MODULAR FORMS

JAMES NEWTON

Contents

1. Introduction

1.1. **Basic notation.** The *modular group*, sometimes denoted $\Gamma(1)$, is

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

The *upper half plane* is $\mathcal{H} = \{\tau \in \mathbb{C} : \mathrm{Im}(\tau) > 0\}$. We can define an action of $\Gamma(1)$ on $\mathcal{H}$ as follows

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \tau = \frac{a\tau + b}{c\tau + d}.$$

Exercise 1. Check that this action preserves $\mathcal{H}$ and is a group action. Hint: first show that

$$\mathrm{Im}(\gamma \cdot \tau) = \frac{\mathrm{Im}(\tau)}{|c\tau + d|^2}.$$

Definition 1.1. Let k be an integer and Γ a finite index subgroup of $\Gamma(1)$. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k and level* Γ if

$$f(\gamma \cdot \tau) = (c\tau + d)^k f(\tau)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $\tau \in \mathcal{H}$.

Remark 1.2. Made more precise later: A function f being weakly modular of weight 0 and level Γ means it gives a meromorphic function on $\Gamma\backslash\mathcal{H}$. A function f being weakly modular of weight 2 means $f(\tau)d\tau$ gives a meromorphic differential on $\Gamma\backslash\mathcal{H}$.

Modular forms will be defined precisely in the next couple of lectures, but for now I will say that a weakly modular function of weight k and level Γ is a *modular form* (of weight k and level Γ) if it is holomorphic on $\mathcal{H}$ and satisfies some other condition.

When $\Gamma = \Gamma(1)$, this 'other condition' is that there exist constants $C, Y \in \mathbb{R}_{>0}$ with

$$|f(\tau)| \leq C$$

for all τ with $\mathrm{Im}(\tau) > Y$.

1.2. **Some motivating examples.**

- **Representation numbers for quadratic forms**

  For an integer $k \geq 1$ and $n \geq 0$ write $r_k(n)$ for the number of distinct ways of writing n as a sum of k squares, allowing zero and counting signs and orderings. For example, we have $r_2(1) = 4$ since $1 = 0^2 + 1^2 = 0^2 + (-1)^2 = 1^2 + 0^2 = (-1)^2 + 0^2$.

  Define a function θ on $\mathcal{H}$ by taking

$$\theta(\tau) = \sum_{-\infty}^{\infty} e^{2\pi i n^2 \tau}.$$

We write q for the variable $e^{2\pi i\tau}$. Then for a positive integer $k \geq 1$

$$\theta(\tau)^k = \sum_{n\geq 0} r_k(n)q^n.$$

It turns out that for even integers k, $\theta(\tau)^k$ is a modular form, and we will see later that one can obtain information about the function $r_k(n)$ using this. It allows you to write $r_k(n)$ as 'nice formula'+'error term'. For example,

$$r_4(n) = 8 \sum_{\substack{0<d|n \\ 4\nmid d}} d$$

(in this case there's no error term!).

- **Complex uniformisation of elliptic curves**

  If we have a lattice (rank two discrete subgroup) $\Lambda \subset \mathbb{C}$ the Weierstrass $\wp$-function $\wp(z, \lambda)$ is a holomorphic function $\mathbb{C}/\Lambda \to \mathbb{P}^1(\mathbb{C})$ and the map

$$z \mapsto (\wp(z, \Lambda), \wp'(z, \Lambda))$$

gives an isomorphism between $\mathbb{C}/\Lambda$ and complex points $E_\Lambda(\mathbb{C})$ of an elliptic curve E_Λ over $\mathbb{C}$ with equation

$$y^2 = 4x^3 - 60G_4(\Lambda) - 140G_6(\Lambda)$$

where

$$G_k(\Lambda) = \sum_{\omega\in\Lambda} \frac{1}{\omega^k}$$

— these are examples of modular forms (if we consider the function $\tau \mapsto G_k(\mathbb{Z}\tau \oplus \mathbb{Z})$). We write Λ_τ for the lattice $\mathbb{Z}\tau \oplus \mathbb{Z}$. Similarly, any homogeneous polynomial in G_4, G_6 is a modular form, for example the discriminant function

$$\tau \mapsto (60G_4(\Lambda_\tau))^3 - 27(140G_6(\Lambda_\tau))^2.$$

- **Dirichlet series and L-functions**

  Recall the Riemann zeta function

$$\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s}$$

(this is the definition for $\mathrm{Re}(s) > 2$). It has a meromorphic continuation to all of $\mathbb{C}$ and satisfies a functional equation relative $\zeta(s)$ and $\zeta(1-s)$. In the course we will prove Hecke's converse theorem: if we are given a set of complex numbers $\{a_n\}_{n\geq 1}$ such that the Dirichlet series

$$Z(s) = \sum_{n\geq 1} \frac{a_n}{n^s}$$

is absolutely convergent for $\mathrm{Re}(s) >> 0$, with suitable analytic continuation and functional equation, then the function

$$f(\tau) = \sum_{n\geq 1} a_n e^{2\pi i n\tau}$$

is a modular form.

Given $E/\mathbb{Q}$ an elliptic curve, the Hasse-Weil L-function of E, $L(E,s)$ is given by

$$\prod_p L_p(E,s) = \sum \frac{a_n}{n^s}$$

where for p a prime of good reduction (with E reducing to $\widetilde{E}_p$) $L_p(E,s) = (1 - a_p p^{-s} + p^{1-2s})^{-1}$, and $a_p = p + 1 - |\widetilde{E}_p(\mathbb{F}_p)|$ (and one can also write down the local factors at the primes of bad reduction). A very deep theorem (due to Wiles, Breuil, Conrad, Diamond and Taylor) is that $f(\tau) = \sum_{n\geq 1} a_n e^{2\pi i n\tau}$ is also a modular form. This is not proved using the converse theorem! Indeed, the only proof that $L(E,s)$ has analytic continuation and functional equation is to first show that $L(E,s)$ comes from a modular form in this way.

2. Modular forms of level one

In this section we will be interested in weakly modular functions of weight k and level $\Gamma(1)$.

2.1. Fourier expansions.

Note that the definition of a weakly modular function of level $\Gamma(1)$ implies that $f(\tau+1) = f(\tau)$ for all $\tau \in \mathcal{H}$. Suppose f is holomorphic on the region $\{\mathrm{Im}(\tau) > Y\}$ for some $Y \in \mathbb{R}$. Denote by D^* the punctured unit disc $\{q \in \mathbb{C} : 0 < |q| < 1\}$.

The map $\tau \mapsto e^{2\pi i\tau}$ defines a holomorphic, surjective, map from $\mathcal{H}$ to D^* and we can define a function F on D^* by $F(q) = f(\tau)$ where $\tau \in \mathcal{H}$ is something satisfying $q = e^{2\pi i\tau}$. The function F is well-defined since the value of $f(\tau)$ is independent of the choice of τ. Moreover, F is holomorphic on the region $\{0 < |q| < e^{-2\pi Y}\}$, since f is holomorphic on the corresponding region in $\mathcal{H}$ and we can define $F(q) = f(\frac{1}{2\pi i}\log(q))$ for appropriate branches of log on open subsets of D^*.

Therefore we obtain a Laurent series expansion $F(q) = \sum_{n\in\mathbb{Z}} a_n(f)q^n$, with $a_n(f) \in \mathbb{C}$. This is called the Fourier expansion, or q-expansion, of f.

2.2. Modular forms.

Definition 2.1. Suppose f is a weakly modular function of weight k and level $\Gamma(1)$.

- We say that f is *meromorphic* (resp. *holomorphic*) at ∞ if f is holomorphic for $\mathrm{Im}\tau >> 0$ and $a_n(f) = 0$ for all $n << 0$ (resp. for all $n < 0$). Equivalently, F extends to a meromorphic (resp. holomorphic) function on an open neighbourhood of 0 in the unit disc D.
- If f is meromorphic at ∞ we say that f is a *meromorphic form of weight k* (non-standard, but will need them later).
- If f is holomorphic on $\mathcal{H}$ and at ∞ we say that it is a *modular form of weight k*, and if moreover $a_0(f) = 0$ we say that it is a *cusp form.*

Lemma 2.2. *Suppose f is weakly modular of level $\Gamma(1)$. Then f is holomorphic at ∞ if and only if there exists $C, Y \in \mathbb{R}$ such that $|f(\tau)| \leq C$ for all τ with $\mathrm{Im}\tau > Y$.*

Proof. Suppose f is holomorphic at ∞. Then the function F extends to a holomorphic function on an open neighbourhood of 0 in D. Therefore F is bounded on some sufficiently small disc in D with centre 0. This implies the desired boundedness statement for f.

Conversely, suppose there exists $C, Y \in \mathbb{R}$ such that $|f(\tau)| \leq C$ for all τ with $\mathrm{Im}\tau > Y$. This implies that f is holomorphic for $\mathrm{Im}\tau > Y$ (since it is bounded and meromorphic) and so we get a holomorphic function F on the region $\{0 < |q| < e^{-2\pi Y}\}$ satisfying $F(e^{2\pi i\tau}) = f(\tau)$. The boundedness condition on f implies that $qF(q)$ tends to zero as q tends to 0, so F has a removable singularity at 0 and we are done. $\square$

Definition 2.3. We denote the set of modular forms of weight k by $M_k(\Gamma(1))$, and denote the set of cusp forms of weight k by $S_k(\Gamma(1))$ (sometimes M_k and S_k for short).

Exercise 2.
(1) M_k and S_k are $\mathbb{C}$-vector spaces (obvious addition and scalar multiplication)
(2) $f \in M_k$, $g \in M_l$, then $fg \in M_{k+l}$
(3) $f \in M_k \implies f(-\tau) = (-1)^k f(\tau)$, so k odd $\implies M_k = \{0\}$.

We will later show that M_k and S_k are finite dimensional and compute their dimensions (main goal of the first half of the course). One of the reasons for imposing the 'holomorphic at ∞' condition is to ensure these spaces are finite dimensional.

2.3. First examples of modular forms.

Definition 2.4. Let $k > 2$ be an even integer. Then the Eisenstein series of weight k is a function on $\mathcal{H}$, defined, for $\tau \in \mathcal{H}$, by

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

Here the $'$ on the summation sign tells us to omit the $(0, 0)$ term.

Lemma 2.5. *This sum is absolutely convergent for $\tau \in \mathcal{H}$, and converges uniformly on compact subsets of $\mathcal{H}$, hence G_k is a holomorphic function on $\mathcal{H}$.*

Proof. Let's fix a compact subset C of $\mathcal{H}$ and think of τ varying over this compact set. Consider the parallelogram P_1 in $\mathbb{C}$ with vertices $1 + \tau$, $1 - \tau$, $-1 - \tau$ and $-1 + \tau$. Denote by $D(\tau)$ the minimum absolute value of a point in the boundary of P_1 (i.e. the length of the shortest line joining the origin and the boundary of P_1). As τ varies over the compact set C, $D(\tau)$ attains a minimum, which we denote by r.

For $m \in \mathbb{Z}_{\geq 1}$ denote by P_m the parallelogram whose vertices are m times the vertices of P_1. It is clear that as τ varies over C the minimum absolute value of a point in the boundary of P_m is mr.

Now let's consider how many points of the lattice $\mathbb{Z} \oplus \mathbb{Z}\tau$ lie in each of the P_m. The parallelogram P_m contains a $(2m+1)\times(2m+1)$ grid of these lattice points, so the boundary of P_m contains $(2m + 1)^2 - (2m - 1)^2 = 8m$ lattice points.

For $M \in \mathbb{Z}_{\geq 1}$ let's consider the sum

$$\sum_{(c,d)\in S_M} \frac{1}{|c\tau + d|^k}$$

where S_M is the set of pairs of integers (c,d) such that $c\tau+d$ is in the boundary of P_m for some $m \geq M$.

For $\tau \in C$ we have

$$\begin{aligned}\sum_{(c,d)\in S_M} \frac{1}{|c\tau+d|^k} &\leq \sum_{m=M}^{\infty} \frac{8m}{(mr)^k} \\ &= \frac{8}{r^k}\sum_{m=M}^{\infty}\frac{1}{m^{k-1}}\end{aligned}$$

and if $k>2$ the final expression tends to 0 as M tends to ∞. Therefore the Eisenstein series is uniformly absolutely convergent for $\tau \in C$.

A slightly different proof of this Lemma is suggested in Exercise 1.1.4 of the book by Diamond and Shurman. ☐

Proposition 2.6. *The holomorphic function G_k is weakly modular of weight k.*

Proof. Let $\gamma = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in \Gamma(1)$. Now

$$G_k(\gamma\tau) = \sum_{(m,n)\in\mathbb{Z}^2}{}' \frac{(c\tau+d)^k}{(m(a\tau+b)+n(c\tau+d))^k} = (c\tau+d)^k \sum_{(m,n)\in\mathbb{Z}^2}{}' \frac{1}{((am+cn)\tau+(bm+dn))^k}.$$

Right multiplication by γ gives a bijection from $\mathbb{Z}^2 - \{0,0\}$ to itself. Therefore the last term in the displayed equation is equal to $(c\tau+d)^k G_k(\tau)$ as required. ☐

Proposition 2.7. *The q-expansion of G_k is*

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

where $\zeta(k) = \sum_{n\geq 1}\frac{1}{n^k}$ is the Riemann zeta function and

$$\sigma_{k-1}(n) = \sum_{\substack{m|n \\ m>0}} m^{k-1}.$$

In particular, G_k is a modular form of weight k (we just had holomorphy at ∞ left to show).

The proof of this is postponed to the end of this section.

Definition 2.8. A normalisation: $E_k(\tau) = G_k(\tau)/2\zeta(k)$.

Fact 2.9. Proved later: $\dim(M_8(\Gamma(1))) = 1$.

Application of this fact: E_4^2 and E_8 are both in M_8 and their q-exapnsions have the same constant term (namely 1), so they are equal.

Corollary 2.10. *We deduce from $E_4^2 = E_8$ that*

- $\zeta(4) = \pi^4/90$, $\zeta(8) = \pi^8/2\cdot 3^3\cdot 5^2\cdot 7$
- $\sigma_7(n) = \sigma_3(n) + 120\sum_{j=1}^{n-1}\sigma_3(j)\sigma_3(n-j)$.

Proof. For first part, compare a_1 and a_2 terms of q-expansions. Then compare general term to get the second part. $\square$

Some more interesting examples. A cusp form:

Definition 2.11. The Ramanujan delta function

$$\Delta(\tau) = \frac{E_4^3 - E_6^2}{1728} = q - 24q^2 + ... = \sum_{n \geq 1} \tau(n) q^n.$$

We will later see that $\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24}$. At any rate, from its definition we have $\Delta \in S_{12}(\Gamma(1))$.

Now I'll give the proof of Proposition 2.7. We will use Poisson summation:

Fact 2.12. Let $h : \mathbb{R} \to \mathbb{C}$ be a continuous function such that

- h is L^1, i.e. $\int_{-\infty}^{\infty} |h(x)| dx < \infty$.
- the sum $S(x) = \sum_{d \in \mathbb{Z}} h(x + d)$ converges absolutely and uniformly as x varies in a compact subset of $\mathbb{R}$, and $S(x)$ is an infinitely differentiable function in x.

Then, if we denote by $\hat{h}$ the Fourier transform

$$\hat{h}(t) = \int_{-\infty}^{\infty} h(x) e^{-2\pi i x t} dx$$

we have

$$\sum_{d \in \mathbb{Z}} h(x + d) = \sum_{m \in \mathbb{Z}} \hat{h}(m) e^{2\pi i m x}.$$

Idea of the proof: the sum $S(x)$ satisfies $S(x) = S(x + 1)$ and the right hand side of the final equality is the Fourier expansion for S.

Now let's apply this to the case we're interested in. We have

$$G_k(\tau) = 2 \sum_{d=1}^{\infty} \frac{1}{d^k} + 2 \sum_{c=1}^{\infty} \sum_{d \in \mathbb{Z}} \frac{1}{(c\tau + d)^k}.$$

For c and τ fixed, let's define

$$h_c(x) = \frac{1}{(c\tau + x)^k}.$$

Now we can compute $\sum_{d \in \mathbb{Z}} \frac{1}{(c\tau + d)^k}$ by applying Poisson summation to h_c. Exercise: check that h_c satisfies the conditions for Poisson summation.

We have

$$\hat{h}_c(m) = \int_{-\infty}^{\infty} \frac{e^{-2\pi i m x}}{(c\tau + x)^k} dx = \frac{1}{c^{k-1}} \int_{-\infty}^{\infty} \frac{e^{-2\pi i m c u}}{(\tau + u)^k} du$$

For the last equality we substitute $x = cu$.

To compute this integral we use Cauchy's residue theorem applied to the complex function $f_n(z) = \frac{e^{-2\pi inz}}{z^k}$. For $n \in \mathbb{Z}$, denote by I_n the integral of $f_n(z)$ along the horizontal line from $\tau - \infty$ to $\tau + \infty$. Then we have

$$\hat{h}_c(m) = \frac{e^{2\pi imc\tau}}{c^{k-1}} I_{cm}.$$

Lemma 2.13. *We have $I_n = 0$ for $n \leq 0$. For $n > 0$, we have $I_n = \frac{(-2\pi i)^k n^{k-1}}{(k-1)!}$.*

Proof. The residue of f_n at $z = 0$ is $\frac{1}{(k-1)!}(-2\pi in)^{k-1}$ (consider the Taylor expansion of $e^{-2\pi inz}$). So it follows that $-2\pi i \mathrm{Res}_{z=0} f_n(z) = \frac{(-2\pi i)^k n^{k-1}}{(k-1)!}$.

For $n \leq 0$ we integrate over rectangles with vertices at $\tau - C, \tau + C, \tau + C + iC, \tau - C + iC$, for $C \in \mathbb{R}$ tending to ∞. These integrals are all equal to zero, and it's easy to see that the integrals over the upper horizontal and vertical sides tend to zero.

For $n > 0$ integrate (clockwise) over rectangles with vertices at $\tau - C, \tau + C, \tau + C - iC, \tau - C - iC$. Now the integrals (for C large enough that the rectangle contains $z = 0$) are all equal to $\frac{(-2\pi i)^k n^{k-1}}{(k-1)!}$ and the integrals over three of the sides tend to zero. $\square$

So now we conclude that $\hat{h}_c(m) = 0$ for $m \leq 0$ and

$$\hat{h}_c(m) = \frac{(-2\pi i)^k m^{k-1}}{(k-1)!} e^{2\pi imc\tau}$$

for $m > 0$. Now applying Poisson summation, we can deduce Proposition 2.7 (recall that k is assumed even, so $(-2\pi i)^k = (2\pi i)^k$).

2.4. **Fundamental domain for $\Gamma(1)$.** We want to study some of the properties of the action of $\Gamma(1)$ on $\mathcal{H}$.

Definition 2.14. Suppose a group G acts continuously on a topological space X. Then a *fundamental domain* for G is an open subset $\mathscr{F} \subset X$ such that no two distinct points of $\mathscr{F}$ are equivalent under the action of G and every point $x \in X$ is equivalent (under G) to a point in the closure $\overline{\mathscr{F}}$.

Proposition 2.15. *The set*

$$\mathscr{F} = \{\tau \in \mathcal{H} : |\tau| > 1, |\mathrm{Re}(\tau)| < \frac{1}{2}\}$$

is a fundamental domain for the action of $\Gamma(1)$ on $\mathcal{H}$.

More precisely, if we set

$$\widetilde{\mathscr{F}} = \mathscr{F} \cup \{\tau \in \mathcal{H} : |\tau| \geq 1, \mathrm{Re}(\tau) = -\frac{1}{2}\} \cup \{\tau \in \mathcal{H} : |\tau| = 1, -\frac{1}{2} \leq \mathrm{Re}(\tau) \leq 0\}$$

then $\widetilde{\mathscr{F}}$ contains a unique representative for every $\Gamma(1)$-orbit.

The group $\Gamma(1)$ is generated by the elements

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

Proof. We let G be the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ generated by S and T. Fix $\tau \in \mathcal{H}$. For $\gamma \in G$ have $\mathrm{Im}(\gamma\tau) = \frac{\mathrm{Im}(\tau)}{|c\tau+d|^2}$. Since c and d are integers, $|c\tau + d|^2$ attains a minimum as c and d vary over possible bottom rows of matrices in G. Therefore $\mathrm{Im}(\gamma\tau)$ attains a maximum as γ varies over G. So there is a $\gamma_0 \in G$ with $\mathrm{Im}(\gamma_0\tau) \geq \mathrm{Im}(\gamma\tau)$ for all $\gamma \in G$.

In particular

$$\frac{\mathrm{Im}(\gamma_0\tau)}{|\gamma_0\tau|^2} = \mathrm{Im}(-\frac{1}{\gamma_0\tau}) = \mathrm{Im}(S\gamma_0\tau) \leq \mathrm{Im}(\gamma_0\tau)$$

which implies that $|\gamma_0\tau| \geq 1$. Since applying T does not change the imaginary part we have $|T^n\gamma_0\tau| \geq 1$ for all n, and for some n we have $\mathrm{Re}(T^n\gamma_0\tau)| \in [-1/2, 1/2)$.

If $T^n\gamma_0\tau \in \overline{\mathscr{F}}\backslash\widetilde{\mathscr{F}}$ then $ST^n\gamma_0\tau \in \widetilde{\mathscr{F}}$ so we have proven that every G-orbit has a representative in $\widetilde{\mathscr{F}}$. This immediately implies that every $\Gamma(1)$-orbit has a representative in $\widetilde{\mathscr{F}}$.

It remains to prove that every $\Gamma(1)$-orbit has a unique representative in $\widetilde{\mathscr{F}}$.

Suppose that we have two distinct but $\Gamma(1)$-equivalent points $\tau_1 \neq \tau_2 = \gamma\tau_1$ in $\widetilde{\mathscr{F}}$. Since both τ_i's have real part $< 1/2$ we have $\gamma \neq \pm T^n$, so $c \neq 0$. Moreover, $\mathrm{Im}\tau \geq \sqrt{3}/2$ for all $\tau \in \mathscr{F}$, so

$$\frac{\sqrt{3}}{2} \leq \mathrm{Im}(\tau_2) = \frac{\mathrm{Im}(\tau_1)}{|c\tau_1 + d|^2} \leq \frac{\mathrm{Im}(\tau_1)}{c^2\mathrm{Im}(\tau_1)^2} \leq \frac{2}{c^2\sqrt{3}},$$

which implies that $c = \pm 1$. So we have

$$\mathrm{Im}(\tau_2) = \frac{\mathrm{Im}(\tau_1)}{| \pm \tau_1 + d|^2}$$

but $|\pm\tau_1 + d| \geq |\tau_1| \geq 1$ which implies that $\mathrm{Im}(\tau_2) \leq \mathrm{Im}(\tau_1)$. But everything was symmetric between τ_1 and τ_2, so we have $\mathrm{Im}(\tau_2) \leq \mathrm{Im}(\tau_1)$ and $|\tau_1| = |\tau_2| = 1$. Since $\tau_1, \tau_2 \in \widetilde{\mathscr{F}}$ this implies that $\tau_1 = \tau_2$. So there are no distinct but $\Gamma(1)$-equivalent points in $\widetilde{\mathscr{F}}$ and $\widetilde{\mathscr{F}}$ indeed contains a unique representative of every $\Gamma(1)$-orbit.

Now we can deduce that $\Gamma(1) = G$. Let $\gamma \in \Gamma(1)$ and consider the action on $2i \in \mathscr{F}$. There exists a $g \in G$ such that $g\gamma(2i) \in \overline{\mathscr{F}}$. We write $g\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and observe that

$$\mathrm{Im}(g\gamma(2i)) = \frac{2}{4c^2 + d^2} \geq \frac{\sqrt{3}}{2}.$$

This implies that $c = 0$ and $d = \pm 1$, hence $g\gamma = \pm T^n$ for some integer n and therefore $\gamma \in G$. $\square$

Exercise 3. Suppose $\tau \in \mathcal{H}$ is such that $\gamma\tau = \gamma$ for $\gamma \in \Gamma(1)$ with $\gamma \neq \pm I$. Then τ is in the $\Gamma(1)$ orbit of i or $\omega = -1/2 + \sqrt{3}/2$. If we define $n_\tau = |\mathrm{Stab}_{\Gamma(1)/\pm I}(\tau)|$ then $n_\tau = 2, 3$ if τ is in the orbit of i, ω respectively.

Note that to do the above exercise, it is enough to compute the stabilisers for $\tau \in \widetilde{\mathscr{F}}$.

Definition 2.16. If τ is a point of $\mathcal{H}$ such that $\mathrm{Stab}_{\Gamma(1)/\pm I}(\tau)$ is non-trivial, we say that τ is an elliptic point, of order $n_\tau = |\mathrm{Stab}_{\Gamma(1)/\pm I}(\tau)|$.

Note that the value of n_τ only depends on the orbit $\Gamma(1)\tau$.

2.5. **Zeros and poles of meromorphic forms.**

Definition 2.17. If f is a weight k meromorphic form and $\tau \in \mathcal{H}$ we write $\mathrm{ord}_\tau(f)$ for the order of vanishing of f at τ (i.e. it is the order of the zero if f vanishes at τ, 0 if f is holomorphic and non-vanishing at τ and it is the negative of the order of the pole if f has a pole at τ).

We write $\mathrm{ord}_\infty(f)$ for the smallest n such that $a_n(f) \neq 0$, where $f(\tau) = \sum_{n\in\mathbb{Z}} a_n(f)q^n$.

Since f is weakly modular, the integer $\mathrm{ord}_\tau(f)$ depends only on the $\Gamma(1)$-orbit $\Gamma(1)\tau$.

Proposition 2.18. *Let f be a non-zero meromorphic form of weight k. Then*

$$\mathrm{ord}_\infty(f) + \sum_{\Gamma(1)\tau\in\Gamma(1)\backslash\mathcal{H}} \frac{1}{n_\tau}\mathrm{ord}_\tau(f) = \frac{k}{12}.$$

Note that the sum in the above has finitely many non-zero terms — fixing $Y \in \mathbb{R}_{>0}$, f has finitely many zeros and poles in the compact region $\tau \in \overline{\mathscr{F}} \cap \{\mathrm{Im}(\tau) \leq Y\}$, and by meromorphy at ∞ it has finitely many zeros and poles in the region $\tau \in \overline{\mathscr{F}} \cap \mathrm{Im}(\tau) \geq Y$.

Proof. We drew a complicated contour and integrated $f'(\tau)/f(\tau)$ around it. See Serre, 'A Course in Arithmetic', Chapter VII, Theorem 3. □

Here are some immediate consequences of Proposition 2.18. Recall the weight 12 cusp form

$$\Delta = \frac{E_4^3 - E_6^2}{1728}.$$

We can also define a meromorphic form of weight 0:

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

Corollary 2.19. *Δ is non-vanishing on $\mathcal{H}$. The weight 0 meromorphic form $j(\tau)$ is holomorphic on $\mathcal{H}$ and induces a bijection*

$$\Gamma(1)\backslash\mathcal{H} \to \mathbb{C}.$$

Proof. The q-expansion of Δ is $q - 24q^2 + \cdots$. In particular, we have $\mathrm{ord}_\infty(\Delta) = 1$. It follows immediately from Proposition 2.18 that Δ has no zeros in $\mathcal{H}$. This implies that j is indeed holomorphic on $\mathcal{H}$. To show that it induces a bijection

$$\Gamma(1)\backslash\mathcal{H} \to \mathbb{C},$$

fix $z \in \mathbb{C}$ and consider the weight 12 modular form $f_z(\tau) = E_4(\tau)^3 - z\Delta(\tau)$. By definition, we have $f_z(\tau) = 0$ if and only if $j(\tau) = z$. So to show that j induces a bijection it suffices to show that the zeros of f_z are given by a single $\Gamma(1)$-orbit.

By considering the q-expansion of f_z, we see that $\mathrm{ord}_\infty(f_z) = 0$. So Proposition 2.18 implies that we have an equality

$$\sum_{\Gamma(1)\tau\in\Gamma(1)\backslash\mathcal{H}} \frac{1}{n_\tau}\mathrm{ord}_\tau(f) = 1.$$

Now we see that the possibilities for the zeros of f_z are that there is a simple zero at a single non-elliptic orbit, a double zero at $\Gamma(1)i$ or a triple zero at $\Gamma(1)\omega$. In any case, the zeros form a single $\Gamma(1)$-orbit, as required. $\square$

2.6. **Dimension formula.**

Lemma 2.20. *$M_k = \{0\}$ for $k < 0$. $M_0 \cong \mathbb{C}$, and is given by constant functions.*

Proof. Suppose $k < 0$. Then if $f \in M_k$ is non-zero, Proposition 2.18 implies that f cannot be holomorphic on $\mathcal{H}$ and at ∞. So f is zero.

Suppose $f \in M_0$. Then the constant term in the q-expansion of f, $a_0(f)$, is also in M_0, so $g = f - a_0(f) \in S_0$. Applying Proposition 2.18 to g tells us that g is zero, so f is constant. $\square$

Lemma 2.21. *For even $k \geq 0$ we have*

$$\dim M_k(\Gamma(1)) \leq \lfloor k/12 \rfloor$$

if $k \equiv 2$ mod 12 and

$$\dim M_k(\Gamma(1)) \leq \lfloor k/12 \rfloor + 1$$

otherwise.

Proof. For general even k, we set $m = \lfloor k/12 \rfloor + 1$, and fix m distinct non-elliptic orbits $P_1, ..., P_m$ in $\Gamma(1)\backslash\mathcal{H}$. Suppose $f_1, ..., f_{m+1} \in M_k(\Gamma(1))$. Then we can find a linear combination of the f_i, denoted f, such that f has a zero at each of the m points P_i. Applying Proposition 2.18 implies that $f = 0$, so $\dim M_k \leq m$.

Now we suppose we are in the special case $k = 12l + 2$, $l \in \mathbb{Z}_{\geq 0}$. We now set $m = l$, choose l non-elliptic points as before, and suppose we have $l + 1$ elements $f_1, ..., f_{l+1}$ of $M_k(\Gamma(1))$. We denote by f a linear combination of the f_i with a zero at the l chosen points, therefore if f is non-zero we now have an equation

$$\mathrm{ord}_\infty(f) + l + \sum_{P \neq P_i} \frac{\mathrm{ord}_P(f)}{n_P} = l + \frac{1}{6}$$

which is impossible. So $f = 0$ and we conclude that $\dim M_k \leq l = \lfloor k/12 \rfloor$. $\square$

Exercise 4. Consider the graded ring $\bigoplus_{k\geq 0} M_k$. Show that this direct sum injects into the ring of holomorphic functions on $\mathcal{H}$. In other words, there are no non-trivial linear dependence relations between modular forms of different weights.

Theorem 2.22. *Let $R : \mathbb{C}[X, Y] \to M$ be the map given by sending X to E_4 and Y to E_6. Then R is an isomorphism of rings (and respects the graded, if we give X degree 4 and Y degree 6).*

Corollary 2.23. *The set*

$$\{E_4^a E_6^b : a, b \geq 0, 4a + 6b = k\}$$

is a basis for M_k.

$$\dim M_k = \lfloor k/12 \rfloor$$

if $k \equiv 2$ *mod* 12 *and*

$$\dim M_k = \lfloor k/12 \rfloor + 1$$

otherwise.

Proof. This is an exercise. □

Proof of Theorem 2.22. The proof of the above Corollary, together with Lemma 2.21 tells us that, appropriately graded, the degree k part of $\mathbb{C}[X, Y]$ has dimension $\geq$ the dimension of M_k. So it's enough to show that R is an injection.

In other words we must show that E_4 and E_6 are algebraically independent, considered as elements of the field of meromorphic functions on $\mathcal{H}$. It is enough to show that E_4^3 and E_6^2 are algebraically independent. Suppose there is a dependence relation

$$\sum_{a,b} \lambda_{a,b} E_4^{3a} E_6^{2b} = 0.$$

By considering parts of fixed degree we can assume that the sum is only over a, b with $3a + 2b$ fixed. Dividing by a suitable power of E_6, we see that E_4^3/E_6^2 is the root of a polynomial with coefficients in $\mathbb{C}$, which implies that E_4^3/E_6^2 is a constant function. This implies that $(E_6/E_4)^2$ is a constant multiple of E_4, but E_4 is holomorphic so this would imply that $E_6/E_4 \in M_2 = \{0\}$, which gives a contradiction. □

3. Modular forms for congruence subgroups

3.1. Definitions.

Definition 3.1. Suppose $N \in \mathbb{Z}_{\geq 1}$, then we define the *principal congruence subgroup* of level N

$$\Gamma(N) = \{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv Id \mod N\}.$$

Definition 3.2. A subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ is a *congruence* subgroup if $\Gamma(N) \subset \Gamma$ for some N.

It follows immediately that congruence subgroups have finite index in $\mathrm{SL}_2(\mathbb{Z})$ (the converse is false — c.f. congruence subgroup problem).

Definition 3.3. $\Gamma_0(N)$: upper triangular mod N $\Gamma_1(N)$: upper triangular with $1, 1$ on diagonal mod N

Recall that we already defined weak modularity with respect to a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$. For this course, we will only consider weak modularity with respect to congruence subgroups, although much of the theory goes through for any finite index subgroup.

Definition 3.4. Let k be an integer and Γ a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$. A meromorphic function $f : \mathcal{H} \to \mathbb{C}$ is *weakly modular of weight k and level* Γ if

$$f(\gamma \cdot \tau) = (c\tau + d)^k f(\tau)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ and $\tau \in \mathcal{H}$.

A fundamental domain for Γ acting on $\mathcal{H}$ can be obtained by taking a union of translates of $\mathscr{F}$ (the level one fundamental domain) by coset representatives for $\mathrm{SL}_2(\mathbb{Z})/\Gamma$. If you look at a picture of such a fundamental domain (e.g. use the applet at `http://www.math.lsu.edu/~verrill/fundomain/` written by Helena Verrill), then you'll see that there are (a finite number of) boundary points, known as cusps, on the real line (actually rational numbers). To get finite dimensional spaces of modular forms, we will need to impose conditions on the behaviour of weakly modular functions as τ approaches each of these limit points, as well as when τ has imaginary part going to ∞.

Definition 3.5. Suppose Γ is a congruence subgroup. We define the *period* of the cusp ∞ by

$$h(\Gamma) = \min\{h \in \mathbb{Z}_{>0} : \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \in \Gamma\}.$$

Definition 3.6. Suppose $f : \mathcal{H} \to \mathbb{C}$ weakly modular of weight k and level Γ, and that f is holomorphic for $\mathrm{Im}(\tau) >> 0$. Set $q_h = e^{2\pi i\tau/h}$, and define a function F on the punctured unit disc by

$$F(q_h) = f(\tau).$$

As before, F is holomorphic and has a Laurent series expansion

$$F(q_h) = \sum_{n\in\mathbb{Z}} a_n q_h^n.$$

We say that f is meromorphic (resp. holomorphic) at ∞ if F extends to a meromorphic (resp. holomorphic) function around 0 (i.e. if the appropriate condition holds on vanishing of the negative coeffients in the Laurent series for F).

Definition 3.7. The *slash operator* of weight k is defined as follows: for $\gamma \in \mathrm{GL}_2^+(\mathbb{R})$, $f : \mathcal{H} \to \mathbb{C}$ and $k \in \mathbb{Z}$, define $f|_{\gamma,k} : \mathcal{H} \to \mathbb{C}$ by $f|_{\gamma,k}(\tau) = (c\tau + d)^{-k} f(\gamma \cdot \tau)$.

Remark 3.8. If $f : \mathcal{H} \to \mathbb{C}$ is meromorphic and Γ is a congruence subgroup, then f is weakly modular of weight k and level Γ if and only if

$$f|_{\gamma,k} = f$$

for all $\gamma \in \Gamma$.

To show that f is weakly modular of weight k and level Γ, it suffices to show that $f|_{\gamma_i,k} = f$ for a set of generators $\gamma_1, ..., \gamma_n$ of Γ.

If $\alpha \in \mathrm{SL}_2(\mathbb{Z})$ and f is weakly modular of weight k and level Γ, then $f|_{\alpha,k}$ is weakly modular of weight k and level $\alpha^{-1}\Gamma\alpha$. Moreover, the function $f|_{\alpha,k}$ only depends on the coset $\Gamma\alpha \in \Gamma\backslash\mathrm{SL}_2(\mathbb{Z})$.

Definition 3.9. Let Γ be a congruence subgroup, $k \in \mathbb{Z}$ and f weakly modular of weight k and level Γ. We say that f is *meromorphic at the cusps* (resp. *holomorphic at the cusps*) if $f|_{\alpha,k}$ is meromorphic at ∞ (resp. holomorphic at ∞) for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$.

Definition 3.10. If f is weakly modular of weight k and level Γ and meromorphic at the cusps, we say that f is a *meromorphic form* of weight k and level Γ.

If f is weakly modular of weight k and level Γ, holomorphic on $\mathcal{H}$ and holomorphic at the cusps, we say that f is a *modular form* of weight k and level Γ. If moreover $a_0(f|_{\alpha,k}) = 0$ for all α we say that f is a cusp form.

We denote the space of modular forms of weight k and level Γ by $M_k(\Gamma)$, and denote the subspace of cusp forms by $S_k(\Gamma)$.

Here is a usual condition in practice for checking that a weakly modular function is a modular form:

Proposition 3.11. *If $\Gamma(N) \subset \Gamma$, f holomorphic on $\mathcal{H}$ and weakly modular of weight k and level Γ, with $f(\tau) = \sum_{n=0}^{\infty} a_n(f)e^{2\pi i\tau n/N}$ and $|a_n(f)| \leq Cn^r$ for some constants $C, r \in \mathbb{R}_{>0}$, then f is holomorphic at the cusps. Therefore f is a modular form of weight k and level Γ.*

Proof. See Diamond and Shurman Exercise 1.2.6 □

3.2. **Examples: θ-functions.** Recall the definition $\theta(\tau) = \sum_{n=-\infty}^{\infty} q^{n^2}$. It is straightforward to show that this series is absolutely and uniformly convergent on compact subsets of $\mathcal{H}$. We have

$$\theta(\tau, k) = \theta(\tau)^k = \sum_{n=0}^{\infty} r(n,k)q^n$$

where $r(n,k)$ is the number of ways of writing n as the sum of k squares.

Proposition 3.12. *We have $\theta(\tau+1) = \theta(\tau)$ and $\theta(-1/4\tau) = \sqrt{2\tau/i}\theta(\tau)$. Here by $\sqrt{\ }$ we mean the branch on $\mathrm{Re}(z) > 0$ extending the positive square root on the positive real axis.*

Proof. The first equality is clear. For the second we use Poisson summation. Set $h(x) = e^{-\pi t x^2}$ with $t \in \mathbb{R}_{>0}$. We have

$$\hat{h}(y) = \int_{-\infty}^{\infty} e^{-\pi t x^2 - 2\pi i x y} dx = e^{-\pi y^2/t} \int_{-\infty}^{\infty} e^{-\pi(\sqrt{t}x + iy/\sqrt{t})^2} dx.$$

We substitute $u = \sqrt{t}x + iy/\sqrt{t}$, use $\int_{-\infty}^{\infty} e^{-\pi u^2} du = 1$, and conclude that $\hat{h}(y) = e^{-\pi y^2/t}/\sqrt{t}$.

So Poisson summation tells us that

$$\sum_{d \in \mathbb{Z}} e^{-\pi t d^2} = \sum_{m \in \mathbb{Z}} e^{-\pi m^2/t}/\sqrt{t}$$

whence $\theta(it/2) = \frac{1}{\sqrt{t}}\theta(i/2t)$ for $t \in \mathbb{R} > 0$. Now uniqueness of analytic continuation implies that the conclusion of the Proposition. □

Corollary 3.13. $\theta(\tau/4\tau+1)^2 = (4\tau+1)\theta(\tau)^2$

Proof. Easy computation. □

We conclude that for even k $\theta(\tau/4\tau+1,k)=(4\tau+1)^{k/2}\theta(\tau,k)$. In particular, for positive integers k, $\theta(\tau,4k)$ is weakly modular of weight $2k$ and level Γ, where Γ is the subgroup of $\mathrm{SL}_2(\mathbb{Z})$ generated by

$$\pm\begin{pmatrix}1&1\\0&1\end{pmatrix},\pm\begin{pmatrix}1&0\\4&1\end{pmatrix}.$$

In the first example sheet, it is shown that $\Gamma=\Gamma_0(4)$, so $\theta(\tau,4k)\in M_{2k}(\Gamma_0(4))$.

3.3. **Examples: old forms.** It's worth noting that if $\Gamma\subset\Gamma'$ then if a function f is a modular form of weight k and level Γ' it is also a modular form of weight k and level Γ.

Lemma 3.14. *Suppose $f\in M_k(\Gamma_0(N))$ and $M\in\mathbb{Z}_{\geq 1}$. Then*

$$f_M:\tau\mapsto f(M\tau)$$

is in $M_k(\Gamma_0(MN))$.

Proof. First we check that f_M is weakly modular. We can write $f_M(\tau)=f|_{\gamma_M}(\tau)$, where

$$\gamma_M=\begin{pmatrix}M&0\\0&1\end{pmatrix}\in\mathrm{GL}_2^+(\mathbb{R}).$$

So f_M is weakly modular of weight k and level $(\gamma_M^{-1}\Gamma_0(N)\gamma_M)\cap\mathrm{SL}_2(\mathbb{Z})=\Gamma_0(MN)$.

To check holomorphy at the cusps, let

$$\alpha=\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z})$$

and let $\alpha'\in\mathrm{SL}_2(\mathbb{Z})$ be such that $\alpha'\infty=M\alpha\infty\in\mathbb{P}^1(\mathbb{Q})$ (i.e. $\alpha'\infty=\gamma_M\alpha\infty$). Observe that

$$f_M|_{\alpha,k}(\tau)=f|_{\alpha'\beta,k}(\tau)$$

where $\beta=\alpha'^{-1}\gamma_M\alpha\in\mathrm{GL}_2^+(\mathbb{Q})$ stabilises $\infty\in\mathbb{P}^1(\mathbb{Q})$. Hence β is an upper triangular matrix, and it is easy to see that $f|_{\alpha'\beta,k}$ has a holomorphic Fourier expansion (since $f|_{\alpha',k}$ does). ☐

Similarly, we have

Lemma 3.15. *Suppose $f\in M_k(\Gamma_1(N))$ and $M\in\mathbb{Z}_{\geq 1}$. Then*

$$f_M:\tau\mapsto f(M\tau)$$

is in $M_k(\Gamma_1(MN))$.

Definition 3.16. Fix $N\in\mathbb{Z}_{\geq 1}$. For each divisor $M\mid N$, let i_M be the map

$$\begin{aligned}i_M:M_k(\Gamma_1(N/M))\oplus M_k(\Gamma_1(N/M))&\to M_k(\Gamma_1(N))\\(f,g)&\mapsto f+g_M.\end{aligned}$$

We define the space of *oldforms* at level N, $M_k(\Gamma_1(N))^{\mathrm{old}}$ to be the span of the union of the images of i_M as $M>1$ varies over divisors of N.

3.4. **Examples: weight 2 Eisenstein series.** We would like to define an Eisenstein series of weight 2 by

$$G_2(\tau) = \sum_{d \neq 0} \frac{1}{d^2} + \sum_{c \neq 0} \sum_{d \in \mathbb{Z}} \frac{1}{(c\tau + d)^2}.$$

The above sum is *not* absolutely convergent, so we cannot interchange the order of summation over c and d (recall that this was used to prove the G_k was weakly modular for $k > 2$). However, the Poisson summation argument still allows us to compute the sums over d and conclude that this sum converges, and

$$G_2(\tau) = 2\zeta(2) + 2(2\pi i)^2 \sum_{n=1}^{\infty} \sigma_1(n) q^n$$

where this latter series is absolutely and uniformly convergent on compact subsets.

We use something known as 'Hecke's trick' to determine how G_2 transforms under the action of $\Gamma(1)$. This will then allow us to define some weight 2 modular forms of higher levels.

Definition 3.17. For $\epsilon \in \mathbb{R}_{>0}$ we define

$$G(\tau, \epsilon) = \sum_{(c,d) \in \mathbb{Z}^2}{}' \frac{1}{(c\tau + d)^2 |c\tau + d|^{2\epsilon}}.$$

The point is that we have perturbed $G_2(\tau)$ a little, to obtain a double sum which is now absolutely convergent. Now in exactly the same way as for the higher weight Eisenstein series we deduce

$$G(\gamma\tau, \epsilon) = (c\tau + d)^2 |c\tau + d|^{2\epsilon} G(\tau, \epsilon)$$

for $\gamma \in \Gamma(1)$.

Theorem 3.18 (non-examinable)**.** *For any $\tau \in \mathcal{H}$, the limit $\varinjlim_{\epsilon \to 0} G(\tau, \epsilon)$ exists and is equal to $G_2^*(\tau) = G_2(\tau) - \frac{\pi}{\mathrm{Im}(\tau)}$.*

As a consequence, we have $G_2^*(\gamma\tau) = (c\tau + d)^2 G_2^*(\tau)$, but note that G_2^* is not holomorphic. However, we can use G_2^* to get higher level modular forms.

Corollary 3.19. *For N a positive integer we let $G_2^{(N)}(\tau) = G_2(\tau) - NG_2(N\tau)$. Then $G_2^{(N)} \in M_2(\Gamma_0(N))$, and its q-expansion is given by*

$$2(1 - N)\zeta(2) - 8\pi^2 \sum_{n=1}^{\infty} \Big(\sum_{\substack{0 < d \mid n \\ N \nmid d}} d \Big) q^n.$$

Proof. We deduce that $G_2^{(N)}$ is weakly modular of weight 2 from the equality $G_{2,N} = G_2^* - N\iota_N(G_2^*)$ and our discussion of oldforms. To show that $G_2^{(N)}$ is holomorphic at the cusps we either show it directly or apply Proposition 3.11. The computation of the q-expansion is easily deduced from the q-expansion of G_2. ☐

Fact 3.20. The space $M_2(\Gamma_0(4))$ has dimension 2.

It follows from this fact that $G_2^{(2)}, G_2^{(4)}$ is a basis for $M_2(\Gamma_0(4))$, and by comparing q-expansions we see that

$$\theta(\tau, 4) = -\frac{1}{\pi^2} G_{2,4}(\tau)$$

and as a consequence we obtain:

Theorem 3.21. *For integers $n \geq 1$*

$$r(n, 4) = 8 \sum_{\substack{0<d|n \\ 4\nmid d}} d.$$

Finally, I should sketch the proof of Theorem 3.18. Unfortunately, it's a little painful...

Similarly to the higher weight case, we apply Poisson summation to the sums

$$\sum_{d \in \mathbb{Z}} \frac{1}{(c\tau + d)^2 |c\tau + d|^{2\epsilon}}.$$

We write $h_{c,\epsilon}(x) = (c\tau + x)^{-2} |c\tau + x|^{-2\epsilon}$ and then we have Fourier coefficients

$$\hat{h}_{c,\epsilon}(m) = \int_{-\infty}^{\infty} \frac{e^{-2\pi i m x}}{(c\tau + x)^2 |c\tau + x|^{2\epsilon}} dx = \frac{1}{c^{1+2\epsilon}} \int_{-\infty}^{\infty} \frac{e^{-2\pi i c m x}}{(\tau + x)^2 |\tau + x|^{2\epsilon}} dx$$

so we can write

$$G_2(\epsilon, \tau) = 2 \sum_{d=1}^{\infty} \frac{1}{d^{2+2\epsilon}} + 2 \sum_{c=1}^{\infty} \sum_{m \neq 0} \hat{h}_{c,\epsilon}(m) + 2 \sum_{c=1}^{\infty} \hat{h}_{c,\epsilon}(0).$$

The following Lemma tells us that the second of these sums is nice enough that we can compute its limit as $\epsilon \to 0$ by exchanging the limit and the summation:

Lemma 3.22. *Suppose $m \neq 0$ and $\epsilon < 1$. Then there exists a constant $K \in \mathbb{R}_{>0}$ (independent of ϵ, c, depending on τ) such that*

$$|\hat{h}_{c,\epsilon}(m)| \leq \frac{K}{c^{3+2\epsilon} m^2}.$$

Proof. It's enough to show that there exists K with

$$\left| \int_{-\infty}^{\infty} \frac{e^{-2\pi i c m x}}{(\tau + x)^2 |\tau + x|^{2\epsilon}} dx \right| \leq \frac{K}{c^2 m^2}.$$

This follows from observing that $|\tau + x|^{2\epsilon} \geq |\mathrm{Im}(\tau)|^{2\epsilon} \geq \min\{1, |\mathrm{Im}(\tau)|^2\}$ and then showing that

$$\left| \int_{-\infty}^{\infty} \frac{e^{-2\pi i c m x}}{(\tau + x)^2} dx \right| \leq \frac{K}{c^2 m^2}.$$

This last estimate is derived by integrating by parts twice — it comes down to the fact that

$$\int_{-\infty}^{\infty} \frac{1}{|\tau + x|^4} dx$$

is finite. □

So for the $m \neq 0$ terms we take the limit inside the sum and we can also interchange the limit with the integral defining $\hat{h}_c(\epsilon, m)$. To prove the theorem, it is now sufficient to show that

$$\varinjlim_{\epsilon \to 0} 2 \sum_{c=1}^{\infty} \hat{h}_c(\epsilon, 0) = -\frac{\pi}{\mathrm{Im}\tau}.$$

We do this as follows: first, translating the variable x and using $|x-i|^{2\epsilon} = (x+i)^\epsilon (x-i)^\epsilon$ (these powers are defined using the principal branch of the logarithm) we obtain

$$\hat{h}_c(\epsilon, 0) = \frac{1}{(c\mathrm{Im}\tau)^{1+2\epsilon}} \int_{-\infty}^{\infty} \frac{1}{(x+i)^{2+\epsilon}(x-i)^\epsilon} dx.$$

Integration by parts tells us that this integral is equal to

$$-\frac{\epsilon}{1+\epsilon} \int_{-\infty}^{\infty} \frac{1}{1+x^2}^{1+\epsilon} dx$$

so we have

$$\sum_{c=1}^{\infty} \hat{h}_c(\epsilon, 0) = -\frac{\zeta(1+2\epsilon)\epsilon}{1+\epsilon} \frac{1}{(\mathrm{Im}\tau)^{1+2\epsilon}} \int_{-\infty}^{\infty} \frac{1}{(1+x^2)^{1+\epsilon}} dx.$$

Since $\zeta(s)$ has a simple pole with residue 1 at $s = 1$ the first fraction tends to $1/2$ as $\epsilon \to 0$, whilst the integral tends to π. This gives us the desired result.

3.5. **Finite dimensionality.** Now we can give a cheap proof that $M_k(\Gamma)$ is finite dimensional for all congruence subgroups Γ. We won't determine the dimensions for a while, however!

Suppose Γ' is a normal subgroup of Γ, and denote the quotient group $\Gamma' \backslash \Gamma$ by G (I'm thinking of the elements as right cosets, hence the notation). We define a right action of G on $M_k(\Gamma')$ by setting $f^g = f|_{\gamma,k}$ for $g = \Gamma'\gamma \in G$. The action of g is well-defined (i.e. independent of the choice of coset representative γ).

Now we can see that $M_k(\Gamma) = M_k(\Gamma')^G$, where by ^{G} we mean the invariants under the action of G (i.e. a function in $M_k(\Gamma')$ is Γ-invariant under the slash operator if and only if it is G-invariant).

Lemma 3.23. *Suppose Γ' is a normal subgroup of Γ and $f \in M_k(\Gamma')$. Then there exist modular forms $h_i \in M_{ik}(\Gamma)$ for $i = 1, \ldots, [\Gamma : \Gamma']$ with*

$$f^n + h_1 f^{n-1} + \cdots + h_n = 0.$$

Proof. Consider the identity

$$\prod_{g \in G} (f - f^g) = 0.$$

Expanding out the product we get a monic polynomial in f with the coefficient of f^{n-i} given by a symmetric polynomial of degree i in the f^g. This coefficient is therefore in $M_{ik}(\Gamma')^G = M_{ik}(\Gamma)$. □

Lemma 3.24. (1) *for $k < 0$, $M_k(\Gamma) = 0$*
(2) *$M_0(\Gamma) = \mathbb{C}$ (the constant functions)*

Proof. Since Γ is a congruence subgroup, we have $\Gamma(N) \subset \Gamma$ for some N. Note that $\Gamma(N)$ is a normal subgroup of $\Gamma(1)$. To show the Lemma, it is enough to prove it for $\Gamma = \Gamma(N)$. Suppose $f \in M_k(\Gamma)$ and $k < 0$. Then Lemma 3.23 gives us some $h_i \in M_{ik}(\Gamma(1))$ which are all zero, since $ik < 0$. so we have $f^n = 0$ for some n. Hence $f = 0$.

If $k = 0$, then $h_i \in M_0(\Gamma(1)) = \mathbb{C}$ for all i, so f is a root of a polynomial with constant coefficients. Hence f is constant. □

To prove that the spaces $M_k(\Gamma)$ are finite dimensional we will use some commutative algebra. The key ingredient is the notion of integral extensions of rings.

Definition 3.25. Let A be a subring of B. An element $b \in B$ is said to be integral over A if b satisfies a monic polynomial $b^n + a_1 b^{n-1} + \cdots a_n = 0$ with coefficients in A.

The ring B is said to be integral over A if every element of B is integral over A.

The integral closure $\widetilde{A}$ of A in B is defined to be the set of elements of B which are integral over A.

Exercise 5. Show that b is integral over A if and only if there exists a ring C with $A \subset C \subset B$ and $b \in C$, such that C is a finitely generated A-module.

Deduce that $\widetilde{A}$ is a subring of B (i.e. the set $\widetilde{A}$ is closed under the ring operations).

Show that if we have an extension of rings $A \subset B$ with B integral over A and B a finitely generated A-algebra, then B is a finitely generated A-module.

Remark 3.26. It follows from Lemma 3.23 and the above exercise that if we set $A = \oplus_{k\geq 0} M_k(\Gamma)$ and $B = \oplus_{k\geq 0} M_k(\Gamma')$ then B is integral over A (we think of A as a subring of B via the natural inclusions $M_k(\Gamma) \subset M_k(\Gamma')$ for each k).

Here's an important general result in commutative algebra (due to E. Noether)

Theorem 3.27. *Let F be a field and let A be a finitely generated F-algebra. Suppose A is an integral domain and denote the field of fractions* Frac(A) *by K. Suppose L is a finite extension field of K and denote by $\widetilde{A}$ the integral closure of A in L. Then $\widetilde{A}$ is a finitely generated A-module (and is in particular a finitely generated F-algebra).*

Proof. See Corollary 13.13 in Eisenbud's book 'Commutative algebra...'. □

As a consequence, we obtain the following useful lemma:

Lemma 3.28. *Let F be a field, B a commutative F-algebra, and assume B is an integral domain. Let $A \subset B$ be a sub F-algebra and assume that B is integral over A and* Frac$(B)/$Frac(A) *is a finite extension of fields. Then B is a finitely generated F-algebra if and only if A is a finitely generated F-algebra.*

Proof. First we suppose that A is a finitely generated F-algebra. Then Theorem 3.27 implies that $\widetilde{A}$, the integral closure of A in Frac(B), is a finite generated A-module. We have $A \subset B \subset \widetilde{A}$.

Now A is Noetherian, so a submodule of a finitely generated A-module is finitely generated, hence B is a finitely generated A-module. Therefore B is a finitely generated F-algebra.

For the reverse implication, we now suppose that B is a finitely generated F-algebra. Pick generators $b_1, ..., b_n$ for B and pick monic polynomials with coefficients in A with b_i as a root. Let C be the finitely generated sub F-algebra of A generated by the coefficients of these polynomials. By construction B is integral over C and it is a finitely generated C-algebra (since it is a finitely generated F-algebra).

By the exercise above, we know B is a finitely generated C-module. So A is a submodule of a finitely generated C-module, and is hence itself a finitely generated C-module. Therefore A is a finitely generated F-algebra. □

Finally, we can give the desired result about finite dimensionality of spaces of modular forms.

Theorem 3.29. *Let Γ be a congruence subgroup. Then*
1. *for $k < 0$, $M_k(\Gamma) = 0$*
2. *$M_0(\Gamma) = \mathbb{C}$ (the constant functions)*
3. *$M(\Gamma) := \oplus_{k \geq 0} M_k(\Gamma)$ is a finitely generated $\mathbb{C}$-algebra*

Proof. Since Γ is a congruence subgroup, we have $\Gamma(N) \subset \Gamma$ for some N. Note that $\Gamma(N)$ is a normal subgroup of $\Gamma(1)$. Set $C = M(\Gamma)$, $B = M(\Gamma(N))$ and $A = M(\Gamma(1))$.

Denote by G the finite group $\Gamma(N)\backslash\Gamma(1)$ which acts on B, with invariants $B^G = A$ (we let G act on each graded piece $M_k(\Gamma(N))$ by the weight k slash operator we discussed earlier).

We can extend the G action to the fraction field $\mathrm{Frac}(B)$ by letting G act on the numerator and denominator of a fraction.

We first check that
$$\mathrm{Frac}(A) = (\mathrm{Frac}(B))^G .$$
This is because, if we write $x \in (\mathrm{Frac}(B))^G$ as $\frac{p}{q}$ with $p, q \in B$, then
$$x \prod_{g \in G} q^g \in B^G = A$$
so x is in $\mathrm{Frac}(A)$.

Now Artin's lemma (as in Galois theory) implies that $\mathrm{Frac}(B)/\mathrm{Frac}(A)$ is a finite extension of fields (indeed, it is Galois with Galois group G). This also implies that $\mathrm{Frac}(B)$ is a finite extension of $\mathrm{Frac}(C)$.

Now we can apply Lemma 3.28: we know that A is a finitely generated $\mathbb{C}$-algebra, so we deduce that B is a finitely generated $\mathbb{C}$-algebra. Applying Lemma 3.28 once more, we deduce that C is a finitely generated $\mathbb{C}$-algebra, as required. □

Corollary 3.30. *For $k \geq 0$, $M_k(\Gamma)$ is a finite dimensional $\mathbb{C}$-vector space.*

Proof. We have shown that $\oplus_{k \geq 0} M_k(\Gamma)$ is a finitely generated $\mathbb{C}$-algebra. By decomposing a generator into its components of fixed weight, we obtain a generating set whose elements are modular forms $f_1, \ldots, f_n$ of weights $k_1 \ldots, k_n$. This implies that $M_k(\Gamma)$ is spanned by the monomials

$$\{\prod_i f_i^{l_i} : l_i \geq 0, \sum_i k_i l_i = k\}$$

$\square$

4. Modular curves as Riemann surfaces

4.1. Recap on Riemann surfaces.

Definition 4.1. Suppose X is a Hausdorff topological space (*topological space* for short). A *complex chart* on X is a homeomorphism $\phi : U \to V$ from $U \subset X$ open to $V \subset \mathbb{C}$ open.

Two charts $\phi_i : U_i \to V_i$ are *compatible* if

$$\phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$$

is biholomorphic.

An *atlas* on X is a family

$$\mathcal{A} = \{\phi_i : U_i \to V_i : i \in I\}$$

of compatible charts, with $X = \cup_{i \in I} U_i$.

We define an equivalence relation on pairs of topological spaces and atlases by $(X, \mathcal{A}) \sim (X, \mathcal{A}')$ if every chart in $\mathcal{A}$ is compatible with every chart in $\mathcal{A}'$.

A Riemann surface is defined to be an equivalence class of pairs $(X, \mathcal{A}')$. We will usually work with connected Riemann surfaces.

For X a Riemann surface, a function $f : Y \to \mathbb{C}$ on an open subset $Y \subset X$ is defined to be *holomorphic* if for all charts (in some atlas) $\phi : U \to V$,

$$f \circ \phi^{-1} : \phi(U \cap Y) \to \mathbb{C}$$

is holomorphic. The set of holomorphic functions on Y is denoted by $\mathscr{O}_X(Y)$.

In fact the assignment $Y \mapsto \mathscr{O}_X(Y)$ determines the Riemann surface structure on the topological space X. We'll develop this viewpoint a little, as it's convenient for talking about quotients of Riemann surfaces.

Definition 4.2. Let X be a topological space. A presheaf (of Abelian groups) on X is a pair $(\mathscr{F}, \rho)$ comprising

- for any $U \subset X$ open, an Abelian group $\mathscr{F}(U)$
- for any $V \subset U \subset X$ open, a group homomorphism

  $$\rho_V^U : \mathscr{F}(U)\mathscr{F}(V)$$

  called 'restriction from U to V' such that $\rho_U^U = id$ and $\rho_W^V \circ \rho_V^U = \rho_W^U$ for $W \subset V \subset U$. For $f \in \mathscr{F}(U)$ we usually write $f|_V$ for $\rho_V^U(f) \in \mathscr{F}(V)$.

The group $\mathscr{F}(U)$ is called the sections of $\mathscr{F}$ on U. Examples of presheafs are given by $\mathscr{F}(U) =$ continuous functions from U to $\mathbb{C}$. Denote this presheaf by $\mathscr{O}_X^{cts}$. Also, for X a Riemann surface, we have the presheaf of holomorphic functions $\mathscr{F}(U) = \mathscr{O}_X(U)$.

Definition 4.3. A presheaf $\mathscr{F}$ on X is a *sheaf* if for every open $U \subset X$, and every covering family $\{U_i\}_{i\in I}$ of U (i.e. $U_i \subset U$ with $U = \cup_{i\in I} U_i$), we have

- if $f, g \in \mathscr{F}(U)$ and $f|_{U_i} = g|_{U_i}$ for all i, then $f = g$
- given $f_i \in \mathscr{F}(U_i), i \in I$ such that $f_i|_{U_i\cap U_j} = f_j|_{U_i\cap U_j}$ for all $i, j \in I$, then there exists a section $f \in \mathscr{F}(U)$ such that $f|_{U_i} = f_i$ for every $i \in I$.

The first point says sections are determined by local data (i.e. restriction to covers by small open sets), the second says that we can define a section on U by 'gluing' sections defined on an open cover. Note that these statements are obviously satisfied by any reasonable presheaf of functions.

In particular, it's easy to check that the presheaves $\mathscr{O}_X^{cts}$ and $\mathscr{O}_X$ defined above are in fact sheaves.

Definition 4.4. A $\mathbb{C}$-space is a Hausdorff topological space X, equipped with a sheaf $\mathscr{F}$ such that $\mathscr{F}(U)$ is a sub $\mathbb{C}$-algebra of $\mathscr{O}_X^{cts}$ for all U, and the restriction maps ρ_V^U are given by restriction of functions.

A morphism of $\mathbb{C}$-spaces $(X, \mathscr{F}) \to (Y, \mathscr{G})$ is a continuous map $f : X \to Y$ such that for all opens $V \subset Y$, $g \in \mathscr{G}(V)$, we have $g \circ f \in \mathscr{F}(f^{-1}(V))$.

For example, if $\phi : U \to V$ is a chart of a Riemann surface X, then $\phi : (U, \mathscr{O}_X|_U) \cong (V, \mathscr{O}_V)$. Here $\mathscr{O}_X|_U$ denotes the sheaf on U given by $\mathscr{O}_X|_U(U') = \mathscr{O}_X(U')$ for $U' \subset U$.

Definition 4.5. A *sheafy Riemann surface* is a $\mathbb{C}$-space $(X, \mathscr{F})$ such that there is an open cover $\cup_{i\in I} U_i = X$ and isomorphisms of $\mathbb{C}-$spaces

$$\phi_i : (U_i, \mathscr{F}|_{U_i}) \cong (V_i, \mathscr{O}_{V_i})$$

with $V_i \subset \mathbb{C}$ open.

Proposition 4.6. *The map sending a Riemann surface X to the sheafy Riemann surface $(X, \mathscr{O}_X)$ identifies Riemann surfaces with sheafy Riemann surfaces, and identifies holomorphic maps between Riemann surfaces with $\mathbb{C}$-space morphisms between their associated sheafy Riemann surfaces.*

Proof. The inverse map takes $(X, \mathscr{O}_X)$ to an atlas provided by the definition of a sheafy Riemann surface. Now just check everything: for example to check compatibility of the charts, we need to show that $\tau = \phi_2 \circ \phi_1^{-1} : \phi_1(U_1 \cap U_2) \to \phi_2(U_1 \cap U_2)$ is holomorphic, but we know that τ identifies the holomorphic functions on these two open subsets of $\mathbb{C}$, and in particular it's composition with the identity map is holomorphic, so τ is holomorphic. □

4.2. **Group actions on Riemann surfaces.** Let X be a Riemann surface, and G a group, with a homomorphism $r : G \to \mathrm{Aut}^{\mathrm{hol}}(X)$ (i.e. if $\gamma \in G$ then $r(\gamma)$ is a biholomorphic map from X to itself.

Definition 4.7. We say that the group G acts *properly* on X if for all compact subsets $A, B \subset X$ the set

$$\{\gamma \in G : r(\gamma)A \cap B \neq \emptyset\}$$

is finite.

In particular for each $x \in X$ the stabiliser G_x is finite.

Exercise 6. The group G acts properly if and only if the map $\alpha : G \times X \to X \times X$ taking (γ, x) to $x, r(\gamma)x$ is *proper*: i.e. when we give G the discrete topology and products the product topology, $\alpha^{-1}(K)$ is compact for any compact subset K of $X \times X$.

Lemma 4.8. *Suppose G acts properly on X. Then for each $x \in X$ there exists a connected open neighbourhood U_x of x with compact closure satisfying*

$$r(\gamma)U_x \cap U_x \neq \emptyset \iff r(\gamma)x = x.$$

Proof. First note that we can find a connected open neighbourhood U of x with compact closure such that $r(\gamma)U \cap U \neq \emptyset$ for only finitely many γ. We do this by taking $A = B$ to be (the pre-image under some chart of) a small closed ball around x in the definition of acting properly (U is then the interior of this closed ball). Let $g_1, ..., g_n$ enumerate the elements of G such that $r(\gamma)U \cap U \neq \emptyset$. We want to show that for each i with $g_i x \neq x$ we can find an open subset $x \in U_i \subset U$ such that $U_i \cap g_i U_i = \emptyset$. We will then set $U_x = \cap U_i$ (or the connected component of x in this intersection, if this intersection is disconnected). By the Hausdorff property of X (so U is also Hausdorff), if $g_i x \neq x$ we can find disjoint open neighbourhoods V_i, V_i' of $x, g_i x$ in U. Since G acts continuously on X we can find an open neighbourhood W_i of x in X such that $g_i W_i \subset V_i'$. We set $U_i = V_i \cap W_i$. Then U_i is disjoint from V_i', yet $g_i U_i \subset V_i'$, so $U_i \cap g_i U_i = \emptyset$. □

Lemma 4.9. *Suppose G acts properly on X. We topologise the set of orbits $G\backslash X$ by saying that a subset of $G\backslash X$ is open if and only if its preimage in X is open. With this definition, the maps $\pi : X \to G\backslash X$ is continuous and open, and the quotient topological space $G\backslash X$ is Hausdorff.*

In fact the above topology is the unique topology such that the quotient map $\pi : X \to G\backslash X$ is continuous and every continuous map of topological spaces $f : X \to Y$ satisfying $f(gx) = f(x)$ for all $g \in G$ factors uniquely (and continuously) through π.

Proof. First we show that π is open (it is obviously continuous). If $U \subset X$ is an open set, then $\pi^{-1}(\pi(U)) = \cup_{g\in G} gU$ is a union of open sets gU, hence it is open, so by definition $\pi(U)$ is open.

Now we show that $G\backslash X$ is Hausdorff. Let Gx, Gy be two distinct points of the quotient $G\backslash X$. Let K_x, K_y be two distinct compact neighbourhoods of x, y (say given by small closed balls with respect to some chart), and denote the interiors by U_x, U_y. We know that $gK_x \cap K_y \neq \emptyset$ only for g in a finite subset $G_0 \subset G$. By shrinking the neighbourhoods K_x, K_y we can assume that y is not in gK_x for any $g \in G_0$.

Let V_y be the open neighbourhood of y given by the intersection $U_y \cap (X \setminus \cup_{g\in G_0} gK_x)$. Now $gU_x \cap V_y = \emptyset$ for all $g \in G$, so $\pi(U_x)$ and $\pi(V_y)$ are disjoint open neighbourhoods of Gx and Gy. □

Note that we are just using the fact that X is a Hausdorff and *locally compact* topological space.

The next lemma tells everything we'll need to know about the structure of these quotient spaces.

Lemma 4.10. *Let G, X be as above, and let $x \in X$. Then there exists an open neighbourhood U_x of x (connected with compact closure) such that $gU_x = U_x$ for all $g \in G_x$ and satisfies*

$$\pi^{-1}(\pi(U_x)) = \coprod_{gG_x \in G/G_x} g(U_x).$$

Proof. First we take U a neighbourhood of x with compact closure such that $gU \cap U \neq \emptyset \iff g \in G_x$. Then we define U_x to be the connected component of x in $\cap_{g \in G_x} gU$ (each $g \in G_x$ maps a connected set containing x to a connected set containing x, so we have $gU_x = U_x$). $\square$

Now for U an open subset in $G\backslash X$, consider the set of holomorphic functions $\mathscr{O}_X(\pi^{-1}(U))$. This set has a natural right action of G, given by

$$f^g : x \mapsto f(gx).$$

If we consider the invariants under the G-action, $\mathscr{O}_X(\pi^{-1}(U))^G$, then we have a set of G-invariant continuous functions on X. By the definition of the quotient topology on $G\backslash X$, this set naturally embeds in $\mathscr{O}^{cts}_{G\backslash X}(U)$. This means that we can make the following definition:

Definition 4.11. We given $G\backslash X$ the structure of a $\mathbb{C}$-space by setting $\mathscr{O}_{G\backslash X}(U) = \mathscr{O}_X(\pi^{-1}(U))^G$ for U an open subset of $G\backslash X$.

It's easy to see that $\mathscr{O}_{G\backslash X}$ is a sheaf on the topological space $G\backslash X$.

Theorem 4.12. *The pair $(G\backslash X, \mathscr{O}_{G\backslash X})$ defines a Riemann surface. The map π is holomorphic, and for $x \in X$, there exist charts around $x, \pi(x)$ such that π is locally given by*

$$z \mapsto z^{n_x}$$

where $n_x = |r(G_x)|$ (moreover, $r(G_x)$ is cyclic of order n_x).

The Riemann surface structure we have defined on $G\backslash X$ satisfies the universal property that every holomorphic map of Riemann surfaces $f : X \to Y$ which satisfies $f(gx) = f(x)$ for all $g \in G$ factors uniquely (and holomorphically) through π.

Proof. We can immediately assume that G is a subgroup of $\mathrm{Aut}(X)$. Let $x \in X$. We are going to define a chart on a neighbourhood of $\pi(x)$. Denote by U the open neighbourhood (connected with compact closure) of x provided by Lemma 4.10. Possibly shrinking U, we can assume that U is biholomorphic to an open subset of $\mathbb{C}$. Denote by V the image $\pi(U)$. Since π is open, this an open neighbourhood of $\pi(x)$. Also, since $\pi^{-1}(V) = \coprod_{gG_x \in G/G_x} g(U)$, we have $\mathscr{O}_{G\backslash X}(V) = \mathscr{O}_X(U)^{G_x}$ (the datum of a G_x invariant function on U is equivalent to a G-invariant function on the disjoint union of the gU). This tells us that it is enough to consider the case where G is a finite group fixing $0 \in X \subset \mathbb{C}$, with X a connected open subset of $\mathbb{C}$ (recall we are interested in the local structure of $G\backslash X$ in a neighbourhood of $\pi(x)$, and here we have mapped x to 0).

Now we claim that there is a biholomorphic map f from a neighbourhood U of 0 in X to the open unit disc D such that $f(0) = 0$, $gU = U$ for all $g \in G$ and for every g,

$f^{-1} \circ g \circ f$ is given by a rotation $z \mapsto \zeta(g)z$ ($\zeta(g)$ a root of unity). In particular, the group G is isomorphic to $\mathbb{Z}/n_x\mathbb{Z}$.

Let's assume this claim for the moment. Then we are reduced to the case where X is the open unit disc and $G \cong \mathbb{Z}/n_x\mathbb{Z}$ acts via $i \cdot z \mapsto \zeta^i z$ with ζ a primitive n_xth root of unity. Now the chart $G\backslash X \to X$ sending Gz to z^{n_x} gives an isomorphism of $\mathbb{C}$-spaces.

Finally we prove the claim. The key point is that for a sufficiently small open disc D_ϵ, centred at 0, in X, the set gD_ϵ is *convex* for every $g \in G$. See the Lemma below for a proof of this.

Then the intersection $U = \cap_{g\in G} gD_\epsilon$ is convex, hence simply connected, and moreover $gU = U$ for all $g \in G$. Since U is simply connected (with compact closure), it is biholomorphic to D via a map sending 0 to 0, and now we use the fact that biholomorphic maps from the unit disc to itself, fixing a point, are given by rotations (the Schwarz lemma). □

Remark 4.13. The fact that this works for non-free group actions is special to one-dimensional complex manifolds. An alternative presentation of this material is given by Miranda III.3.

Lemma 4.14. *Let X be an open subset of $\mathbb{C}$, containing 0, and suppose that f is an automorphism of X with $f(0) = 0$. Then there is an $\epsilon \in \mathbb{R}_{>0}$ such that f maps every disc $D_r = \{|z| \leq r\}$ with $r < \epsilon$ onto a convex region (of course, we take ϵ small enough so that all the D_r are contained in X).*

Proof. See Farkas-Kra, III.7.7. The region $f(D_r)$ is convex if and only if the curves $C_r = \{f(z) : |z| = r\}$ are all convex. Suppose $\arg(z) + \arg(f'(z))$ is an increasing function of $\arg(z)$ on $\{|z| = r\}$. Then we claim that the curve C_r is convex — this is because the tangent to the curve C_r at $f(z)$ has direction

$$\frac{d}{d\theta} f(z) = izf'(z),$$

where $z = re^{i\theta}$.

So we compute the derivative of $\theta + \arg(f'(re^{i\theta})) = \theta + \mathrm{Re}(\log(-if'(re^{i\theta})))$ with respect to θ, and get $1 + \mathrm{Re}(zf''(z)/f'(z))$. Since $f'(z) \neq 0$ (as f is biholomorphic), this derivative is positive for $|z|$ small. □

Proposition 4.15. *The action of $\Gamma(1)$ on $\mathcal{H}$ is proper.*

Proof. Recall that $\mathrm{Im}(\gamma\tau) = \mathrm{Im}(\tau)/|c\tau + d|^2$. Suppose we have A, B compact subsets of $\mathcal{H}$. We are interested in the set G_0 of $\gamma \in \Gamma(1)$ such that $\gamma A \cap B \neq \emptyset$. Now, since B is compact, the set $\{\mathrm{Im}(\tau) : \tau \in B\}$ is contained in some compact interval $I = [c_1, c_2] \subset \mathbb{R}_{>0}$ (with $c_1 > 0$). So if $\gamma_\tau \in B$ then $\mathrm{Im}(\gamma\tau) \in I$ so we have inequalities

$$\mathrm{Im}(\tau)/c_2 \leq |c\tau + d|^2 \leq \mathrm{Im}(\tau)/c_1,$$

which imply that there are only finitely many possibilities for the integers c and d (since the real and imaginary parts of τ are also bounded). Suppose that two elements γ, δ of $\Gamma(1)$ have the same c, d. Then a computation shows that

$$\gamma\delta^{-1} = \pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

for some $n \in \mathbb{Z}$. Since A and B are compact, there are only finitely many possibilities for n. So we have shown that G_0 is finite. $\square$

Definition 4.16. For Γ a congruence subgroup, we denote by $Y(\Gamma)$ the Riemann surface obtained from the quotient $\Gamma\backslash\mathcal{H}$.

Corollary 4.17. *The map $j : Y(\Gamma(1)) \to \mathbb{C}$ is a biholomorphic map.*

Proof. We showed before that j is a bijection. It is holomorphic since it is induced by a $\Gamma(1)$-invariant holomorphic function on $\mathcal{H}$, and holomorphic bijections are biholomorphisms. $\square$

4.3. **Cusps and compactifications.**

Definition 4.18. Suppose Γ is a congruence subgroup. Then the set of *cusps* of $Y(\Gamma)$ is defined to be the set of orbits $C_\Gamma := \Gamma\backslash\mathbb{P}^1(\mathbb{Q})$ where the action of an element of $\mathrm{SL}_2(\mathbb{Z})$ on $(x : y) \in \mathbb{P}^1(\mathbb{Q})$ is given by

$$\gamma(x : y) = (ax + by : cx + dy).$$

We denote the cusp $\Gamma(1 : 0)$ by ∞. As usual, we think of $\mathbb{P}^1(\mathbb{Q})$ bijecting with $\mathbb{Q} \cup \{\infty\}$ by sending $(x : y)$ to x/y (or ∞ if $y = 0$).

For $s \in C_\Gamma$, we define the *width* of the cusp $s = \Gamma x$ to be the index of the subgroup $\{\pm I\}\Gamma_x$ in the stabiliser $\Gamma(1)_x$. We denote this positive integer by h_s (Exercise: this is independent of the choice of representative x for s).

For example, when $\Gamma = \Gamma(1)$ we have a single cusp, since the action of $\mathrm{SL}_2(\mathbb{Z})$ on $P^1(\mathbb{Q})$ is transitive.

For $\Gamma = \Gamma_0(p)$ there are two cusps, one of width 1 and the other of width p.

If we denote by $\mathcal{H}^*$ the disjoint union $\mathcal{H} \coprod \mathbb{P}^1(\mathbb{Q})$ then we define (first as a set) $X(\Gamma) = \Gamma\backslash cH^* = Y(\Gamma) \coprod C_\Gamma$.

Now we make $\mathcal{H}^*$ into a topological space. We will list a bunch of open sets, and take the topology generated by them. First we let the usual open sets in $\mathcal{H}$ be open in $\mathcal{H}^*$. The sets $U_A = \{\tau \in \mathcal{H} : \mathrm{Im}(\tau) > A\} \cup \{\infty\}$are also declared to be open: they are the preimages of the open discs centred at 0 under the map $\tau \mapsto e^{2\pi i\tau}$).

Finally, we declare to be open sets of the form gU_A for $g \in \Gamma(1)$ — these will be open neighbourhoods of the point $g(1 : 0) \in \mathbb{P}^1(\mathbb{Q})$, and they are regions bounded by circles touching the real line at $g(1 : 0)$ (if g does not stabilise $(1 : 0)$).

We define a topology on the quotient $X(\Gamma)$ as usual, by saying an set is open if and only if its preimage in $\mathcal{H}^*$ is open.

Lemma 4.19. *Let x be an element of $\mathbb{P}^1(\mathbb{Q})$. Then there exists an open neighbourhood U of x in $\mathcal{H}^*$ such that*

$$\Gamma_x := \{g \in \gamma : gx = x\} = \{g \in \Gamma : gU \cap U \neq \emptyset\}.$$

Proof. First we do this for $x = \infty$. Let $A \in \mathbb{R}_{>0}$. We have

$$gU_A = \{\tau \in \mathcal{H} : \mathrm{Im}(g^{-1}\tau) > A\} = \{\tau \in \mathcal{H} : \frac{\mathrm{Im}(\tau)}{|-c\tau + d|^2} > A\}.$$

Since $|-c\tau+d|^2 \geq c^2\mathrm{Im}(\tau)^2$, if $c \neq 0$ we have

$$\tau \in gU_A \implies \mathrm{Im}(\tau) > A\mathrm{Im}(\tau)^2 \implies \mathrm{Im}(\tau) < 1/A$$

so, for large enough A, U and gU_A are disjoint for all g with $c \neq 0$. Now $c = 0$ if and only if $g\infty = \infty$, so we are done.

Now for general x we fix g_0 with $g_0\infty = x$, and take $U = g_0U_A$ for A large enough (as in the above paragraph). ☐

Proposition 4.20. *Let Γ be a congruence subgroup, then the topological space $X(\Gamma)$ is connected, Hausdorff and compact.*

Proof. First we check that $\mathcal{H}^*$ is connected, since $\mathcal{H}$ is connected, and each element of $\mathbb{P}^1(\mathbb{Q})$ has a base of open neighbourhoods having non-trivial intersection with $\mathcal{H}$. It follows that the continuous image $X(\Gamma)$ of $\mathcal{H}^*$ is connected.

To show that $X(\Gamma)$ is Hausdorff, first recall that $X(\Gamma(1))$ is homeomorphic to the Riemann sphere (elementary way to see this is to stare at the funamental domain). For general Γ we know we can separate points of $Y(\Gamma)$. Suppose we have a cusp s and a point $y \in Y(\Gamma)$. The image of s in $X(\Gamma(1))$ is ∞ and the image of y in $X(\Gamma(1))$ is in $Y(\Gamma(1))$. Since the image of these points can be separated by open neighbourhoods, s and y can be separated by the pre-images of these opens. The fact that two cusps can be separated by open neighbourhoods follows from Lemma 4.19.

For compactness, first note that the extended fundamental domain $\overline{\mathscr{F}}^* = \overline{\mathscr{F}} \cup \{\infty\}$ is a compact subset of $\mathcal{H}^*$. Now $X(\Gamma)$ is a continuous image of the finite union of compact sets $\cup_{\Gamma\gamma\in\Gamma\backslash\Gamma(1)}\gamma\overline{\mathscr{F}}^*$, so it is compact. ☐

Lemma 4.21. *Let Γ be a congruence subgroup. There exist open neighbourhoods U_s of each cusp s in $X(\Gamma)$ such that the U_s are pairwise disjoint, and are all homeomorphic to the unit disc D, via maps sending s to 0 which are biholomorphisms from $U_s\backslash\{s\} \subset Y(\Gamma)$ to the punctured unit disc D^*.*

Proof. Let π be the quotient map $\mathcal{H}^* \to X(\Gamma)$. It follows from Lemma 4.19 that for A large enough, if we choose $g_s \in \Gamma(1)$ with $g_s\infty = x$ and $\Gamma x = s$ for each cusp s, then $U(g_s, A) = \pi(g_sU_A)$ gives a pairwise disjoint set of open neighbourhoods of the cusps.

Set $U^* = g_sU_A$ and denote by U the intersection $U^* \cap \mathcal{H}$. The natural inclusion $U^* \to \mathcal{H}^*$ induces a map $\Gamma_x\backslash U^* \to X(\Gamma)$. For A large enough, Lemma 4.19 tells us that this map is injective. Its image is the open neighbourhood $U(g_s, A)$ of the cusp s. Recall that a holomorphic function on $V = U(g_s, A)\backslash s$ is by definition a Γ-invariant holomorphic function on $\pi^{-1}(V)$ which is the same thing as a Γ_x-invariant holomorphic function on U.

Let h_s be the width of the cusp s. Then

$$\{\pm I\}\Gamma_x = \{\pm I\}g_s \begin{pmatrix} 1 & h_s\mathbb{Z} \\ 0 & 1 \end{pmatrix} g_s^{-1}$$

and the map sending $\tau \mapsto e^{2\pi i(g_s^{-1}\tau)/h_s}$ for $\tau \in \Gamma_x\backslash U$ and x to 0 sends $\Gamma_x\backslash U^*$ homeomorphically to an open disc.

Rescaling gives a homeomorphism from U^*/Γ_x to the unit disc, and so we get a homeomorphism from $U(g_s, A)$ to the unit disc. Restricting this map to $U(g_s, A)\backslash s$ gives a biholomorphism to the punctured unit disc, since it is a homeomorphism induced by a Γ_x-invariant holomorphic function on U. $\square$

Definition 4.22. We define a Riemann surface structure on $X(\Gamma)$ extending the Riemann surface structure on $Y(\Gamma)$ by adding the charts on the neighbourhoods of the cusps given by Lemma 4.21.

Since a continuous function on the unit disc which is holomorphic on the punctured unit disc is holomorphic everywhere, we can define the Riemann surface structure sheaf theoretically by saying that a continuous function on an open subset U of $X(\Gamma)$ is holomorphic if and only if its restriction to $U \cap Y(\Gamma)$ is holomorphic. Equivalently, we say such a function is holomorphic if and only if it defines a holomorphic function of D when we apply the homeomorphisms of Lemma 4.21.

Note that it follows from the proof of Lemma 4.21 that the map $X(\Gamma) \to X(\Gamma(1))$ has the form $z \mapsto z^{h_s}$ with respect to some charts around s and ∞.

5. Differentials and divisors on Riemann surfaces

5.1. Meromorphic differentials.

Definition 5.1. For $U \subset \mathbb{C}$ open, $n \in \mathbb{Z}_{>0}$ we define the space of meromorphic differentials of degree n on U by

$$\Omega^{\otimes n}(U) := \{f(z)dz^n : f \text{ meromorphic on } U\}.$$

For $\phi : U_1 \to U_2$ holomorphic, define

$$\phi^* : \Omega^{\otimes n}(U_2) \to \Omega^{\otimes n}(U_1)$$

by $\phi^*(f(z_2)(dz_2)^n) = f(\phi(z_1))(\phi'(z_1))^n(dz_1)^n$.

So $\Omega^{\otimes n}(U)$ is a $\mathbb{C}$-vector space, isomorphic to the space of meromorphic functions on U (but note that the pullback by ϕ^* of a differential is not the same as the pullback of a function).

Definition 5.2. Suppose X is a Riemann surface. Suppose we have two open subsets U_1, U_2 of X, with charts $\phi_i : U_i \cong D_i \subset \mathbb{C}$. Denote by τ_{ij} the transition functions $\phi_j \circ \phi_i^{-1} : \phi_i(U_i \cap U_j) \cong \phi_j(U_i \cap U_j)$. Then a meromorphic differential (of degree n) on X is a rule sending charts $\phi : U \to D$ on X to meromorphic differentials of $\omega(\phi)$ of degree n on D, such that for any two charts ϕ_1, ϕ_2 the differentials $\omega(\phi_1)$ and $\omega(\phi_2)$ are *compatible*: i.e. $\tau_{ij}^*(\omega(\phi_j)|_{\phi_j(U_i \cap U_j)}) = \omega(\phi_i)|_{\phi_i(U_i \cap U_j)}$ for $i, j \in \{1, 2\}$.

The set of differentials on X has an obvious structure of a $\mathbb{C}$-vector space.

Remark 5.3. By sending an open subset U of X to the $\mathbb{C}$-vector space of degree n meromorphic differentials on U, we can define a sheaf $\Omega_X^{\otimes n}$ on X. To check the sheaf property you need the following lemma:

Lemma 5.4. *Let X be a Riemann surface, and $\mathcal{A}$ an atlas on X. Suppose we have a collection of compatible meromorphic differentials for just the charts in $\mathcal{A}$. Then there exists a unique meromorphic differential on X extending the given meromorphic differentials on the charts.*

Proof. Given any chart $\phi : U \to D$ on X we can define a meromorphic differential on D by appropriate tranformations of the meromorphic differentials associated to charts in the atlas: if $\phi_i : U_i \to D_i$ is a chart then we get a biholomorphism $\phi(U \cap U_i) \cong \phi_i(U_i \cap U)$ and pulling back a differential on $\phi_i(U_i \cap U)$ gives a differential on $\phi(U \cap U_i)$. Doing this for all i gives a collection of compatible differentials on an open cover of D, which glue to the desired differential on D. □

It is now straightforward, given a holomorphic map of Riemann surfaces $\phi : X \to Y$ to define the pullback map $\phi^* : \Omega_Y^{\otimes n}(V) \to \Omega_X^{\otimes n}(\phi^{-1}(V))$ for any open $V \subset Y$.

5.2. **Meromorphic differentials and meromorphic forms.** In this section we let Γ be a congruence subgroup. Recall that we have a holomorphic map $\pi : \mathcal{H} \to Y(\Gamma) \hookrightarrow X(\Gamma)$.

Definition 5.5. Suppose ω is an element of $\Omega^{\otimes} k(X(\Gamma))$. We denote by f_ω the meromorphic function on $\mathcal{H}$ given by $\pi^*\omega = f_\omega(\tau)(d\tau)^k$.

Theorem 5.6. *Suppose $\omega \in \Omega^{\otimes k}(X(\Gamma))$. Then f_ω is a meromorphic form of weight $2k$ and level Γ. Moreover, the map $\omega \mapsto f_\omega$ is an isomorphism of $\mathbb{C}$-vector spaces from $\Omega^{\otimes k}(X(\Gamma))$ to the space of meromorphic forms of weight $2k$ and level Γ.*

Proof. First we check that f_ω is weakly modular of weight $2k$ and level Γ. Let $\gamma \in \Gamma$. We have a biholomorphism γ from $\mathcal{H}$ to $\mathcal{H}$ which descends to the identity map on $Y(\Gamma)$. Consider the meromorphic differential on $\mathcal{H}$ given by $\gamma^*\pi^*\omega$. On the one hand, this is equal to $(\pi \circ \gamma)^*(\omega) = \pi^*\omega$, since $\pi \circ \gamma = \pi$. On the other hand, we have

$$\gamma^*(f_\omega(\tau)(d\tau)^k) = f_\omega(\gamma\tau)\left(\frac{d\gamma\tau}{d\tau}\right)^k (d\tau)^k = f_\omega(\gamma\tau)(c\tau + d)^{-2k}(d\tau)^k$$

so we deduce that $f_\omega(\gamma\tau)(c\tau + d)^{-2k} = f(\tau)$ and f_ω is indeed weakly modular of weight $2k$ and level Γ.

Next we check that f_ω is meromorphic at the cusps. Let $\alpha \in \mathrm{SL}_2(\mathbb{Z})$. The map $\alpha : \mathcal{H}^* \to \mathcal{H}^*$ descends to a biholomorphism $\alpha : X(\alpha^{-1}\Gamma\alpha) \cong X(\Gamma)$. It follows from a calculation as above that $f_{\alpha^*\omega} = f_\omega|_{\alpha,k}$. So it suffices to show that f_ω is meromorphic at ∞. For large enough A, the image of $U_A \cup \{\infty\}$ in $X(\Gamma)$ is biholomorphic to an open disc via $\tau \mapsto e^{2\pi i\tau/h} = q_h$ so the fact that f_ω is meromorphic at ∞ follows immediately from the fact that ω is a meromorphic differential: if the differential on this chart in a neighbourhood of ∞ is $g(q_h)(dq_h)^k$ then f_ω satisfies

$$f_\omega(\tau) = \left(\frac{2\pi i q_h}{h}\right)^k g(q_h).$$

The map $\omega \mapsto f_\omega$ is clearly $\mathbb{C}$-linear. We show it is an isomorphism by writing down an inverse. For f a meromorphic form of weight $2k$ and level Γ we want to define a meromorphic differential $\omega(f)$ on $X(\Gamma)$ which pulls back to $f(\tau)(d\tau)^k$ on $\mathcal{H}$.

Let $x \in \mathcal{H}$. Then, from the proof of Theorem 4.12 we know that there are charts from neighbourhoods U of x and V of $\pi(x)$ to the unit disc such that π is the map $z \mapsto z^{n_x}$ in this coordinate and Γ_x acts via $z \mapsto \zeta^i z$, with ζ a primitive n_xth root of unity.

Since $f(\tau)(d\tau)^k$ defines a Γ_x-invariant meromorphic differential on U, in the new coordinate we have a meromorphic differential $g(z)(dz)^k$ on the open unit disc, such that $g(\zeta^i z)(d(\zeta^i z))^k = g(z)(dz)^k$ for all i. Therefore we have $g(\zeta^i z) = \zeta^{-ik} g(z)$ for all i, so the function $z^k g(z)$ is Γ_x-invariant and is equal to $h(z^{n_x})$ for a meromorphic function h on the open unit disc.

Now we define a meromorphic differential on the open unit disc by

$$\omega = (n_x z)^{-k} h(z)(dz)^k.$$

This pulls back under $z \mapsto z^{n_x}$ to

$$(n_x z^{n_x})^{-k} h(z^{n_x})(n_x z^{n_x - 1})^k (dz)^k = z^{-k} h(z^{n_x})(dz)^k = g(z)(dz)^k,$$

so we define $\omega(f)$ on the chart from V to the open unit disc to be given by ω — by construction, it pulls back to $f(\tau)(d\tau)^k$ on the neighbourhood U of x.

Finally, we need to define our meromorphic differential in neighbourhoods of the cusps. It is enough to consider the cusp $\Gamma\infty$, since for a general cusp $s = \Gamma x$ with $x = \alpha\infty$ we can define a meromorphic differential ω in a neighbourhood of s by taking the meromorphic differential $\omega(f|_{\alpha,2k})$ defined in a neighbourhood of $(\alpha^{-1}\Gamma\alpha)\infty$ in $X(\alpha^{-1}\Gamma\alpha)$ and pulling back by the biholomorphism $X(\Gamma) \cong X(\alpha^{-1}\Gamma\alpha)$.

Recall that associated to $f(\tau)$ we have a meromorphic function on the unit disc, extending the holomorphic function F on the punctured unit disc defined by $F(e^{2\pi i\tau/h}) = f(\tau)$. Here h is the width of the cusp ∞.

Recall that the chart of Lemma 4.21 is also given in terms of the biholomorphism $\tau \mapsto e^{2\pi} i\tau/h = q_h$ from U_A/Γ_∞ to an open disc. We define a meromorphic differential on this disc

$$\omega = \left(\frac{2\pi i q_h}{h}\right)^{-k} F(q_h)(dq_h)^k.$$

Then ω pulls back to the meromorphic differential $f(\tau)(d\tau)^k$ on U_A under the map $\tau \mapsto e^{2\pi} i\tau/h$.

We just have to check that the meromorphic differentials we have defined are all compatible. However, they all pull back to restrictions of the same meromorphic differential, $f(\tau)(d\tau)^k$ on $\mathcal{H}$, so this follows from the next lemma. □

Lemma 5.7. *Suppose $\pi : X \to Y$ is a non-constant morphism of Riemann surfaces, with Y connected. Then the map π^* is an injection from $\Omega^{\otimes n}(Y)$ to $\Omega^{\otimes n}(X)$.*

Proof. We can assume that X and Y are open subsets of $\mathbb{C}$, with Y connected. So we have $\omega = f(z)(dz)^{\otimes n}$ and $f(\pi(z))(\pi'(z))^n = 0$ for all $z \in X$. Since π is non-constant, the zeroes of π' are discrete, and so $f(\pi(z)) = 0$ on an open subset of X. Hence f is zero on an open subset of Y (since π is an open map by the open mapping theorem), and f is therefore zero. □

There is a natural definition of the order of vanishing of a meromorphic differential at a point:

Definition 5.8. Let X be a Riemann surface, and ω a meromorphic differential on X. Let $x \in X$ be a point. Then we define $v_x(\omega)$ to be the order of vanishing of $f(z)$ at $z = z_0$, where ω is given by $f(z)(dz)^n$ on a chart defined on a neighbourhood of x sending x to z_0.

It's an exercise to check that v_x is well-defined (i.e. it doesn't matter what chart we choose). Recall that for $f \in M_k(\Gamma(1))\backslash\{0\}$ we previously defined ord_x for $x \in X(\Gamma(1))$ — this was just the order of vanishing of f at $x \in Y(\Gamma(1))$, or the natural order of vanishing of f at ∞ defined in terms of its q-expansion.

Exercise 7. By considering the proof of Theorem 5.6, show that for $f \in M_{2k}(\Gamma(1))$ non-zero we have

$$\begin{aligned} v_\infty(\omega_f) &= \mathrm{ord}_\infty(f) - k \\ v_x(\omega_f) &= \frac{\mathrm{ord}_x(f) - k(n_x - 1)}{n_x} \end{aligned}$$

for $x \in Y(\Gamma)$.

Deduce that the equality

$$\mathrm{ord}_\infty(f) + \sum_{x \in Y(\Gamma)} \frac{1}{n_x} \mathrm{ord}_x(f) = \frac{2k}{12}$$

of Proposition 2.18 is equivalent to the equality

$$\sum_{x \in X(\Gamma)} v_x(\omega_f) = -2k.$$

The last equality is a statement that the degree of the divisor associated to ω_f is equal to $-2k = (2g-2)k$, where $g = 0$ is the genus of $X(\Gamma(1))$.

5.3. **Divisors.** We assume X is a compact connected Riemann surface throughout this section.

Definition 5.9. Recall the definition of the group of divisors on a Riemann surface X: it is the free Abelian group generated by the points of X, i.e. formal sums $\sum_{x \in X} a_x[x]$ with $a_x = 0$ for almost all x. A divisor D has a degree $\deg(D) = \sum a_x$, where $D = \sum_{x \in X} a_x[x]$.

We say a divisor is *effective* if $a_x \geq 0$ for all x, and write $D \geq 0$ if D is effective.

For f a non-zero meromorphic function on X we define the *divisor of* f to be

$$\mathrm{div}(f) = \sum_{x \in X} v_x(f)[x]$$

where v_x is the order of vanishing at x. Note that since X is compact this sum is finite.

Similarly, we define $\mathrm{div}(\omega)$ for a meromorphic differential ω.

For D a divisor on X, we also define a $\mathbb{C}$-vector space

$$L(D) = \{f \text{ a non-zero meromorphic function on } X : \mathrm{div}(f) + D \geq 0\} \cup \{0\}.$$

The vector space structure is given by scalar multiplication and addition of meromorphic functions.

In fact $L(D)$ is the group of sections of a sheaf on X.

Definition 5.10. For $D = \sum a_x[x]$ a divisor on X and U an open subset of X, denote by $D|_U$ the divisor on U given by $\sum_{x \in U} a_x[x]$ and define

$$\mathscr{O}_X(D)(U) = \{f \text{ a non-zero meromorphic function on } U : \mathrm{div}(f) + D|_U \geq 0\} \cup \{0\}.$$

Note that we're abusing notation a bit, since U is not necessarily compact, so $\mathrm{div}(f)$ might be an infinite formal sum.

We can also denote the sheaf of holomorphic differentials on X by Ω^1_X (it is the subsheaf of the meromorphic differentials of degree 1 given by demanding that in the local expressions $f(z)dz$, f is *holomorphic*).

Exercise 8. Supposing ω_0 is a non-zero meromorphic differential of degree 1 on X, then we have

$$\Omega^1_X(U) \cong \mathscr{O}_X(\mathrm{div}(\omega_0))(U)$$

via the map $\omega \mapsto \omega/\omega_0$.

Remark 5.11. Every compact Riemann surface has a non-constant meromorphic function. In fact, for the Riemann surfaces $X(\Gamma)$ this is easy to see: we have the function $X(\Gamma) \to X(\Gamma(1)) \cong \mathbb{P}^1$. We obtain a non-zero meromorphic differential by pulling back a non-zero meromorphic differential on $\mathbb{P}^1$.

The Riemann-Roch theorem:

Theorem 5.12. *Let X be a compact connected Riemann surface, and let D be a divisor on X. Denote by g the genus of X. Denote by K the divisor* $\mathrm{div}(\omega_0)$ *for a non-zero meromorphic differential of degree* 1 *on X. Then*

$$\dim L(D) - \dim L(K - D) = \deg(D) + 1 - g.$$

Remark 5.13. By Serre duality we can write the above equality as

$$\dim H^0(X, \mathscr{O}_X(D)) - \dim H^1(X, \mathscr{O}_X(D)) = \deg(D) + 1 - g$$

so this is an Euler characteristic formula for the cohomology of the sheaf $\mathscr{O}_X(D)$.

Corollary 5.14. *We have* $\dim H^0(X, \Omega^1_X) = g$ *and* $\deg(K) = 2g - 2$.

Proof. Set $D = 0$ and $D = K$. □

Lemma 5.15. *Suppose $\omega \in \Omega^{\otimes n}(X)$. Then* $\mathrm{div}(\omega)$ *has degree* $(2g - 2)n$.

Proof. Let ω_0 be a non-zero meromorphic differential of degree one. If ω_0 is locally given by $f(z)dz$ it is easy to check that the local expressions $f(z)^n(dz)^n$ define a non-zero meromorphic differential of degree n, which we denote by ω_0^n. Now ω/ω_0^n defines a meromorphic function, whose associated divisor has degree 0, so the degree of $\mathrm{div}(\omega)$ is n times the degree of $\mathrm{div}(\omega_0)$ which is $(2g - 2)n$. □

5.4. **The genus of modular curves.** We can compute the genus of modular curves using the Riemann-Hurwitz formula:

Theorem 5.16. *Suppose $f : X \to Y$ is a holomorphic map between compact connected Riemann surfaces. Then*

$$2 - 2g(X) = \deg(f)(2 - 2g(Y)) - \sum_{x \in X}(e_x - 1).$$

Here $g()$ denotes the genus and e_x denotes the ramification index of the map f at the point x (i.e. locally around x, f looks like $z \mapsto z^{e_x}$).

Our modular curves $X(\Gamma)$ come equipped with maps to $X(\Gamma(1)) \cong \mathbb{P}^1$, so we apply Riemann-Hurwitz to these maps.

Set $r_2 = |\{x \in Y(\Gamma) : n_x = 2\}|$ and $r_3 = |\{x \in Y(\Gamma) : n_x = 3\}|$. As before, we set r_∞ to be the number of cusps of $X(\Gamma)$.

Theorem 5.17. *We have*

$$g(X(\Gamma)) = 1 + \frac{[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]}{12} - \frac{r_2}{4} - \frac{r_3}{3} - \frac{r_\infty}{2}.$$

Proof. Let $f : X(\Gamma) \to X(\Gamma(1))$ be the natural map. First note that $\deg(f) = [\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}]$. We denote this integer by d for the rest of the proof. Set $g = g(X(\Gamma)$.

Now we compute some ramification indexes. Let $x \in Y(\Gamma)$ with image $f(x) \in Y(\Gamma(1))$. Recall that the quotient map $\mathcal{H} \to Y(\Gamma)$ looks like $z \mapsto z^{n_x}$ around a pre-image of x, whilst the map $\mathcal{H} \to Y(\Gamma(1))$ looks like $z \mapsto z^{n_{f(x)}}$. This implies that the map f looks like $z \mapsto z^{n_{f(x)}/n_x}$, so we have $e_x = n_{f(x)}/n_x$.

For a cusp s a similar computation shows that we have $e_s = h_s$, the width of s.

Now Riemann-Hurwitz says that

$$2 - 2g = 2d - \sum_{x \in X(\Gamma)}(e_x - 1) = 2d - \sum_{f(x)=i}(e_x - 1) - \sum_{f(x)=\omega}(e_x - 1) - \sum_{f(x)=\infty}(e_x - 1).$$

For $P = i$ or ω we have $\sum_{f(x)=P}(e_x - 1) = (n_P - 1)(|f^{-1}(P)| - r_{n_P})$. On the other hand, we have $d = \sum_{f(x)=P} e_x = n_P(|f^{-1}(P)| - r_{n_P}) + r_{n_P}$. So we deduce that

$$\sum_{f(x)=P}(e_x - 1) = \frac{n_P - 1}{n_P}(d - r_{n_P}).$$

Therefore we have

$$2 - 2g = 2d - \frac{1}{2}(d - r_2) - \frac{2}{3}(d - r_{n_3}) - d + r_\infty$$

which rearranges to give the desired result. □

Note that is follows from this result that

$$2g - 2 + \frac{r_2}{2} + \frac{2r_3}{3} + r_\infty = \frac{d}{6} > 0.$$

5.5. **Riemann-Roch and dimension formulae.** Now we have everything we need to compute the dimensions of the $\mathbb{C}$-vector spaces $M_{2k}(\Gamma)$.

Recall that we have proved that meromorphic forms of weight $2k$ and level Γ correspond to meromorphic differentials of degree k on $X(\Gamma)$. We want to identify the image of the subspace of modular forms.

Suppose f is a meromorphic form and $\omega(f)$ its associated differential. It follows just as in Exercise 7 that for $x \in Y(\Gamma)$ we have

$$v_x(\omega(f)) = \frac{\mathrm{ord}_x(f) - k(n_x - 1)}{n_x}$$

and for $s = \Gamma x$ a cusp of $X(\Gamma)$ with $\alpha x = \infty$ ($\alpha \in \Gamma(1)$) we have

$$v_s(\omega(f)) = \mathrm{ord}_\infty(f|_{\alpha,2k}) - k.$$

Here the order of vanishing of $f|_{\alpha,2k}$ at ∞ is in terms of the variable q_{h_s}, where h_s is the width of the cusp s. Note that $f|_{\alpha,2k}$ has a Fourier expansion in the variable q_{h_s} because its weight is even.

We deduce the following

Proposition 5.18. *Suppose f is a meromorphic form of weight $2k$ and level Γ, with associated differential $\omega(f)$. Then f is a modular form if and only if $n_x v_x(\omega(f)) + k(n_x - 1) \geq 0$ for all $x \in Y(\Gamma)$ and $v_s(\omega(f)) + k \geq 0$ for all cusps s.*

The modular form f is cuspidal if and only if we moreover have $v_s(\omega(f)) + k - 1 \geq 0$ for all cusps s.

Note that the first condition is equivalent to

$$v_x(\omega(f)) + \lfloor \frac{k(n_x - 1)}{n_x} \rfloor \geq 0.$$

Definition 5.19. Let ω_0 be a non-zero differential on $X(\Gamma)$, set $K = \mathrm{div}(\omega_0)$ and define divisors

$$D(k) = kK + k \sum_{s \in X(\Gamma) \setminus Y(\Gamma)} [s] + \sum_{x \in Y(\Gamma)} \lfloor \frac{k(n_x - 1)}{n_x} [x]$$

and

$$D_c(k) = kK + (k-1) \sum_{s \in X(\Gamma) \setminus Y(\Gamma)} [s] + \sum_{x \in Y(\Gamma)} \lfloor \frac{k(n_x - 1)}{n_x} [x]$$

on $X(\Gamma)$.

Theorem 5.20. *We have isomorphisms $M_{2k}(\Gamma) \cong L(D(k))$ and $S_{2k}(\Gamma) \cong L(D_c(k))$ given by $f \mapsto \omega(f)/\omega_0^k$.*

Proof. This follows immediately from Proposition 5.18. $\square$

Corollary 5.21. *Denote by r_∞ the number of cusps of $X(\Gamma)$, and denote by g the genus of $X(\Gamma)$. Then we have, for $k \geq 1$*

$$\dim M_{2k}(\Gamma) = kr_\infty + \sum_{x \in Y(\Gamma)} \lfloor \frac{k(n_x - 1)}{n_x} \rfloor + (2k-1)(g-1)$$

and for $k \geq 2$

$$\dim S_{2k}(\Gamma) = (k-1)r_\infty + \sum_{x \in Y(\Gamma)} \lfloor \frac{k(n_x - 1)}{n_x} \rfloor + (2k-1)(g-1).$$

We have $\dim M_0(\Gamma) = 1$*,* $\dim S_0(\Gamma) = 0$ *and* $\dim S_2(\Gamma) = g$*.*

Proof. This all follows from Riemann-Roch and the fact that a holomorphic function on a compact connected Riemann surface is constant. We also use the observation that

$$2g - 2 + \frac{r_2}{2} + \frac{2r_3}{3} + r_\infty = \frac{d}{6} > 0.$$

□

Recall that we used the fact that $\dim(M_2(\Gamma_0(4))) = 2$ a few lectures ago. We can now prove this:

Corollary 5.22. *Let* $k \geq 0$*. We have* $\dim M_{2k}(\Gamma_0(4)) = k + 1$*.*

Proof. We have $[\Gamma(1) : \Gamma_0(4)] = 6$, so $g = 1 + \frac{1}{2} - \frac{r_2}{4} - \frac{r_3}{3} - \frac{r_\infty}{2}$. We also have two forms $G_2^{(2)}$ and $G_2^{(4)}$ which give linearly independent elements of the quotient vector space $M_2(\Gamma_0(4))/S_2(\Gamma_0(4))$ (see this by looking at the constant terms of their q-expansions at the cusps), so $\dim M_2(\Gamma) - \dim S_2(\Gamma) = r_\infty - 1 \geq 2$.

So $g \leq -\frac{r_2}{4} - \frac{r_3}{3}$, which implies that $g = r_2 = r_3 = 0$ and $r_\infty = 3$. Now apply the dimension formula.

In fact, it's probably easier just to determine the cusps of $X(\Gamma_0(4))$, which immediately gives $r_\infty = 3$... □

6. Hecke operators

In this section we will define Hecke operators on spaces of modular forms of level $\Gamma_1(N)$. To give the cleanest description, we begin by giving an alternative description of modular forms in terms of functions on lattices.

6.1. Modular forms and functions on lattices.

Definition 6.1. A *lattice* in $\mathbb{C}$ is a $\mathbb{Z}$-module $L \subset \mathbb{C}$ generated by two elements of $\mathbb{C}$ which are linearly independent over $\mathbb{R}$.

For $N > 1$ a $\Gamma_1(N)$-*level structure* on a lattice $L \subset \mathbb{C}$ is a point $t \in \mathbb{C}/L$ of exact order N (i.e. a point of the elliptic curve $\mathbb{C}/L$ of exact order N).

Denote the set $\{(L, t)\}$ comprising pairs of lattices with a $\Gamma_1(N)$-level structure by $\mathcal{L}_N$.

Suppose $k \in \mathbb{Z}$ and F is a function from $\mathcal{L}_N$ to $\mathbb{C}$. We say that F has *weight* k if $F(\lambda L, \lambda t) = \lambda^{-k} F(L, t)$ for all $(L, t) \in \mathcal{L}_N$ and $\lambda \in \mathbb{C}^\times$.

Remark 6.2. For example, the function $G_k(L) = \sum_{0 \neq l \in L} l^{-k}$ for $k > 2$ even is a function of weight k on $\mathcal{L}_1$. Note that $G_k(\mathbb{Z}\tau \oplus \mathbb{Z}) = G_k(\tau)$ where $G_k(\tau)$ is previously defined usual Eisenstein series.

Denote by M the set of pairs $\omega = (\omega_1, \omega_2)$ of elements of $\mathbb{C}^\times$ such that $\omega_1/\omega_2 \in \mathcal{H}$. To such a pair we can associate a lattice $L(\omega_1, \omega_2) = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ with $\Gamma_1(N)$-level structure $t(\omega_1, \omega_2) = \omega_2/N + L(\omega_1, \omega_2)$. This defines a surjective map from M to $\mathcal{L}_N$. The group $\mathrm{GL}_2^+(\mathbb{R})$ acts on M by sending (ω_1, ω_2) to $(a\omega_1 + b\omega_2, c\omega_1 + d\omega_2)$.

Lemma 6.3. *The map $M \to \mathcal{L}_N$ identifies $\mathcal{L}_N$ with the quotient of M by the action of $\Gamma_1(N)$.*

Proof. We first check that the map is surjective. Suppose $(L, t) \in \mathcal{L}_N$.

Let ω_1', ω_2' be any basis for L. We have

$$t = \frac{1}{N}(a\omega_1' + b\omega_2') + L$$

where $\gcd(a, b, N) = 1$. We can then find a', b', congruent to $a, b \bmod N$, such that a', b' are coprime. Now set $\omega_2 = a'\omega_1' + b'\omega_2'$. Since a' and b' are coprime, we have a basis ω_1, ω_2 for L, and $t = \frac{\omega_2}{N} + L$. If ω_1/ω_2 is not in $\mathcal{H}$, replace ω_1 with $-\omega_1$.

To show that this map identifies $\mathcal{L}_N$ with the quotient of M by $\Gamma_1(N)$, suppose we have two elements $(\omega_1, \omega_2), (\omega_1', \omega_2')$ of M with the same image in $\mathcal{L}_N$. In particular, we have $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ with

$$\gamma \cdot \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} = \begin{pmatrix} \omega_1' \\ \omega_2' \end{pmatrix}$$

the two elements of M span the same lattice L. Moreover, since $\omega_2'/N = \omega_2/N \bmod L$, the matrix γ lies in $\Gamma_1(N)$. $\square$

We also have an action of $\lambda \in \mathbb{C}^\times$ on M by mapping (ω_1, ω_2) to $(\lambda\omega_1, \lambda\omega_2)$, so we can define a notion of weight k for complex functions on M. The quotient of M by this action can be identified with $\mathcal{H}$ via the map $(\omega_1, \omega_2) \mapsto \omega_1/\omega_2$. This identifies $\mathcal{L}_N/\mathbb{C}^\times$ with the quotient $\Gamma_1(N)\backslash\mathcal{H}$.

The left action of $\mathrm{GL}_2^+(\mathbb{R})$ on M induces a right action on functions on M, by setting $\tilde{F} \cdot \gamma(\omega) = \tilde{F}(\gamma\omega)$.

On making the above observations, there is a natural way to pass between functions on $M, \mathcal{L}_N$ and $\mathcal{H}$. Given $F : \mathcal{L}_N \to \mathbb{C}$ of weight k we first define $\tilde{F} : M \to \mathbb{C}$ by $\tilde{F}(\omega) = F(L(\omega), t(\omega))$. Then we define $f(\tau) = \tilde{F}(\tau, 1)$ for $\tau \in \mathcal{H}$.

Proposition 6.4. *Let $k \in \mathbb{Z}$. The above association of F with $\tilde{F}$ and f gives a bijective correspondence between the following sets of complex-valued functions:*

1. *functions $F : \mathcal{L}_N \to \mathbb{C}$ of weight k*
2. *functions $\tilde{F} : M \to \mathbb{C}$ of weight k which are invariant under the action of* $\Gamma_1(N)$
3. *functions $f : \mathcal{H} \to \mathbb{C}$ which are invariant under the slash operator $|_{\gamma,k}$ for* $\gamma \in \Gamma_1(N)$

Proof. Exercise. $\square$

Now we say that a function F on $\mathcal{L}_N$ of weight k is weakly modular/a modular form/a cusp form if the associated function f on $\mathcal{H}$ is.

6.2. **Hecke operators.**

Definition 6.5. Suppose F is a function $\mathcal{L}_N \to \mathbb{C}$ and $n \in \mathbb{Z}_{\geq 1}$. Then we define a function T_nF by

$$T_nF(L,t) = \frac{1}{n}\sum_{L'} F(L',t)$$

where the sum is over lattices $L' \supset L$ with index $[L' : L] = n$ such that $t + L'$ is a point of exact order N in $\mathbb{C}/L'$.

For n coprime to N, we also define

$$T_{n,n}F(L,t) = \frac{1}{n^2}F(\frac{1}{n}L,t).$$

Proposition 6.6. *We have the following identities:*

(1) *if m and n are coprime, then $T_m \circ T_n = T_{mn}$*
(2) *if p is prime and divides N and $n \geq 1$ then $T_{p^n} = T_p^n$*
(3) *for p prime and coprime to N, $n \geq 1$, $T_{p^n} \circ T_p = T_{p^{n+1}} + pT_{p^{n-1}} \circ T_{p,p}$*
(4) *$T_n \circ T_{m,m} = T_{m,m} \circ T_n$*
(5) *$T_{m,m} \circ T_{n,n} = T_{mn,mn}$.*

Proof. The last two properties are easy to check, and are left to the reader. Now we consider the first claim. Let $(L,t) \in \mathcal{L}_N$. We observe that $T_{mn}(L,t)$ is a sum over lattices L'' containing L with index mn, such that t still has exact order N when reduced modulo L''. Since m and n are coprime, there is a unique lattice L' such that

$$L \subset L' \subset L''$$

and L has index n in L'. Indeed, L'/L is the unique subgroup of L''/L of order n. Clearly t has exact order N when reduced modulo L''.

Conversely, given

$$L \subset L' \subset L''$$

where L has index n in L' and L' has index m in L'', and $t \in \mathbb{C}/L$ such that t has exact order N modulo L'' we see that $L \subset L''$ has index mn. So we see that the elements of $\mathcal{L}_N$ occuring in $T_{mn}(L,t)$ and $T_m \circ T_n(L,t) = \frac{1}{m}\sum_{L'} T_n(L',t)$ are the same and we have $T_n = T_m \circ T_n$.

Now we consider the second item. By induction, it suffices to show that $T_{p^{n-1}}T_p = T_{p^n}$ for $n \geq 2$. Let $t' = (N/p)t$. Then $T_{p^n}(L,t) = p^{-n}\sum[(L',t)]$ where the summation is over $L' \supset L$ such that $L'/L \subset \frac{1}{p^n}L/L$ has order p^n and does not contain t'. Now we claim that L'/L is cyclic. This is because if it is not cyclic it contains $\frac{1}{p}L/L$ which contains t' (since $pt' = 0$). We may now argue as in the previous part, since a cyclic subgroup of order p^n contains a unique subgroup of order p.

For the third claim, first note that $T_{p^n} \circ T_p(L,t)$, $T_{p^{n+1}}(L,t)$ and $T_{p^{n-1}}T_{p,p}(L,t)$ are all given by linear combinations of lattices containing L with index p^{n+1}. Let L'' be such a lattice. Denote its coefficient in the three terms by a, b, c. Then we want to show that $a = b + pc$. We can immediately observe that $b = 1$. There are now two cases:

(1) $L'' \not\supset \frac{1}{p}L$: this implies that $c = 0$. Now a is the number of lattices L' contained in L'' with index p^n. Such an L' is contained in $L'' \cap \frac{1}{p}L$. Since $L'' \not\supset \frac{1}{p}L$ we actually have $L' = L'' \cap \frac{1}{p}L$ and so $a = 1$ and we are done.
(2) $L'' \supset \frac{1}{p}L$: in this case $c = 1$, and L' as above can be any sublattice of $\frac{1}{p}L$ of index p. So $a = 1 + p$ and we are again done.

$\square$

Corollary 6.7. *The T_n are polynomials in the elements T_p and $T_{p,p}$, for $p|n$.*

Proof. This follows from induction on n. $\square$

Corollary 6.8. *The $\mathbb{C}$-subalgebra $\mathbb{T}$ of $\mathrm{End}(\{F : \mathcal{L}_N \to \mathbb{C}\})$ generated by the T_p and $T_{p,p}$ for p prime is commutative and contains all the T_n and $T_{n,n}$.*

Proof. This follows from the above proposition and corollary. $\square$

The above relations between the T_n and $T_{n,n}$ can be nicely summarised as indentities of formal power series with coefficients in $\mathbb{T}$. For $p|N$ we have (in the ring $\mathbb{T}[[X]]$) an identity

$$\sum_{n=0}^{\infty} T_{p^n} X^n = \frac{1}{1 - T_p X}.$$

For $p \nmid N$ we have

$$\sum_{n=0}^{\infty} T_{p^n} X^n = \frac{1}{1 - T_p X + p T_{p,p} X^2}.$$

If we replace X in these identities with p^{-s} we get

$$\sum_{n=1}^{\infty} T_n n^{-s} = \prod_p \sum_{n=0}^{\infty} T_{p^n} p^{-ns} = \prod_{p|N} \frac{1}{1 - T_p p^{-s}} \prod_{p\nmid N} \frac{1}{1 - T_p p^{-s} + T_{p,p} p^{1-2s}}.$$

Definition 6.9. For $d \in \mathbb{Z}$ coprime to N and $F : \mathcal{L}_N \to \mathbb{C}$ denote by $\langle d \rangle F$ the function defined by $\langle d \rangle F(L, t) = F(L, dt)$. Since t has order N this depends only on the class of d in $(\mathbb{Z}/N\mathbb{Z})^\times$.

Lemma 6.10. *The actions of T_n, $T_{n,n}$ and $\langle d \rangle$ take weight k functions on $\mathcal{L}_N$ to weight k functions on $\mathcal{L}_N$. If F is a weight k function on $\mathcal{L}_N$ then $T_{n,n}F = n^{k-2}\langle n \rangle F$.*

Proof. Left to the reader. $\square$

As a consequence, we have actions of T_n, $T_{n,n}$ and $\langle d \rangle$ on the vector space of functions on $\mathcal{H}$ invariant under $\Gamma_1(N)$ acting by the weight k slash operator. There is a more matrix theoretic description of the Hecke operators, as follows.

Definition 6.11. Let S_n^N be the set of matrices (with integer entries) $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with $ad = n$, $a \geq 1$, a coprime to N and $0 \leq b < d$.

Suppose $\sigma \in S_n^N$ and $(L, t) \in \mathcal{L}_N$.Write $L = \mathbb{Z}\omega_1 \oplus \mathbb{Z}\omega_2$ with $\omega_1/\omega_2 \in \mathcal{H}$ and $t = \omega_2/N$. Then we denote by L_σ the lattice with basis $(\frac{a}{n}\omega_1 + \frac{b}{n}\omega_2, \frac{d}{n}\omega_2)$.

Remark. The choice of ω_1, ω_2 in the above definition is well-defined up to multiplication of the column vector $\begin{pmatrix}\omega_1\\ \omega_2\end{pmatrix}$ by an element of $\Gamma_1(N)$ (this is Lemma 6.3). The lattices L_σ depend on the choice of ω_1, ω_2 but the following lemma shows that the *set* of lattices $\{L_\sigma : \sigma \in S_n^N\}$ is independent of this choice.

Lemma 6.12. *The map $\sigma \mapsto L_\sigma$ is a bijection from S_n^N to the set of lattices $L' \supset L$ with $[L' : L] = n$ such that t has order N when reduced modulo L'.*

Proof. Since $\det(\sigma) = n$ we see that L has index n in L_σ. Since a is coprime to N, ω_2/N still has order N in $\mathbb{C}/L_\sigma$. Conversely, suppose L has index n in a lattice L' and t has order N modulo L'. Then we let a and d be the cardinality of $Y_1 = \frac{1}{n}L/(L' + \frac{1}{n}\mathbb{Z}\omega_2)$ and $Y_2 = \frac{1}{n}\mathbb{Z}\omega_2/L' \cap \frac{1}{n}\mathbb{Z}\omega_2$ respectively.

There is a short exact sequence of abelian groups

$$0 \to Y_2 \to \frac{1}{n}L/L' \to Y_1 \to 0$$

so $ad = n$. Since $t = \omega_2/N$ and $L' \cap \frac{1}{n}\mathbb{Z}\omega_2 = \frac{d}{n}\mathbb{Z}\omega_2 = \frac{1}{a}\mathbb{Z}\omega_2$, the condition that t has order N modulo L' is equivalent to the condition that a is coprime to N. Since $\frac{a}{n}\omega_1$ has image zero in Y_1, there exists $b \in \mathbb{Z}$ such that $\frac{a}{n}\omega_1 + \frac{b}{n}\omega_2 \in L'$. Since $\frac{d}{n}\omega_2 \in L'$, we can find a unique such b in the range $0 \le b < d$. We have now associated a, b, d to L' such that $\frac{a}{n}\omega_1 + \frac{b}{n}\omega_2$ and $\frac{d}{n}\omega_2$ are in L'. Since these elements span a lattice which contains L with index n, they span L'. Now we have constructed a map $L' \mapsto \begin{pmatrix} a & b\\ 0 & d\end{pmatrix}$ which is an inverse to the map $\sigma \mapsto L_\sigma$. □

Proposition 6.13. *The actions of T_n, $T_{n,n}$ and $\langle d\rangle$ preserve the spaces $M_k(\Gamma_1(N))$ and $S_k(\Gamma_1(N))$.*

Proof. We saw above that $T_{n,n}F = n^{k-2}\langle n\rangle F$ for weight k functions F, so it is enough to consider the operators $\langle d\rangle$ and T_n.

First we consider $\langle n\rangle$, for n coprime to N. Let $f \in M_k(\Gamma_1(N))$, with associated weight k function F on $\mathcal{L}_N$. Let $\sigma_n = \begin{pmatrix} a & b\\ c & d\end{pmatrix} \in \Gamma(1)$ be an element which is congruent mod N to $\begin{pmatrix} n^{-1} & 0\\ 0 & n\end{pmatrix}$. Such an element exists because $\Gamma(1)$ surjects onto $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. Observe that

$$f|_{\sigma_n,k}(\tau) = (c\tau+d)^{-k}F(\mathbb{Z}\sigma_n\tau+\mathbb{Z}, 1/N) = F(\mathbb{Z}(a\tau+b)+\mathbb{Z}(c\tau+d), n/N) = F(\mathbb{Z}\tau+\mathbb{Z}, n/N).$$

Now we have

$$\langle n\rangle f(\tau) = \langle n\rangle F(\mathbb{Z}\tau + \mathbb{Z}, 1/N) = F(\mathbb{Z}\tau + \mathbb{Z}, n/N) = f|_{\sigma_n,k}(\tau),$$

and the statement of the proposition for $\langle n\rangle$ follows.

The case of T_n uses the set of matrices S_n^N. For $\tau \in \mathcal{H}$ we have the lattice $L_\tau = \mathbb{Z}\tau + \mathbb{Z}$. By definition, $f(\tau) = F(L_\tau, 1/N)$ and

$$T_n f(\tau) = T_n F(L_\tau, 1/N) = \frac{1}{n} \sum_{\sigma \in S_n^N} F(L_{\tau,\sigma}, 1/N).$$

For $\sigma = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in S_n^N$ we have $(L_{\tau,\sigma}, 1/N) = \frac{1}{a}(L_{\sigma\tau}, a/N)$ so

$$F(L_{\tau,\sigma}, 1/N) = a^k F(L_{\sigma\tau}, a/N) = a^k \langle a \rangle F(L_{\sigma\tau}, 1/N) = a^k \langle a \rangle f(\sigma\tau).$$

We can write this as

$$T_n f(\tau) = n^{k-1} \sum_{\sigma \in S_n^N} (\langle a \rangle f)|_{\sigma,k}(\tau).$$

From here we can deduce the proposition. $\square$

Remark 6.14. The above proof shows that the action of $(\mathbb{Z}/N\mathbb{Z})^\times$ on $M_k(\Gamma_1(N))$ can be identified with the action of $\Gamma_0(N)/\Gamma_1(N)$ by the weight k slash operator, via the isomorphism $(\mathbb{Z}/N\mathbb{Z})^\times \cong \Gamma_0(N)/\Gamma_1(N)$ given by

$$d + N\mathbb{Z} \mapsto \begin{pmatrix} d^{-1} & 0 \\ 0 & d \end{pmatrix} \Gamma_1(N).$$

The finite Abelian group $(\mathbb{Z}/N\mathbb{Z})^\times$ now acts on the finite dimensional $\mathbb{C}$-vector space $M_k(\Gamma_1(N))$ by the $\langle \cdot \rangle$ action. So this vector space decomposes into a direct sum indexed by characters $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$:

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

where $M_k(\Gamma_1(N), \chi)$ denotes the subspace of $M_k(\Gamma_1(N))$ on which $\langle d \rangle$ acts as multiplication by $\chi(d)$ for every $d \in (\mathbb{Z}/N\mathbb{Z})^\times$. We write $M_k(N, \chi)$ to abbreviate $M_k(\Gamma_1(N), \chi)$.

We now consider the effect of the Hecke operator T_n on the q-expansion at ∞ of a modular form $f \in M_k(N, \chi)$.

Proposition 6.15. *Let $f \in M_k(N, \chi)$ with $f(\tau) = \sum_{i=0}^\infty a_n q^n$ and let $T_p f(\tau) = \sum_{i=0}^\infty b_n q^n$. Then*

$$b_n = a_{np} + \chi(p) p^{k-1} a_{n/p}$$

where we take $\chi(p) = 0$ if $p|N$ and $a_{n/p} = 0$ if $p \nmid n$.

Proof. The set of matrices S_p has a simple description. If $p|N$ then S_p consists of matrices $\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$ with $a = 1$, $d = p$ and $b = 0, 1, ..., p-1$. Therefore we have

$$T_p f(\tau) = \frac{1}{p} \sum_{b=0}^{p-1} f\left(\frac{\tau + b}{p}\right).$$

Since the sum

$$\frac{1}{p}\sum_{b=0}^{p-1} e^{2\pi i n(\frac{\tau+b}{p})} = \frac{1}{p} q^{n/p} \sum_{b=0}^{p-1} e^{2\pi i n b/p}$$

is equal to zero if p does not divide n and equals $q^{n/p}$ if p does divide n, we see that $b_n = a_{np}$ as required.

Now suppose that $p \nmid N$. Then we have one additional element of S_p, given by $a = p$, $d = 1$ and $b = 0$. So we have

$$T_p f(\tau) = \frac{1}{p}\left(\sum_{b=0}^{p-1} f\left(\frac{\tau+b}{p}\right) + p^k \langle p \rangle f(p\tau)\right).$$

Since $\langle p \rangle f = \chi(p) f$ we obtain the desired result. □

We will write the above formula for the effect of T_p on q-expansions in terms of some operators on the ring of formal power series $\mathbb{C}[[q]]$. For $m \geq 1$ an integer we write U_m for the operator which takes $\sum_{n\geq 0} a_n q^n$ to $\sum_{n\geq 0} a_n q^{n/m}$ (where $q^{n/m}$ is zero if $m \nmid n$). We write V_m for the operator which takes $\sum_{n\geq 0} a_n q^n$ to $\sum_{n\geq 0} a_n q^{mn}$. We have $U_m \circ V_m$ equals the identity and $V_m \circ U_m$ equals the operator given by retaining only the q^n terms where $m|n$. The above Proposition just says that

$$T_p = U_p + \chi(p) p^{k-1} V_p$$

as operators on q-expansions. This allows us to write a formal factorisation

$$1 - T_p X + \chi(p) p^{k-1} X^2 = (1 - U_p X)(1 - \chi(p) p^{k-1} V_p X).$$

If you like, this equality takes place in the (non-commutative) ring of $\mathbb{C}$-linear endomorphisms of $\mathbb{C}[[q]]$.

Recall that we also have a formal identity

$$\sum_{n=1}^{\infty} T_n n^{-s} = \prod_{p|N} \frac{1}{1 - T_p p^{-s}} \prod_{p\nmid N} \frac{1}{1 - T_p p^{-s} + T_{p,p} p^{1-2s}}.$$

Here the right hand side is equal to

$$\prod_p [(1 - U_p p^{-s})(1 - \chi(p) p^{k-1} V_p p^{-s})]^{-1} = \prod_p (1 - \chi(p) p^{k-1} V_p p^{-s})^{-1} (1 - U_p p^{-s})^{-1}.$$

Note the change in order of the product when we compute the inverse, since U_p and V_p do not commute. Since for distinct p and p', U_p and $V_{p'}$ do commute, we can collect all the V_p and U_p terms. Doing the standard geometric series expansion, we obtain an equality

$$\sum_{n=1}^{\infty} T_n n^{-s} = \left(\sum_{n=1}^{\infty} \chi(n) n^{k-1} V_n n^{-s}\right)\left(\sum_{n=1}^{\infty} U_n n^{-s}\right).$$

From this we deduce

Proposition 6.16.

$$T_n = \sum_{0<d|n} \chi(d) d^{k-1} V_d \circ U_{n/d}.$$

Corollary 6.17. *Let $f \in M_k(N,\chi)$ with $f(\tau) = \sum_{i=0}^{\infty} a_n q^n$ and let $T_m f(\tau) = \sum_{i=0}^{\infty} b_n q^n$. Then*

$$b_n = \sum_{d \mid \gcd(m,n)} \chi(d) d^{k-1} a_{mn/d^2}.$$

Proof. We have

$$T_m f = \sum_{d \mid m} \chi(d) d^{k-1} V_d \circ U_{m/d} f = \sum_{d \mid m} \chi(d) d^{k-1} V_d \sum_{m/d \mid n} a_n q^{dn/m} = \sum_{d \mid m} \chi(d) d^{k-1} \sum_{m/d \mid n} a_n q^{d^2 n/m}.$$

Now we set $r = d^2 n/m$ so we have

$$T_m f = \sum_{d \mid m} \chi(d) d^{k-1} \sum_{d \mid r} a_{mr/d^2} q^r.$$

From here we can immediately obtain the statement of the corollary. □

Definition 6.18. We say that $f \in M_k(N,\chi)$ is an *eigenform* if $T_n f = \lambda_n f$ for some $\lambda_n \in \mathbb{C}$, for all $n \in \mathbb{Z}_{\geq 1}$. We also say thaat f is *normalised* if $a_1 = 1$.

Lemma 6.19. *Suppose f is a non-constant eigenform with Hecke eigenvalues λ_n. Then $a_1(f) \neq 0$ and $\lambda_n = a_n(f)/a_1(f)$. Moreover, if $a_0(f) \neq 0$ then $\lambda_n = \sum_{d \mid n} \chi(d) d^{k-1}$ for all $n \geq 1$.*

Proof. Since $\lambda_n a_1(f) = a_1(T_n f) = a_n(f)$ (by Corollary 6.17), if $a_1(f) = 0$ and f is an eigenform then $a_n(f) = 0$ for all $n \geq 1$, so f is constant.

Suppose that $a_0(f) \neq 0$. We have $\lambda_n a_0(f) = a_0(T_n f) = \sum_{d \mid n} \chi(d) d^{k-1} a_0(f)$, so $\lambda_n = \sum_{d \mid n} \chi(d) d^{k-1}$ as required. □

Here are some examples of eigenforms:

(1) The level one Eisenstein series $E_k(\tau)$ for $k > 2$ even. The Hecke eigenvalues λ_n are equal to $\sum_{d \mid n} \chi(d) d^{k-1}$.
(2) For characters χ with $\chi(-1) = (-1)^k$ and $k > 2$ there are Eisenstein series $E_k^\chi \in M_k(N,\chi)$ with Hecke eigenvalues $\sum_{d \mid n} \chi(d) d^{k-1}$.
(3) Whenever a space of modular forms (or cusp forms) is one-dimensional, an element of this space is automatically a eigenform. For example $\Delta \in S_{12}(\Gamma(1))$ and $\theta(\tau)^2 \in M_1(\Gamma_1(4))$.

Let's consider the final example of $f = \theta^2 \in M_1(4,\chi)$ a little more closely. Here χ is the unique non-trivial character of $(\mathbb{Z}/4\mathbb{Z})^\times$. We have $a_0(f) = 1$ and $a_1(f) = 4$, so $T_n(f) = (\sum_{d \mid n} \chi(d)) f$. In particular, for odd primes p the Hecke eigenvalue λ_p is equal to $1 + \left(\frac{-1}{p}\right)$ where $\left(\frac{-1}{p}\right)$ denotes the Legendre symbol (it is 1 if -1 is a square mod p and -1 otherwise).

This means we can interpret the Hecke eigenvalues as the traces of certain elements of $\mathrm{Gal}(\mathbb{Q}(i)/\mathbb{Q})$ acting on a two-dimensional $\mathbb{C}$-vector space by the direct sum of characters $\mathbf{1} \oplus \widetilde{\chi}$. Here $\widetilde{\chi}$ is the unique non-trivial character of $\mathrm{Gal}(\mathbb{Q}(i)/\mathbb{Q})$.

These elements are Frobenius elements at places dividing p. More explicitly, if p splits in $\mathbb{Q}(i)$, i.e. if $\left(\frac{-1}{p}\right) = 1$, then we take the identity element. If p is inert in $\mathbb{Q}(i)$, i.e. $\left(\frac{-1}{p}\right) = -1$, then we take the non-trivial element (given by complex conjugation).

This is an example of a general theorem of Deligne and Serre which attaches two-dimensional Galois representations to all eigenforms $f \in M_k(N,\chi)$.

Proposition 6.20. *Let $f \in M_k(N,\chi)$. Then f is a normalised (i.e. $a_1(f) = 1$) eigenform if and only if:*

- $a_1(f) = 1$
- *for p prime, $n \geq 1$, $a_{p^n}(f)a_p(f) = a_{p^{n+1}}(f) + \chi(p)p^{k-1}a_{p^{n-1}}(f)$*
- *$a_{mn}(f) = a_m(f)a_n(f)$ when m and n are coprime.*

Proof. We already know that if f is a normalised eigenform then these properties are satisfied. For the other direction, we now suppose that f satisfies these properties. It is enough to show that f is an eigenform for every T_p, or indeed to show that $a_n(T_pf) = a_p(f)a_n(f)$ for every $n \in \mathbb{Z}_{\geq 1}$.

Let's suppose $n \geq 1$. We know that we have $a_n(T_pf) = a_{pn}(f)$ if $p \nmid n$ and $a_n(T_pf) = a_{pn} + \chi(p)p^{k-1}a_{n/p}(f)$ if $p|n$.

In the first case, we get $a_p(f)a_n(f)$ as desired. In the second case, we write $n = p^r m$ with $p \nmid m$ and then we have

$$\begin{aligned} a_n(T_pf) &= a_{p^{r+1}m} + \chi(p)p^{k-1}a_{p^{r-1}m}(f) = a_m(f)(a_{p^{r+1}}(f) + \chi(p)p^{k-1}a_{p^{r-1}}(f)) \\ &= a_m(f)a_{p^r}(f)a_p(f) = a_p(f)a_n(f). \end{aligned}$$

So we have proved that $T_pf - a_p(f)f$ is a constant. It is also an element of $M_k(N,\chi)$, so if $k > 0$ we have $T_pf = a_p(f)f$ as desired. If $k = 0$ then everything is constant and there are no normalised eigenforms! (Since $a_1 = 1$ is impossible). $\square$

6.3. **Petersson inner product.**

We define a measure $d\mu$ on $\mathcal{H}$ by $d\mu(\tau) = \frac{dxdy}{y^2}$, where $\tau = x + iy$. This measure is actually $\mathrm{GL}_2^+(\mathbb{R})$-invariant, so defines a measure on $\Gamma\backslash\mathcal{H}$ for any congruence subgroup $\Gamma \subset \Gamma(1)$. Integrating over $Y(\Gamma)$ is the same as integrating over a fundamental domain for Γ in $\mathcal{H}$.

For simplicity we will only consider fundamental domains of the form $\coprod_{\alpha_i} \alpha_i\mathscr{F}(1)$, where $\mathscr{F}(1)$ is the standard fundamental domain for $\Gamma(1)$ and $\alpha_i \in \mathrm{PSL}_2(\mathbb{Z})$ are coset representatives for $\mathrm{PSL}_2(\mathbb{Z})/\overline{\Gamma}$.

Lemma 6.21. *Let $\mathscr{F}$ be a fundamental domain for Γ and define $\mu(\Gamma) = \int_{\mathscr{F}} \frac{dxdy}{y^2}$. Then*

1. *The integral $\mu(\Gamma)$ converges and is independent of the choice of $\mathscr{F}$.*
2. $[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}] = \mu(\Gamma)/\mu(\Gamma(1))$.

Proof. Write $[\mathrm{PSL}_2(\mathbb{Z}) : \overline{\Gamma}] = d$. Let $\mathrm{PSL}_2(\mathbb{Z}) = \coprod_{i=1}^d \overline{\Gamma}\alpha_i$ and take $\mathscr{F}(1)$ to be the standard fundamental domain for $\Gamma(1)$. Then we let $\mathscr{F} = \cup_i alpha_i\mathscr{F}(1)$. This is a fundamental domain for Γ. We can easily bound $\mu(\Gamma(1))$ by

$$\int_{\mathscr{F}(1)} \frac{dxdy}{y^2} < \int_{-1/2}^{1/2} \int_{\sqrt{3}/2}^{\infty} y^{-2} dydx = \frac{2}{\sqrt{3}}.$$

Under the change of variables $z \mapsto \alpha_i z$ the measure $\frac{dxdy}{y^2}$ is invariant so we get $\mu(\Gamma) = d\mu(\Gamma(1))$. The invariance of the measure under the action of $\Gamma(1)$ likewise enables us to easily show independence of the integral on the choice of a fundamental domain. $\square$

Definition 6.22. Let $f, g \in M_k(\Gamma)$, with at least one of f, g a cusp form. We define

$$\langle f, g \rangle = \frac{\mu(\Gamma(1))}{\mu(\Gamma)} \int_{Y(\Gamma)} f(\tau)\overline{g(\tau)} y^k d\mu.$$

Lemma 6.23. *The integral in the above definition is absolutely convergent, and can be computed as an integral over a fundamental domain $\mathscr{F}$ for Γ. If $\Gamma' \subset \Gamma$ is another congruence subgroup then the definition of $\langle f, g \rangle$ is independent of whether f, g are considered in $M_k(\Gamma)$ or $M_k(\Gamma')$.*

Proof. The convergence follows from the following lemma, since $fg \in S_{2k}(\Gamma)$. We leave the rest of the lemma as an exercise. $\square$

Lemma 6.24. *Suppose $f \in S_k(\Gamma)$. Then $|f(\tau)| \leq C(\mathrm{Im}(\tau))^{-k/2}$ for some constant C independent of τ.*

Proof. First we set $\phi(x + iy) = |f(x + iy)|y^{k/2}$. It is easy to check that ϕ is Γ-invariant, so we just need to show that it is bounded on $Y(\Gamma)$. Suppose s is a cusp with $\alpha\infty = s$. We have $f|_{\alpha,k}(\tau) = \sum_{n=1}^{\infty} b_n q_{h'}^n = q_{h'}\theta(q_{h'})$ for some h', and a holomorphic function θ on the open unit disc. So

$$\phi(\alpha\tau) = |f_{\alpha,k}(\tau)||y^{k/2} = |\theta(q_{h'})|e^{-2\pi y/h'} y^{k/2}$$

tends to zero uniformly in x as y tends to ∞. So in fact ϕ defines a continuous function on the compact topological space $X(\Gamma)$ (with value zero at the cusps), so it is in particular bounded on $Y(\Gamma)$. $\square$

Here is another useful corollary of this lemma:

Corollary 6.25. *Suppose $f \in S_k(\Gamma)$, with $f(\tau) = \sum_{n=1}^{\infty} a_n q_h^n$. Then there is a constant C such that $|a_n| \leq Cn^{k/2}$ for all $n \geq 1$.*

Proof. Exercise. $\square$

Lemma 6.26. *For $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$*

$$\langle f, g \rangle = (\det \alpha)^k \langle f|_{\alpha,k}, g|_{\alpha,k} \rangle.$$

Proof. Set $\Gamma' = \Gamma \cap \alpha\Gamma\alpha^{-1}$. We have $f, g \in M_k(\Gamma')$ and $f|_{\alpha,k}, g|_{\alpha,k} \in M_k(\alpha^{-1}\Gamma'\alpha)$.

Suppose $\mathscr{F}(\Gamma')$ is a fundamental domain for Γ'. Then $\alpha^{-1}\mathscr{F}(\Gamma')$ is a fundamental domain for $\alpha^{-1}\Gamma'\alpha$.

It follows from the $\mathrm{GL}_2^+(\mathbb{Q})$-invariance of the measure μ that $\mu(\alpha^{-1}\Gamma'\alpha) = \mu(\Gamma')$. We compute

$$\begin{aligned}(\det\alpha)^k\langle f|_{\alpha,k}, g|_{\alpha,k}\rangle &= \frac{\mu(\Gamma(1))}{\mu(\Gamma')}\int\limits_{\alpha^{-1}\mathscr{F}(\Gamma')} f|_{\alpha,k}\overline{g|_{\alpha,k}}(\det\alpha\ \mathrm{Im}\tau)^k d\mu \\ &= \frac{\mu(\Gamma(1))}{\mu(\Gamma')}\int\limits_{\mathscr{F}(\Gamma')} \frac{f\overline{g}}{|c\tau+d|^{2k}}(\det\alpha\ \mathrm{Im}\alpha^{-1}\tau)^k d\mu \\ &= \frac{\mu(\Gamma(1))}{\mu(\Gamma')}\int\limits_{\mathscr{F}(\Gamma')} f\overline{g}(\mathrm{Im}\tau)^k d\mu = \langle f, g\rangle,\end{aligned}$$

since $\det\alpha\ \mathrm{Im}\tau = |c\tau+d|^2\mathrm{Im}\alpha\tau$. □

I messed up the proof of the following corollary in lectures, sorry!

Corollary 6.27. *If $f, g \in M_k(N,\chi)$, with one of f, g a cusp form, and $n \in \mathbb{Z}_{\geq 1}$ coprime to N, then $\langle T_n f, g\rangle = \chi(n)\langle f, T_n g\rangle$.*

Proof. It suffices to prove the corollary when $n = p$ is prime. Recall that we have a matrix $\sigma_p \in \Gamma_0(N)$ with $\sigma_p = \begin{pmatrix}\alpha & \beta \\ N & p\end{pmatrix}$.

Consider the set of matrices

$$\Delta_p^N := \{\gamma = \begin{pmatrix} a & b \\ c & d\end{pmatrix} \in M_2(\mathbb{Z}) : N \mid c, N \mid (a-1), \det(\gamma) = p, \}.$$

We can check that

$$\begin{aligned}\Delta_p^N &= \Gamma_1(N)\begin{pmatrix}1 & 0 \\ 0 & p\end{pmatrix}\Gamma_1(N) = \{u\begin{pmatrix}1 & 0 \\ 0 & p\end{pmatrix}v : u, v \in \Gamma_1(N)\} \\ &= \coprod_{j=0,\dots,p-1}\Gamma_1(N)\begin{pmatrix}1 & j \\ 0 & p\end{pmatrix}\coprod\Gamma_1(N)\sigma_p\begin{pmatrix}p & 0 \\ 0 & 1\end{pmatrix}.\end{aligned}$$

Now suppose $\gamma \in \Delta_p^N$. We have $\langle f, g|_{\gamma^{-1},k}\rangle = p^k\langle f|_{\gamma,k}, g\rangle$ by the above Lemma. Moreover, the values of $\langle f|_{\gamma,k}, g\rangle$ and $\langle f, g|_{\gamma,k}\rangle$ are actually independent of the choice of $\gamma \in \Delta_p^N$: if $\gamma' = u\gamma v$ with $u, v \in \Gamma_1(N)$, then

$$\langle f|_{u\gamma v,k}, g\rangle = \langle f|_{\gamma,k}, g|_{v^{-1},k}\rangle = \langle f|_{\gamma,k}, g\rangle.$$

A similar argument applies to $\langle f, g|_{\gamma,k}\rangle$.

Since $T_pf = p^{k-1}\left(\sum_{j=0}^{p-1} f|_{\left(\begin{smallmatrix}1 & j\\ 0 & p\end{smallmatrix}\right),k} + f|_{\sigma_p\left(\begin{smallmatrix}p & 0\\ 0 & 1\end{smallmatrix}\right),k}\right)$ we have

$$\begin{aligned}
\langle T_pf, g\rangle &= p^{k-1}(p+1)\langle f|_{\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right),k}, g\rangle = p^{-1}(p+1)\langle f, g|_{\left(\begin{smallmatrix}1 & 0\\ 0 & p^{-1}\end{smallmatrix}\right),k}\rangle\\
&= p^{k-1}(p+1)\langle f, g|_{\sigma_p^{-1}\sigma_p\left(\begin{smallmatrix}p & 0\\ 0 & 1\end{smallmatrix}\right),k}\rangle = \chi(p)p^{k-1}(p+1)\langle f, g|_{\sigma_p\left(\begin{smallmatrix}p & 0\\ 0 & 1\end{smallmatrix}\right),k}\rangle\\
&= \chi(p)p^{k-1}(p+1)\langle f, g|_{\left(\begin{smallmatrix}1 & 0\\ 0 & p\end{smallmatrix}\right),k}\rangle = \chi(p)\langle f, T_pg\rangle.
\end{aligned}$$

Here we use the observation that $\sigma_p\left(\begin{smallmatrix}p & 0\\ 0 & 1\end{smallmatrix}\right) \in \Delta_p^N$.

□

Corollary 6.28. *The space $S_k(N,\chi)$ has a basis (orthonormal with respect to the Petersson inner product) consisting of simultaneous eigenvectors for the Hecke operators T_n with n coprime to N.*

Proof. For each n choose a square root c_n of $\overline{\chi(n)}$. Then for all $f, g \in S_k(N,\chi)$ we have

$$\langle c_nT_nf, g\rangle = \langle f, c_nT_ng\rangle.$$

So the operators c_nT_n are Hermitian and so have an orthonormal basis of eigenvectors. Since all the Hecke operators T_n with n coprime to N commute, we have a basis of simultaneous eigenvectors. □

Remark 6.29. Note that the eigenvalues of c_nT_n are real. In particular, if χ is trivial then the eigenvalues of the T_n are real.

If we want to find a basis of eigenforms (i.e. eigenvectors for *all* the Hecke operators) then we have to restrict to certain subspaces of $S_k(N,\chi)$. Recalling definition 3.16, we can define $S_k(N,\chi)^{\mathrm{old}} = M_k(\Gamma_1(N))^{\mathrm{old}} \cap S_k(\Gamma_1(N))$.

Definition 6.30. Define $S_k(N,\chi)^{\mathrm{new}}$ to be the orthogonal complement of $S_k(N,\chi)^{\mathrm{old}}$ under the Petersson inner product.

An important property of the new subspace is that it has a basis of eigenforms:

Theorem 6.31. *The space $S_k(N,\chi)^{\mathrm{new}}$ is stable under the action of the Hecke operators. If $f \in S_k(N,\chi)^{\mathrm{new}}$ is an eigenvector for the T_n with n coprime to N, then f is an eigenform (i.e. an eigenvector for all then T_n).*

Corollary 6.32. *Suppose two non-zero elements f, g of $S_k(N,\chi)^{\mathrm{new}}$ are eigenvectors for the T_n with n coprime to N with the same eigenvalues. Then f and g are scalar multiples of each other.*

Proof. The Theorem implies that both f and g are eigenforms. By rescaling we can assume that they are both normalised eigenforms. Now $f - g$ is also an eigenform, but has first q-expansion coefficient $a_1 = 0$. Therefore $f - g = 0$. □

Remark 6.33. The above Corollary is a version of a 'multiplicity one' theorem. Various stronger forms of this theorem can be proven: for example, in Miyake's book 'Modular Forms' it is proven by fairly elementary arguments that if f, g have the same T_p eigenvalue for all but finitely many primes p, then f and g are scalar multiples of each other.

7. L–functions

Definition 7.1. For $f \in M_k(\Gamma)$ and $s \in \mathbb{C}$ set $L(f, s) = \sum_{n \geq 1} \frac{a_n}{n^s}$.

Lemma 7.2. *Suppose $f \in S_k(\Gamma)$ The series defining $L(f, s)$ converges absolutely and uniformly on compact subsets of $\{\mathrm{Re}(s) > k/2 + 1\}$.*

Proof. By Corollary 6.25, we have $|a_n| \leq Cn^{k/2}$. This suffices to prove the lemma. □

Remark 7.3. We can explicitly write down Eisenstein series which give the rest of the space $M_k(\Gamma)$. These have Fourier coefficents of order n^{k-1}, so for $f \in M_k(\Gamma)$ the L-function converges nicely for $\mathrm{Re}(s) > k$.

7.1. **Functional equation.** Now we are going to find a functional equation for $L(f, s)$. For simplicity we will assume that $\Gamma = \Gamma_1(N)$. Set $w_N = \begin{pmatrix} 0 & -1/\sqrt{N} \\ \sqrt{N} & 0 \end{pmatrix}$. Note that $w_N^{-1}\Gamma_1(N)w_N = \Gamma_1(N)$, so $f|_{w_N,k} \in S_k(\Gamma_1(N))$.

Explicitly, we have $f|_{w_N,k}(\tau) = N^{-k/2}\tau^{-k}f(-1/N\tau)$.

Theorem 7.4. *Let $f \in S_k(\Gamma_1(N))$ and set $g = i^k f|_{w_N,k}$. If $f = \sum_{n\geq 1} a_n q^n$ then the Dirichlet series $L(f, s) = \sum_{n\geq 1} a_n/n^s$ can be extended to a holomorphic function on $s \in \mathbb{C}$. Setting*

$$\Lambda(f, s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f, s)$$

we have a functional equation

$$\Lambda(f, s) = \Lambda(g, k - s).$$

Here $\Gamma(s)$ is meromorphic continuation of the function defined by

$$\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt.$$

Proof. We let ϕ be the function on $\mathbb{R}_{>0}$ given by $\phi(y) = f(iy)$. Consider the *Mellin transform*

$$F(s) = \int_0^\infty \phi(y)y^{s-1}dy.$$

Now $\phi(y)$ tends to zero exponentially fast as y tends to ∞. We also have $\phi(1/y) = f(-1/iy) = (iy)^k f|_{\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right),k}(iy)$. Since $f|_{\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right),k}$ is also a cusp form, $\phi(1/y)$ tends to zero like y^k times an exponential in $-y$ as y tends to infinity. So this integral converges absolutely (at both upper and lower limits) for all s.

We have $\phi(y) = \sum_{n\geq 1} a_n e^{-2\pi n y}$, and we can switch the sum and integral in $F(s)$ to get

$$F(s) = \sum_{n=1}^{\infty} a_n \int_0^{\infty} e^{-2\pi n y} y^{s-1} dy.$$

Substituting $t = 2\pi n y$ into the integral gives us

$$\int_0^{\infty} e^{-2\pi n y} y^{s-1} dy = (2\pi n)^{-s} \int_0^{\infty} e^{-t} t^{s-1} dt = (2\pi n)^{-s} \Gamma(s).$$

Therefore the switched expression also converges absolutely for $\mathrm{Re}(s) > k/2 + 1$, to $N^{-s/2}\Lambda(f,s)$, and we have $F(s) = N^{-s/2}\Lambda(f,s)$ (for $\mathrm{Re}(s) > k/2+1$). Since $F(s)$ extends to a holomorphic function for all $s \in \mathbb{C}$, $\Lambda(f,s)$ does. Moreover, $\Gamma(s)$ has no zeroes, so $L(f,s)$ also extends to a holomorphic function on the whole complex plane.

To prove the functional equation, let's substitute $u = 1/Ny$ in the integral defining $F(s)$. We get

$$\begin{aligned} N^{-s/2}\Lambda(f,s) = F(s) = N^{-s} \int_0^{\infty} \phi(1/Nu) u^{-1-s} du &= N^{-s} \int_0^{\infty} f(-1/Niu) u^{-1-s} du \\ = N^{k/2-s} \int_0^{\infty} g(iu) u^{k-1-s} du &= N^{k/2-s} N^{-(k-s)/2} \Lambda(g, k-s) = N^{-s/2}\Lambda(g,k-s). \end{aligned}$$

□

7.2. **Euler products.**

Theorem 7.5. *Suppose $f \in S_k(N,\chi)$. Then f is a normalised eigenform if and only if (for* $\mathrm{Re}(s)$ *sufficiently large)*

$$L(s,f) = \prod_p (1 - a_p p^{-s} + \chi(p) p^{k-1-2s})^{-1}.$$

Proof. By Proposition 6.20, it is enough to show that

$$L(s,f) = \prod_p (1 - a_p p^{-s} + \chi(p) p^{k-1-2s})^{-1}$$

if and only if

- $a_1(f) = 1$
- for p prime, $n \geq 1$, $a_{p^n}(f) a_p(f) = a_{p^{n+1}}(f) + \chi(p) p^{k-1} a_{p^{n-1}}(f)$
- $a_{mn}(f) = a_m(f) a_n(f)$ when m and n are coprime.

We leave it as an exercise to show this, using the following lemma. □

Lemma 7.6. *Suppose we have two Dirichlet series $\sum_{n\geq 1} \frac{a_n}{n^s}$ and $\sum_{n\geq 1} \frac{b_n}{n^s}$ which converge absolutely to the same function on* $\mathrm{Re}(s) > \sigma$ *for some positive real σ. Then $a_n = b_n$ for all n.*

Proof. Exercise. □

7.3. **Converse theorems.** Suppose $f \in S_k(N,\chi)$. We have shown $F(s) = N^{-s/2}\Lambda(f,s)$, where $F(s)$ is the Mellin transform

$$F(s) := \int_0^\infty f(iy)y^{s-1}dy.$$

The following Proposition establishes an inversion formula for the Mellin transform.

Proposition 7.7. *Suppose $g : \mathbb{R}_{>0} \to \mathbb{C}$ is twice continuously differentiable, and that c is a real number such that $y^{c-1}g(y)$, $y^c g'(y)$ and $y^{c+1}g''(y)$ are all in $L^1(\mathbb{R}_{>0})$. Then the integral*

$$G(s) := \int_0^\infty y^{s-1}g(y)dy$$

converges for $\mathrm{Re}(s) = c$ and satisfies $G(c+it) = \mathscr{O}((1+|t|)^{-2})$ (i.e. it is bounded and as t approaches ∞ it decays like $|t|^{-2}$).

Moreover, we have

$$g(y) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} y^{-s}G(s)ds$$

where the integral is up the vertical line $\mathrm{Re}(s) = c$.

Proof. If we set $s = c - 2\pi ix$ and substitute $y = e^u$ then we have

$$G(s) = F(x) = \int_{\mathbb{R}} e^{cu}g(e^u)e^{-2\pi ixu}du.$$

Now everything follows from standard properties of the Fourier transform applied to the function $f(u) = e^{cu}g(e^u)$. In particular, f is twice continuously differentible and f, f' and f'' are absolutely integrable. The Fourier transform of f'' is $(2\pi ix)^2F(x)$, so $x^2F(x)$ is bounded, which gives the growth condition on $G(s)$. $\square$

The following theorem, which is a converse to Theorem 7.4 (when the level $N = 1$), is now a simple consequence of Mellin inversion.

Theorem 7.8. *Let a_n be a sequence in $\mathbb{C}$ with $a_n \le Cn^{sigma}$ for some $\sigma \in \mathbb{R}_{>0}$. Set*

$$Z(s) := \sum_{n\ge 0}\frac{a_n}{n^s}$$

and $\Lambda(s) := (2\pi)^{-s}\Gamma(s)Z(s)$. Suppose that $\Lambda(s)$ extends to a holomorphic function on $\mathbb{C}$ which is bounded on vertical strips (i.e. regions of the form $\mathrm{Re}(s) \in [a,b]$) and satisfies

$$\Lambda(s) = i^k\Lambda(k-s).$$

Then $f(\tau) = \sum_{n\ge 1} a_nq^n$ is in $S_k(\Gamma(1))$.

Proof. We define a holomorphic function on $\mathcal{H}$ by $f(\tau) := \sum_{n\geq 1} a_n q^n$. We need only to prove that $f(-1/\tau) = (\tau)^k f(\tau)$. By uniqueness of analytic continuation, it suffices to prove that $f(i/y) = (iy)^k f(iy)$ for $y \in \mathbb{R}_{>0}$.

Set $\phi(y) = f(iy)$. We have $|\phi(y)| \leq C \sum_{n\geq 1} n^\sigma e^{-2\pi n y}$, and (possibly increasing σ) we can assume that σ is a natural number. Since

$$\sum_{n\geq 0} e^{-2\pi n y} = \frac{1}{1-e^{-2\pi y}} = \frac{1}{2\pi y} + g(y)$$

with $g(y)$ holomorphic at $y = 0$, differentiating σ times gives

$$\sum_{n\geq 1} n^\sigma e^{-2\pi n y} = \mathscr{O}(y^{-(1+\sigma)})$$

as y approaches 0. A similar argument shows that $\phi(y)$ is $\mathscr{O}(e^{-2\pi y})$ as y approaches ∞, so the hypotheses of Proposition 7.7 are satisfied by ϕ (for any $c > \sigma + 1$). The Mellin transform of ϕ is given by $\Lambda(s)$. Therefore

$$\phi(y) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} y^{-s} \Lambda(s) ds$$

for $c > 1 + \sigma$. Fix such a c (which we also assume is $> k/2$) and consider the strip $\mathrm{Re}(s) \in [k-c, c]$. For $\mathrm{Re}(s) = c$ we have

$$|\Lambda(k-s)| = |\Lambda(s)| = \mathscr{O}((1 + |\mathrm{Im}(s)|)^{-2}),$$

and by hypothesis $\Lambda(s)$ is bounded on the region $\mathrm{Re}(s) \in [k-c, c]$. So the Phragmén–Lindelöf principle (see Lemma 4.3.4 in Miyake) implies that we have $\Lambda(s) = \mathscr{O}((1 + |\mathrm{Im}(s)|)^{-2})$ uniformly for $\mathrm{Re}(s) \in [k-c, c]$. This allows us to move the line of integration to get

$$\phi(y) = \frac{1}{2\pi i} \int_{k/2-i\infty}^{k/2+i\infty} y^{-s} \Lambda(s) ds = \frac{1}{2\pi i} \int_{k/2-i\infty}^{k/2+i\infty} y^{-s} i^k \Lambda(k-s) ds.$$

Substituting $t = k - s$ we have

$$\phi(y) = \frac{1}{2\pi i} \int_{k/2-i\infty}^{k/2+i\infty} y^{-(k-t)} i^k \Lambda(t) dt = i^k y^{-k} \phi(1/y).$$

This establishes that $f(-1/\tau) = (-1)^k (\tau)^k f(\tau)$. We didn't assume a priori that k was even, so we need to check this. We have

$$f(\tau) = f(-1/(-1/\tau)) = (-1)^k (-1/\tau)^k f(-1/\tau) = (-1)^k f(\tau)$$

so if f is non-zero then k is even. This completes the proof. $\square$