\alpha \in \Gamma(1), \alpha\infty \in \mathbb{P}^1_{\mathbb{Q}}$. Since f is meromorphic, there exists a meromorphic function $g_\alpha : D \to \hat{\mathbb{C}}$ where D is an open disk centered at 0 such that $\forall \tau \in V_2, f(\pi(\alpha\tau)) = F(\alpha\tau) = F|_0[\alpha](\tau) = g(e^{2\pi i\tau/h}) = g(q_h)$. This exactly means that F is meromorphic at the cusp $\Gamma \cdot \alpha\infty$.
Conversely, suppose F is a modular function of weight 0 and level Γ. Then

$F : \mathfrak{h} \to \hat{\mathbb{C}}$ factors through $f : Y(\Gamma) \to \hat{\mathbb{C}}$. We need to show that f extends to $X(\Gamma)$ as a morphism. The previous computation (but done in reverse) reveals that f is meromorphic at the cusps. As for $\tau \in \mathfrak{h}$, we need to check that f is meromorphic in a chart $\phi_\tau(\pi(z)) = \psi_\tau(z)^{n_\tau}$ in a neighbourhood U_τ of a point $\pi(\tau)$. Explicitly, we need to show that the function $w = \psi_\tau(z)^{n_\tau} \mapsto F(z) = (F \circ \psi_\tau^{-1})(\psi_\tau(z))$ is meromorphic in D. $F \circ \psi_\tau^{-1}$ is meromorphic in an open disk D' centered at 0 and satisfies $F \circ \psi_\tau^{-1}(\zeta_{n_\tau} z) = F \circ \psi_\tau^{-1}(z)$ for every n_τ^{th} root of unity ζ_{n_τ}. The result then follows from Lemma 4.5 which will be stated shortly. □

Lemma 4.5. *If D' is a disk centered at 0, g is a meromorphic function in D' and $\forall z \in D', g(z) = g(e^{2\pi i/n} z)$, then there exists a meromorphic function h in D' such that $h(z^n) = g(z)$ for all $z \in D'$.*

Example 4.2. $j(\tau)$ is a modular function of weight 0 and level $\Gamma(1)$, so it defines a morphism $X(1) = X(\Gamma(1)) \to \hat{\mathbb{C}}$. This is bijective, so it is actually an isomorphism.

4.2 Genus and the Generalised Valence Formula

Definition 4.3. If $U \subset \mathbb{C}$ is an open subset, then a holomorphic differential is a symbol of the form $\omega = f(z)\,\mathrm{d}z$ where $f : U \to \mathbb{C}$ is a holomorphic function. We write $\Omega^1(U)$ for the set of holomorphic differentials in U.

If $g : U \to V$ is a holomorphic map between two open subsets of $\mathbb{C}$, then we write $g^* : \Omega^1(V) \to \Omega^1(U)$ for the map (the "pullback") given by $g^*(f(w)\,\mathrm{d}w) = f(g(z))g'(z)\,\mathrm{d}z$. If X is a Riemann surface, then a holomorphic differential on X is the data for every chart $\phi_i : U_i \to V_i$ of a holomorphic differential $\omega_i \in \Omega^1(V_i)$ such that for any i, j, the pullback of $\omega_j|_{\phi_j(U_i \cap U_j)} \in \Omega^1(\phi_j(U_i \cap U_j))$ under the map $\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$ is equal to $\omega_i|_{\phi_i(U_i \cap U_j)}$. Write $\Omega^1(X)$ for the $\mathbb{C}$-vectpr space of holomorphic differentials.

Theorem 4.6. *If X is compact, then $\Omega^1(X)$ is finite dimensional.*

Its dimension g_X is called the genus of X.

Example 4.3. $g_{X(1)} = g_{\hat{\mathbb{C}}} = 0$.

Definition 4.4. An elliptic point of Γ is a point $\pi(\tau)$ such that $\mathrm{Stab}_\Gamma(\tau)/(\Gamma \cap \{\pm 1\})$ is nontrivial.

If $\pi(\tau)$ is an elliptic point, then $\mathrm{Stab}_\Gamma(\tau)/(\Gamma \cap \{\pm 1\}) \hookrightarrow \mathrm{Stab}_{\Gamma(1)}(\tau)/\{\pm 1\}$. This means that τ is $\Gamma(1)$-conjugate to $\rho = e^{2\pi i/3}$ or i and the stabiliser of τ is isomorphic to a cyclic group of order 2 or 3. The order $n_\tau = n_{\pi(\tau)} = |\mathrm{Stab}_\Gamma(\tau)/(\Gamma \cap \{\pm 1\})|$ of it is called the period of the elliptic point.

Theorem 4.7 (Generalised Valence Formula)**.** *Suppose $\Gamma \leq \Gamma(1)$ is a congruence subgroup. Let ϵ_∞ be its number of cusps, ϵ_3 the number of elliptic points of period 3, and ϵ_2 the number of elliptic points of period 2, then*

$$g_{X(\Gamma)} = 1 + \frac{[\Gamma(1) : \Gamma\{\pm 1\}]}{12} - \frac{\epsilon_2}{4} - \frac{\epsilon_3}{3} - \frac{\epsilon_\infty}{2}$$

We'll prove this theorem using the Riemann-Hurwitz formula for Riemann surfaces. To state it, we need the notion of ramification. For a nonconstant morphism $f : X \to Y$ between compact Riemann surfaces and $x \in X$, choose charts $\phi_x : U_x \to V_x$ around X and $\phi_y : U_y \to V_y$ around $y = f(x)$ such that $\phi_x(x) = \phi_y(y) = 0$, then $\phi_y \circ f \circ \phi_x^{-1}$ is a holomorphic function around 0 that fixes 0, therefore takes the form $z^n g(z)$ for some $n \geq 1$ and g holomorphic and nonvanishing at 0.

Definition 4.5. This number n is the ramification index $e_x = e_{x,f}$ of f at x. The degree $\deg f$ is defined to be $\deg f = \sum_{x \in f^{-1}(y)} e_x$ for any $y \in Y$.

It is a fact from the study of Riemann surfaces that these notions are well-defined.

Theorem 4.8 (Riemann-Hurwitz)**.** *If $f : X \to Y$ is a nonconstant morphism of compact Riemann surfaces, then*

$$2 - 2g_X = (\deg f)(2 - 2g_Y) - \sum_{x \in X} (e_{x,f} - 1)$$

We shall prove our theorem by applying this formula to the morphism $X(\Gamma) \to X(1)$ given by the projection $\Gamma \cdot \tau \mapsto \Gamma(1) \cdot \tau$. The key thing to compute is, of course, the various ramification indices.
If $x = \pi(\tau)$ for some $\tau \in \mathfrak{h}$, then a chart around x is given by $\pi(z) \mapsto \psi_\tau(z)^{n_x}$ and a chart around $f(x)$ has the form $\psi_\tau(z)^{n_{f(x)}}$. In these charts, f has the form $w \mapsto w^{n_{f(x)}/n_x}$, hence $e_x = n_{f(x)}/n_x$. If $x = \pi(\alpha\infty)$ for some $\alpha \in \Gamma(1)$, then a chart around x is $\pi(\alpha\tau) \mapsto e^{2\pi i\tau/h}$ (where h is the width of the cusp $\alpha\infty$) for $\tau \in V_2$ and a chart around $f(x)$ is $\pi(\alpha\tau) \mapsto e^{2\pi i\tau} = (e^{2\pi i/h})^h$, so f has the form $w \mapsto w^h$, i.e. $e_x = h$.

Proof of Theorem 4.7. As $g_{X(1)} = 0$, we have

$$2 - 2g_{X(\Gamma)} = 2\deg f - \sum_{x \in X(\Gamma)} (e_x - 1)$$

Let's compute $\deg f$ first: Take $\tau = 2i$ (or literally any other point that's generic enough), then $\forall \gamma \in \Gamma(1), e_{\pi(\gamma\tau)} = 1$ as $n_{f(\pi(\gamma\tau))} = 1$. So $d = \deg f = |f^{-1}(f(\pi(2i)))| = |\Gamma\backslash(\Gamma(1) \cdot (2i))| = |\Gamma\backslash\Gamma(1)/\{\pm 1\}| = [\Gamma(1) : \Gamma\{\pm 1\}]$.
If $e_x > 1$, then either x is a cusp or $x \in Y(\Gamma)$ and $n_{f(x)} > 1$, in which case $f(x) = \Gamma(1) \cdot i$ or $\Gamma(1) \cdot \rho$. For the former, we have

$$\sum_{x \in f^{-1}(\Gamma(1) \cdot i)} (e_x - 1) = d - |f^{-1}(\Gamma(1) \cdot i)|$$

Let $a_i = |\{x \in f^{-1}(\Gamma(1) \cdot i) : e_x = 1\}|, b_i = |\{x \in f^{-1}(\Gamma(1) \cdot i) : e_x = 2\}|$, then $a_i + 2b_i = d, a_i + b_i = |f^{-1}(\Gamma(1) \cdot i)|$ and $a_i = \epsilon_2$. Substitution gives $d - |f^{-1}(\Gamma(1) \cdot i)| = (d - \epsilon_2)/2$.
Similarly, for ρ we have

$$\sum_{x \in f^{-1}(\Gamma(1) \cdot \rho)} (e_x - 1) = \frac{2}{3}(d - \epsilon_3), \quad \sum_{x \in f^{-1}(\Gamma(1) \cdot \infty)} (e_x - 1) = d - \epsilon_\infty$$

Putting these back to Riemann-Hurwitz delivers the result. □

Example 4.4. Let $\Gamma = \langle \Gamma(2), S \rangle$. We know that $\epsilon_\infty = 2$, $[\gamma(1) : \Gamma] = 3$ and $\epsilon_2 \geq 1$ (since $Si = i$). The formula gives $g_{X(\Gamma)} = (1/4) - (\epsilon_2/4) - (\epsilon_3/3) \implies g_{X(\Gamma)} = 0, \epsilon_2 = 1, \epsilon_3 = 0 \implies X(\Gamma) \cong \hat{\mathbb{C}}$.

4.3 Modular Functions as Differentials

Definition 4.6. Let $n \in \mathbb{Z}$ and $U \subset \mathbb{C}$ be open. A holomorphic (resp. meromorphic) n-differential is a symbol $f(z)(\mathrm{d}z)^{\otimes n}$ where $f : U \to \mathbb{C}$ is holomorphic (resp. meromorphic).
We write $\Omega^{\otimes n}(U)$ (resp. $\Omega_!^{\otimes n}(U)$) to denote the $\mathbb{C}$-vector space of holomorphic (resp. meromorphic) n-differentials in U.
If $g : U \to V$ is a holomorphic map between open subsets of $\mathbb{C}$, we define a pullback map $g^* : \Omega^{\otimes n}(V) \to \Omega^{\otimes n}(U)$ (resp. $g^* : \Omega_!^{\otimes n}(V) \to \Omega_!^{\otimes}(U)$) induced by g via $g^*(f(w)(\mathrm{d}w)^{\otimes n}) = f(g(z))(g'(z))^n(\mathrm{d}z)^{\otimes n}$.
If X is a Riemann surface, a holomorphic (resp. meromorphic) n-differential on X is the set of data $\omega_i \in \Omega^{\otimes n}(V_i)$ (resp. $\omega_i \in \Omega_!^{\otimes n}(V_i)$) for every chart $\phi_i : U_i \to V_i$ such that $\forall i, j, \phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \to \phi_j(U_i \cap U_j)$ satisfies $(\phi_j \circ \phi_i^{-1})^*(\omega_j|_{\phi_j(U_i \cap U_j)}) = \omega_i|_{\phi_i(U_i \cap U_j)}$.
We write $\Omega^{\otimes n}(X)$ (resp. $\Omega_!^{\otimes n}(X)$) for the $\mathbb{C}$-vector space of holomorphic (resp. meromorphic) n-differentials on X.

Proposition 4.9. *Let $n \in \mathbb{Z}$, then the map $\Omega_!^{\otimes n}(X(\Gamma)) \to \Omega_!^{\otimes n}(\mathfrak{h})$ given by $\omega \mapsto (\pi|_{\mathfrak{h}})^*(\omega) = f(\tau)(\mathrm{d}\tau)^{\otimes n}$ defines a bijection between $\omega \in \Omega_!^{\otimes n}(X(\Gamma))$ and the set of modular functions $f(\tau)$ of weight $k = 2n$ and level Γ.*

Proof. Let $\omega \in \Omega_!^{\otimes n}(X(\Gamma))$ and write $(\pi|_{\mathfrak{h}})^*(\omega) = f(\tau)(\mathrm{d}\tau)^{\otimes n}$. For any $\gamma \in \Gamma$, we have $(\pi|_{\mathfrak{h}} \circ \gamma) = \pi|_{\mathfrak{h}} \implies (\pi|_{\mathfrak{h}})^* = \gamma^* \circ (\pi|_{\mathfrak{h}})^*$. Therefore

$$f(\tau)(\mathrm{d}\tau)^{\otimes n} = \gamma^*(f(\tau)(\mathrm{d}\tau)^{\otimes n}) = f(\gamma\tau)\gamma'(\tau)^n(\mathrm{d}\tau)^{\otimes n} = f(\gamma\tau)j(\gamma,\tau)^{-2n}(\mathrm{d}\tau)^{\otimes n}$$

That is, f is invariant under the weight k action of Γ.
We also need to check that f is meromorphic at each cusp $\Gamma \cdot \alpha\infty, \alpha \in \Gamma(1)$. There exists a meromorphic function g on a disk D around $q_h = 0$ such that $\omega = g(q_h)(\mathrm{d}q_h)^{\otimes n}$ in the chart $\pi(\alpha V_2) \to D, \pi(\alpha\tau) \mapsto e^{2\pi i \tau/h}$. In αV_2, $\pi^*(\omega) = f(\tau)(\mathrm{d}\tau)^{\otimes n}$ is given by the pullback of $g(q_h)(\mathrm{d}q_h)^{\otimes n}$ under the map $\tau \mapsto e^{2\pi i \alpha^{-1}\tau/h}$. Expanding this relation gives

$$f(\tau) = g(e^{2\pi i \alpha^{-1}\tau/h})(2\pi i/h)j(\alpha^{-1},\tau)^{-2n}(e^{2\pi i \alpha^{-1}\tau/h})^n$$

for $\tau \in \alpha V_2$. So $f|_k[\alpha](\tau) = g(q_h)(2\pi i/h)q_h^n$ which is a meromorphic function of q_h, so indeed f is meromorphic at the cusp $\alpha\infty$.
The rest of the proof is left as exercise. □

Definition 4.7. Let X be a Riemann surface and $\omega \in \Omega_!^{\otimes n}(X)$. For $x \in X$, choose a chart $\phi_x : U_x \to V_x$ around x. Say ω is reprensented by $f(z)(\mathrm{d}z)^{\otimes n} \in \Omega_!^{\otimes n}(V_x)$ in this chart where $f : V_x \to \hat{\mathbb{C}}$ is a meromorphic function. The order $v_x(\omega)$ of ω at x is the order of f at $\phi_x(x)$.

One can check that this is independent of the choice of the chart.

Proposition 4.10. *Let $\omega \in \Omega_!^{\otimes n}(X(\Gamma))$ correspond to a modular function f of weight $k = 2n$ and level Γ, then:*
1. If $x = \pi(\alpha\infty)$, then $v_x(\omega) = v_\infty(f|_k[\alpha]) - n$.
2. If $x = \pi(\tau), \tau \in \mathfrak{h}$, then $n_x v_x(\omega) = v_\tau(f) - n(n_x - 1)$.

Proof. 1. From the proof of the preceding proposition we know that $f|_k[\alpha](\tau) = g(q_h)(2\pi i/h)q_h^n$ where $\omega = g(q_h)(\mathrm{d}q_h)^{\otimes n}$ near the cusp $\pi(\alpha\infty)$. This immediately gives $v_x(\omega) = v_x(g) = v_\infty(f|_k[\alpha]) - n$.
2. Take a chart $\phi_x : U_x \to V_x$ around x given by $\phi_x(\pi(z)) = \psi_\tau(z)^{n_x}$ (where $x = \pi(\tau)$). Say ω has the form $g(w)(\mathrm{d}w)^{\otimes n}$ in the chart for some $g : V_x \to \hat{\mathbb{C}}$ meromorphic, then by computing $(\pi|_{\mathfrak{h}})^*(\omega)$ we know that the modular form that ω corresponds to

$$f(z) = g(\psi_\tau(z)^{n_x})n_x\psi_\tau(z)^{n(n_x-1)}\psi'_\tau(z)^n$$

in a neighbourhood of τ. Hence $v_\tau(f) = n_x v_x(\omega) + n(n_x - 1)$. □

Corollary 4.11. *Take $n = 1$. The map $\omega \mapsto (\pi|_{\mathfrak{h}})^*(\omega) = f(\tau)\,\mathrm{d}\tau$ restricts to an isomorphism $\Omega^1(X(\Gamma)) \to S_2(\Gamma)$. In particular,* $\dim_{\mathbb{C}} S_2(\Gamma) = g_{X(\Gamma)}$.

Proof. If $\omega \in \Omega^1_!(X(\Gamma))$, then ω is holomorphic iff $v_x(\omega) \geq 0$ for all $x \in X(\Gamma)$. If f is a modular function, then f is a cuspidal modular form iff $\forall\alpha \in \Gamma(1), v_\infty(f|_k[\alpha]) \geq 1$ and $\forall\tau \in \mathfrak{h}, v_\tau(f) \geq 0$.
These conditions do match up by the proposition since $v_x(\omega) = v_\infty(f|_k[\alpha] - 1)$ if $x = \pi(\alpha\infty)$ and $v_\tau(f) = n_x v_x(\omega) + n_x - 1$ if $x = \pi(\tau)$. □

Can we do something in general? Needless to say, Hurwitz isn't the only mathematician who shared the credit of a theorem with Riemann.

Definition 4.8. Let X be a Riemann surface. A divisor on X is a formal sum $D = \sum_x n_x x$ where $n_x \in \mathbb{Z}$ for all x and $n_x = 0$ for all but finitely many x. The degree of the divisor D is $\deg D = \sum_{x\in X} n_x$. We define $\Omega^{\otimes n}(D)(X) = \{\omega \in \Omega^{\otimes n}_!(X) : \forall x \in X, v_x(\omega) + n_x \geq 0\} \leq \Omega^{\otimes n}_!(X)$.

Example 4.5. $\Omega^{\otimes 0}_!(x)$ is the space of meromorphic functions on X. So if $f \in \Omega^{\otimes 0}_!(x)$ and $n_x = 0$, then $v_x(f) + n_x \geq 0$ is equivalent to say that f is holomorphic at x; if $n_x > 0$, then $v_x(f) + n_x \geq 0$ is to say that f has a pole of order at most n_x at x; if $n_x < 0$, then $v_x(f) + n_x \geq 0$ is to say that f has a zero of order at least $-n_x$ at x.

Corollary 4.12. *For any n, there are isomorphisms of vector spaces $M_k(\Gamma) \cong \Omega^{\otimes n}(D_M)(X(\Gamma)), S_k(\Gamma) \cong \Omega^{\otimes n}(D_S)(X(\Gamma))$, where $k = 2n$ and*

$$D_M = \sum_{\Gamma\cdot z\in\Gamma\backslash\mathbb{P}^1_{\mathbb{Q}}} n(\Gamma\cdot z) + \sum_{x\in X(\Gamma)} \left\lfloor n\left(1 - \frac{1}{n_x}\right)\right\rfloor x$$

$$D_S = \sum_{\Gamma\cdot z\in\Gamma\backslash\mathbb{P}^1_{\mathbb{Q}}} (n-1)(\Gamma\cdot z) + \sum_{x\in X(\Gamma)} \left\lfloor n\left(1 - \frac{1}{n_x}\right)\right\rfloor x$$

Proof. Recall that if ω corresponds to f, then $v_\infty(f|_k[\alpha]) = v_{\pi(\alpha\infty)}(\omega) + n$ and $v_\tau(f) = n_x v_x(\omega) + n(n_x - 1)$. So $f \in M_k(\Gamma)$ iff $\forall\tau \in \mathfrak{h}, v_\tau(f) \geq 0, \forall\alpha \in \Gamma(1), v_\infty(f|_k[\alpha]) \geq 0$, which happens iff $\forall x \in Y(\Gamma), v_x(\omega) \geq -n(1 - 1/n_x), \forall z \in \mathbb{P}^1_{\mathbb{Q}}, v_{\Gamma\cdot z}(\omega) \geq -n$ which translates to the first isomorphism. The proof for $S_k(\Gamma)$ is analogous. □

This allows us to compute the dimensions of $M_{2n}(\Gamma)$ and $S_{2n}(\Gamma)$ if we know dimensions of some of these $\Omega^{\otimes n}(D)(X)$. The latter is, however, usually hard – but we have a strong enough theorem to make it work.

Theorem 4.13 (Riemann-Roch). *Suppose X be a compact Riemann surface and $n \in \mathbb{Z}$. Let D be a divisor on X such that* $\deg D + (2g-2)(n-1) > 0$*, then*

$$\dim_{\mathbb{C}} \Omega^{\otimes n}(D)(X) = \deg D + (2n-1)(g-1)$$

Let's try this on $M_{2n}(\Gamma)$ for $n \geq 1$. We need $C = \deg D_M + (2g-2)(n-1) > 0$ to apply Riemann-Roch in the form we stated it. Indeed,

$$\begin{aligned} C &= n\epsilon_\infty + \epsilon_2 \left\lfloor \frac{n}{2} \right\rfloor + \epsilon_3 \left\lfloor \frac{2n}{3} \right\rfloor + \left(\frac{d}{6} - \frac{\epsilon_2}{2} - \frac{2\epsilon_3}{3} - \epsilon_\infty \right)(n-1) \\ &= \epsilon_\infty + \epsilon_2 \left(\left\lfloor \frac{n}{2} \right\rfloor - \frac{n-1}{2} \right) + \epsilon_3 \left(\left\lfloor \frac{2n}{3} \right\rfloor - \frac{2(n-1)}{3} \right) + (n-1)\frac{d}{6} > 0 \end{aligned}$$

Riemann-Roch then gives

$$\dim_{\mathbb{C}} M_{2n}(\Gamma) = \dim_{\mathbb{C}} \Omega^{\otimes n}(D_M)(X(\Gamma)) = \deg D_M + (2n-1)(g-1)$$

As for $S_{2n}(\Gamma)$, the same computation gives $\deg D_S + (2g-2)(n-1) > 0$ when $n \geq 2$ (the case $n = 1$ has already been dealt with in Corollary 4.11), so we similarly have

$$\dim_{\mathbb{C}} S_{2n}(\Gamma) = \dim_{\mathbb{C}} \Omega^{\otimes n}(D_S)(X(\Gamma)) = \deg D_S + (2n-1)(g-1)$$

We conclude these in the following theorem:

Theorem 4.14. *If $n \geq 1$, then*

$$\dim_{\mathbb{C}} M_{2n}(\Gamma) = n\epsilon_\infty + \epsilon_2 \left\lfloor \frac{n}{2} \right\rfloor + \epsilon_3 \left\lfloor \frac{2n}{3} \right\rfloor + (2n-1)(g_{X(\Gamma)} - 1)$$

If $n \geq 2$, then

$$\dim_{\mathbb{C}} S_{2n}(\Gamma) = (n-1)\epsilon_\infty + \epsilon_2 \left\lfloor \frac{n}{2} \right\rfloor + \epsilon_3 \left\lfloor \frac{2n}{3} \right\rfloor + (2n-1)(g_{X(\Gamma)} - 1)$$

Example 4.6. 1. Take $\Gamma = \Gamma(1)$, then $g = 0, \epsilon_2 = \epsilon_3 = \epsilon_\infty = 1$, so

$$\dim_{\mathbb{C}} S_k(\Gamma(1)) = \left\lfloor \frac{k}{4} \right\rfloor + \left\lfloor \frac{k}{3} \right\rfloor - \frac{k}{2} = \begin{cases} \lfloor k/12 \rfloor - 1, \text{ if } k \equiv 2 \pmod{12} \\ \lfloor k/12 \rfloor, \text{ otherwise} \end{cases}$$

which is consistent with Theorem 1.4.
2. Take $\Gamma = \langle \Gamma(2), S \rangle$, then we have $\epsilon_\infty = 2, \epsilon_2 \geq 1$, therefore

$$g = \frac{1}{4} - \frac{\epsilon_2}{4} - \frac{\epsilon_3}{3} \implies g = 0, \epsilon_2 = 1, \epsilon_3 = 0$$

So if k is even and is at least 4, then $\dim_{\mathbb{C}} M_k(\Gamma) = 1 + \lfloor k/4 \rfloor$ and $\dim_{\mathbb{C}} S_k(\Gamma) = -1 + \lfloor k/4 \rfloor$. In particular, when $k = 4$, we have $\dim_{\mathbb{C}} M_4(\Gamma) = 2, \dim_{\mathbb{C}} S_4(\Gamma) = 0$ which implies that $M_4(\Gamma)$ is spanned by $E_4(\tau), E_4((\tau+1)/2)$ – a result we quoted a few sections ago.

4.4 A Formula of Ramanujan

Recall that we used the result $\dim_{\mathbb{C}} M_4(\Gamma) = 2$ for $\Gamma = \langle \Gamma(2), S\rangle$ to prove some properties of r_8. Naturally, knowing the explicit formulae of $\dim_{\mathbb{C}} M_k(\Gamma)$ and $\dim_{\mathbb{C}} S_k(\Gamma)$, we shall attempt to investigate $r_\ell(n)$ for $8 \mid n$. Let $k = \ell/2$, then we know that $\theta^\ell \in M_k(\Gamma)$ where as before $\theta(\tau) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 \tau}$. Also, $\theta^\ell = \sum_{n\geq 0} r_\ell(n) q_2^n = 1 + 3\ell q_2 + \cdots$ and θ^ℓ vanishes at the cusp $\gamma \cdot 1$.
We shall use these to get an approximate formula for $r_\ell(n)$. We know that

$$E_k(\tau) = 1 - \frac{2k}{B_k}\sum_{n\geq 1}\sigma_{k-1}(n)q^n \in M_k(\Gamma)$$

and that

$$E_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right](\tau) = \frac{1}{2}E_k\left(\frac{\tau+1}{2}\right) = \frac{1}{2}\left(1 - \frac{2k}{B_k}\sum_{n\geq 1}(-1)^n\sigma_{k-1}(n)q_2^n\right) \in M_k(\Gamma)$$

since $E_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right]$ is invariant under $\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)^{-1}\Gamma(1)\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)$ which contains Γ. It follows that

$$M_k(\Gamma) = S_k(\Gamma) \oplus \operatorname{span}\{E_k, E_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right]\}$$

We want to find $a, b \in \mathbb{Q}$ such that $\theta^\ell - aE_k - bE_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right] \in S_k(\Gamma)$. The constant term of the q-expansion of this expression at ∞ is $1 - a - b/2$. As for the cusp $\Gamma\cdot 1$, take $\alpha = \left(\begin{smallmatrix}1&-1\\1&0\end{smallmatrix}\right)$ (so that $\alpha\infty = 1$). $E_k|_k[\alpha] = E_k$ since $\alpha \in \Gamma(1)$, while

$$E_k|_k\left[\begin{pmatrix}1&1\\0&2\end{pmatrix}\right]|_k[\alpha] = E_k|_k\left[\begin{pmatrix}1&1\\0&2\end{pmatrix}\alpha\right] = E_k|_k\left[\begin{pmatrix}2&0\\0&1\end{pmatrix}\right] = 2^{k-1}E_{2\tau}$$

Hence $\theta^\ell - aE_k - bE_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right]$ vanishes at $\Gamma\cdot 1$ provided that $a + 2^{k-1}b = 0$. Gathering all these information, we have $a = 2^k/(2^k-1), b = 2/(1-2^k)$.

Theorem 4.15. *If $8|\ell$, then*

$$r_\ell(n) = \frac{2^k}{2^k-1}\frac{-2k}{B_k}\sigma_{k-1}\left(\frac{n}{2}\right) + \frac{1}{1-2^k}\frac{-2k}{B_k}(-1)^n\sigma_{k-1}(n) + O(n^{k/2})$$

Proof. Let $f = \theta^\ell - aE_k - bE_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right] \in S_k(\Gamma)$. The q-expansion of f is $f(\tau) = \sum_{n\geq 1} a_n q_2^n$ where $a_n = r_\ell(n) - aa_n(E_k) - ba_n\left(E_k|_k\left[\left(\begin{smallmatrix}1&1\\0&2\end{smallmatrix}\right)\right]\right)$. Using techniques from example sheet, one can show that $|a_n| = O(n^{k/2})$ which gives the formula. □

We now consider the case $\ell = 24$, in which case $k = 12, \dim_{\mathbb{C}} S_k(\Gamma) = 2$. Note that $\Delta = \sum_{n\geq 1}\tau(n)q^n \in S_{12}(\Gamma)$ and

$$\Delta|_k\left[\begin{pmatrix}1&1\\0&2\end{pmatrix}\right] = \frac{1}{2}\Delta\left(\frac{\tau+1}{2}\right) = \frac{1}{2}\sum_{n\geq 1}(-1)^n\tau(n)q_2^n \in S_{12}(\Gamma)$$

These two modular forms span $S_{12}(\Gamma)$, so we should be able to find $c, d \in \mathbb{Q}$ such that

$$\theta^\ell = aE_k + bE_k|_k\left[\begin{pmatrix}1&1\\0&2\end{pmatrix}\right] + c\Delta + d\Delta|_k\left[\begin{pmatrix}1&1\\0&2\end{pmatrix}\right]$$

We first compare the coefficients of the q-expansion of both sides at ∞. The left side is $1+48q_2+\cdots$ and the right side is $(a+b/2)+((b/2)(-2k/B_k)(-1)+(d/2)\tau(n)(-1))q_2+\cdots$. Then we look at the q-expansion of them after action by $\left(\begin{smallmatrix}1 & -1\\ 1 & 0\end{smallmatrix}\right)$ to deal with the cusp $\Gamma\cdot 1$. But $\theta^8|_k\left[\left(\begin{smallmatrix}1 & -1\\ 1 & 0\end{smallmatrix}\right)\right]$ vanishes at ∞, so $\theta^2 4|_k\left[\left(\begin{smallmatrix}1 & -1\\ 1 & 0\end{smallmatrix}\right)\right]$ vanishes to order at least 3 at ∞. The q-expansion of the right side after the action is, on the other hand,

$$aE_k(\tau)+2^{k-1}bE_k(2\tau)+c\Delta(\tau)+2^{k-1}d\Delta(2\tau)=(a+2^{k-1}b)+\left(a\frac{-2k}{B_k}+c\right)q+\cdots$$

Equating all these coefficients solves to

$$a=\frac{4096}{4095},b=-\frac{2}{4095},c=-\frac{65536}{691},d=-\frac{66304}{691}$$

In conclusion,

$$\theta^{24}=\frac{4096}{4095}E_{12}-\frac{2}{4095}E_{12}|_{12}\left[\begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}\right]-\frac{65536}{691}\Delta-\frac{66304}{691}\Delta|_{12}\left[\begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}\right]$$

Expanding the coefficients gives

Theorem 4.16 (Ramanujan)**.** *If $n\in\mathbb{N}$, we have*

$$r_{24}(n)=\frac{65536}{691}\sigma_{11}\left(\frac{n}{2}\right)-\frac{16}{691}(-1)^n\sigma_{11}(n)-\frac{65536}{691}\tau\left(\frac{n}{2}\right)-\frac{33152}{691}(-1)^n\tau(n)$$

Corollary 4.17. *If p is an odd prime, then*

$$r_{24}(p)=\frac{16}{691}(1+p^n)+\frac{33152}{691}\tau(p)$$

Recall that we stated Ramanujan's conjecture that $|\tau(p)|\leq 2p^{11/2}$ which was proved by Deligne in 1973. This implies the optimal estimate

$$\left|r_{24}(p)-\frac{16}{691}(1+p^{11})\right|\leq\frac{66304}{691}p^{11/2}$$

We also stated the Sato-Tate conjecture, proved by Barnet-Lamb, Geraghty, Harris and Taylor in 2011, which implies that the normalised error terms

$$\frac{r_{24}(p)-(16/691)(1+p^{11})}{(66304/691)p^{11/2}}\in[-1,1]$$

are equidistributed with respect to the Sato-Tate measure $(2/\pi)\sqrt{1-t^2}\,\mathrm{d}t$.