Part III — Modular Forms and L-functions

Based on lectures by A. J. Scholl

Notes taken by Dexter Chua

Lent 2017

Modular Forms are classical objects that appear in many areas of mathematics, from number theory to representation theory and mathematical physics. Most famous is, of course, the role they played in the proof of Fermat's Last Theorem, through the conjecture of Shimura-Taniyama-Weil that elliptic curves are modular. One connection between modular forms and arithmetic is through the medium of L-functions, the basic example of which is the Riemann ζ-function. We will discuss various types of L-function in this course and give arithmetic applications.

Pre-requisites

Prerequisites for the course are fairly modest; from number theory, apart from basic elementary notions, some knowledge of quadratic fields is desirable. A fair chunk of the course will involve (fairly 19th-century) analysis, so we will assume the basic theory of holomorphic functions in one complex variable, such as are found in a first course on complex analysis (e.g. the 2nd year Complex Analysis course of the Tripos).

Contents

0 Introduction

One of the big problems in number theory is the so-called Langlands programme, which is relates "arithmetic objects" such as representations of the Galois group and elliptic curves over $\mathbb{Q}$, with "analytic objects" such as modular forms and more generally automorphic forms and representations.

Example. $y^2 + y = x^3 - x$ is an elliptic curve, and we can associate to it the function

$$f(z) = q\prod_{n\geq 1}(1-q^n)^2(1-q^{11n})^2 = \sum_{n=1}^{\infty} a_n q^n, \quad q = e^{2\pi i z},$$

where we assume $\operatorname{Im} z > 0$, so that $|q| < 1$. The relation between these two objects is that the number of points of E over $\mathbb{F}_p$ is equal to $1 + p - a_p$, for $p \neq 11$. This strange function f is a modular form, and is actually cooked up from the slightly easier function

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

by

$$f(z) = (\eta(z)\eta(11z))^2.$$

This function η is called the *Dedekind eta function*, and is one of the simplest examples of a modular forms (in the sense that we can write it down easily). This satisfies the following two identities:

$$\eta(z+1) = e^{i\pi/12}\eta(z), \quad \eta\left(\frac{-1}{z}\right) = \sqrt{\frac{z}{i}}\eta(z).$$

The first is clear, and the second takes some work to show. These transformation laws are exactly what makes this thing a modular form.

Another way to link E and f is via the *L-series*

$$L(E, s) = \sum_{n=1}^{\infty}\frac{a_n}{n^s},$$

which is a generalization of the Riemann ζ-function

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

We are in fact not going to study elliptic curves, as there is another course on that, but we are going to study the modular forms and these L-series. We are going to do this in a fairly classical way, without using algebraic number theory.

1 Some preliminary analysis

1.1 Characters of abelian groups

When we were young, we were forced to take some "applied" courses, and learnt about these beasts known as Fourier transforms. At that time, the hope was that we can leave them for the engineers, and never meet them ever again. Unfortunately, it turns out Fourier transforms are also important in "pure" mathematics, and we must understand them well.

Let's recall how Fourier transforms worked. We had two separate but closely related notions. First of all, we can take the Fourier transform of a function $f : \mathbb{R} \to \mathbb{C}$. The idea is that we wanted to write any such function as

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

One way to think about this is that we are expanding f in the basis $\chi_y(x) = e^{2\pi i y x}$. We also could take the Fourier series of a periodic function, i.e. a function defined on $\mathbb{R}/\mathbb{Z}$. In this case, we are trying to write our function as

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x}.$$

In this case, we are expanding our function in the basis $\chi_n(x) = e^{2\pi i n x}$. What is special about these basis $\{\chi_y\}$ and $\{\chi_y\}$?

We observe that $\mathbb{R}$ and $\mathbb{R}/\mathbb{Z}$ are not just topological spaces, but in fact abelian topological groups. These χ_y and χ_n are not just functions to $\mathbb{C}$, but continuous group homomorphisms to $\mathrm{U}(1) \subseteq \mathbb{C}$. In fact, these give *all* continuous group homomorphisms from $\mathbb{R}$ and $\mathbb{R}/\mathbb{Z}$ to $\mathrm{U}(1)$.

Definition (Character)**.** Let G be an abelian topological group. A (unitary) *character* of G is a continuous homomorphism $\chi : G \to \mathrm{U}(1)$, where $\mathrm{U}(1) = \{z \in \mathbb{C} \mid |z| = 1\}$.

To better understand Fourier transforms, we must understand characters, and we shall spend some time doing so.

Example. For any group G, there is the trivial character $\chi_0(g) \equiv 1$.

Example. The product of two characters is a character.

Example. If χ is a character, then so is χ^*, and $\chi\chi^* = 1$.

Thus, we see that the collection of all characters form a group under multiplication.

Definition (Character group)**.** Let G be a group. The *character group* (or *Pontryagin dual*) $\hat{G}$ is the group of all characters of G.

It is usually not hard to figure out what the character group is.

Example. Let $G = \mathbb{R}$. For $y \in \mathbb{R}$, we let $\chi_y : \mathbb{R} \to \mathrm{U}(1)$ be

$$\chi_y(x) = e^{2\pi i x y}.$$

For each $y \in \mathbb{R}$, this is a character, and all characters are of this form. So $\hat{\mathbb{R}} \cong \mathbb{R}$ under this correspondence.

Example. Take $G = \mathbb{Z}$ with the discrete topology. A character is uniquely determined by the image of 1, and any element of U(1) can be the image of 1. So we have $\hat{G} \cong \mathrm{U}(1)$.

Example. Take $G = \mathbb{Z}/N\mathbb{Z}$. Then the character is again determined by the image of 1, and the allowed values are exactly the Nth roots of unity. So

$$\hat{G} \cong \mu_N = \{\zeta \in \mathbb{C}^\times : \zeta^N = 1\}.$$

Example. Let $G = G_1 \times G_2$. Then $\hat{G} \cong \hat{G}_1 \times \hat{G}_2$. So, for example, $\hat{\mathbb{R}^n} = \mathbb{R}^n$. Under this correspondence, a $y \in \mathbb{R}^n$ corresponds to

$$\chi_y(x) = e^{2\pi x \cdot y}.$$

Example. Take $G = \mathbb{R}^\times$. We have

$$G \cong \{\pm 1\} \times \mathbb{R}^\times_{>0} \cong \{\pm 1\} \times \mathbb{R},$$

where we have an isomorphism between $\mathbb{R}^\times_{>0} \cong \mathbb{R}$ by the exponential map. So we have

$$\hat{G} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{R}.$$

Explicitly, given $(\varepsilon, \sigma) \in \mathbb{Z}/2\mathbb{Z} \times \mathbb{R}$, then character is given by

$$x \mapsto \mathrm{sgn}(x)^\varepsilon |x|^{i\sigma}.$$

Note that $\hat{G}$ has a natural topology for which the evaluation maps $(\chi \in \hat{G}) \mapsto \chi(g) \in \mathrm{U}(1)$ are all continuous for all g. Moreover, evaluation gives us a map $G \to \hat{\hat{G}}$.

Theorem (Pontryagin duality)**.** *Pontryagin duality* If G is locally compact, then $G \to \hat{\hat{G}}$ is an isomorphism.

Since this is a course on number theory, and not topological groups, we will not prove this.

Proposition. Let G be a finite abelian group. Then $|\hat{G}| = |G|$, and G and $\hat{G}$ are in fact isomorphic, but not canonically.

Proof. By the classification of finite abelian groups, we know G is a product of cyclic groups. So it suffices to prove the result for cyclic groups $\mathbb{Z}/N\mathbb{Z}$, and the result is clear since

$$\widehat{\mathbb{Z}/N\mathbb{Z}} = \mu_N \cong \mathbb{Z}/N\mathbb{Z}.$$

□

1.2 Fourier transforms

Equipped with the notion of characters, we can return to our original goal of understand Fourier transforms. We shall first recap the familiar definitions of Fourier transforms in specific cases, and then come up with the definition of Fourier transforms in full generality. In the mean time, we will get some pesky analysis out of the way.

Definition (Fourier transform)**.** Let $f : \mathbb{R} \to \mathbb{C}$ be an L^1 function, i.e. $\int |f| \, \mathrm{d}x < \infty$. The *Fourier transform* is

$$\hat{f}(y) = \int_{-\infty}^{\infty} e^{-2\pi i x y} f(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} \chi_y(x)^{-1} f(x) \, \mathrm{d}x.$$

This is a bounded and continuous function on $\mathbb{R}$.

We will see that the "correct" way to think about the Fourier transform is to view it as a function on $\hat{\mathbb{R}}$ instead of $\mathbb{R}$.

In general, there is not much we can say about how well-behaved $\hat{f}$ will be. In particular, we cannot expect the "Fourier inversion theorem" to hold for general L^1 functions. If we like analysis, then we can figure out exactly how much we need to assume about $\hat{f}$. But we don't. We chicken out and only consider functions that decay rely fast at infinity. This makes our life much easier.

Definition (Schwarz space)**.** The *Schwarz space* is defined by

$$\mathcal{S}(\mathbb{R}) = \{f \in C^\infty(\mathbb{R}) : x^n f^{(k)}(x) \to 0 \text{ as } x \to \pm\infty \text{ for all } k, n \geq 0\}.$$

Example. The function

$$f(x) = e^{-\pi x^2}.$$

is in the Schwarz space.

One can prove the following:

Proposition. If $f \in \mathcal{S}(\mathbb{R})$, then $\hat{f} \in \mathcal{S}(\mathbb{R})$, and the *Fourier inversion formula*

$$\hat{\hat{f}} = f(-x)$$

holds.

Everything carries over when we replace $\mathbb{R}$ with $\mathbb{R}^n$, as long as we promote both x and y into vectors.

We can also take the Fourier transform of functions defined on $G = \mathbb{R}/\mathbb{Z}$. For $n \in \mathbb{Z}$, we let $\chi_n \in \hat{G}$ by

$$\chi_n(x) = e^{2\pi i n x}.$$

These are exactly all the elements of $\hat{G}$, and $\hat{G} \cong \mathbb{Z}$. We then define the Fourier coefficients of a periodic function $f : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ by

$$c_n(f) = \int_0^1 e^{-2\pi i n x} f(x) \, \mathrm{d}x = \int_{\mathbb{R}/\mathbb{Z}} \chi_n(x)^{-1} f(x) \, \mathrm{d}x.$$

Again, under suitable regularity conditions on f, e.g. if $f \in C^\infty(\mathbb{R}/\mathbb{Z})$, we have

Proposition.

$$f(x) = \sum_{n \in \mathbb{Z}} c_n(f) e^{2\pi i n x} = \sum_{n \in \mathbb{Z} \cong \hat{G}} c_n(f) \chi_n(x).$$

This is the Fourier inversion formula for $G = \mathbb{R}/\mathbb{Z}$.

Finally, in the case when $G = \mathbb{Z}/N\mathbb{Z}$, we can define

Definition (Discrete Fourier transform)**.** Given a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we define the *Fourier transform* $\hat{f} : \mu_N \to \mathbb{C}$ by

$$\hat{f}(\zeta) = \sum_{a \in \mathbb{Z}/N\mathbb{Z}} \zeta^{-a} f(a).$$

This time there aren't convergence problems to worry with, so we can quickly prove this result:

Proposition. For a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we have

$$f(x) = \frac{1}{N} \sum_{\zeta \in \mu_N} \zeta^x \hat{f}(\zeta).$$

Proof. We see that both sides are linear in f, and we can write each function f as

$$f = \sum_{a \in \mathbb{Z}/N\mathbb{Z}} f(a)\delta_a,$$

where

$$\delta_a(x) = \begin{cases} 1 & x = a \\ 0 & x \neq a \end{cases}.$$

So we wlog $f = \delta_a$. Thus we have

$$\hat{f}(\zeta) = \zeta^{-a},$$

and the RHS is

$$\frac{1}{N} \sum_{\zeta \in \mu_N} \zeta^{x-a}.$$

We now note the fact that

$$\sum_{\zeta \in \mu_N} \zeta^k = \begin{cases} N & k \equiv 0 \pmod{N} \\ 0 & \text{otherwise} \end{cases}.$$

So we know that the RHS is equal to δ_a, as desired. □

It is now relatively clear what the general picture should be, except that we need a way to integrate functions defined on an abelian group. Since we are not doing analysis, we shall not be very precise about what we mean:

Definition (Haar measure)**.** Let G be a topological group. A *Haar measure* is a left translation-invariant Borel measure on G satisfying some regularity conditions (e.g. being finite on compact sets).

Theorem. Let G be a locally compact abelian group G. Then there is a Haar measure on G, unique up to scaling.

Example. On $G = \mathbb{R}$, the Haar measure is the usual Lebesgue measure.

Example. If G is discrete, then the Haar measure is the counting measure, so that

$$\int f \, \mathrm{d}g = \sum_{g \in G} f(g).$$

Example. If $G = \mathbb{R}_{>0}^{\times}$, then the integral given by the Haar measure is

$$\int f(x) \frac{\mathrm{d}x}{x},$$

since $\frac{\mathrm{d}x}{x}$ is invariant under multiplication of x by a constant.

Now we can define the general Fourier transform.

Definition (Fourier transform)**.** Let G be a locally compact abelian group with a Haar measure $\mathrm{d}g$, and let $f : G \to \mathbb{C}$ be a continuous L^1 function. The *Fourier transform* $\hat{f} : \hat{G} \to \mathbb{C}$ is given by

$$\hat{f}(\chi) = \int_G \chi(g)^{-1} f(g) \, \mathrm{d}g.$$

It is possible to prove the following theorem:

Theorem (Fourier inversion theorem)**.** Given a locally compact abelian group G with a fixed Haar measure, there is some constant C such that for "suitable" $f : G \to \mathbb{C}$, we have

$$\hat{\hat{f}}(g) = Cf(-g),$$

using the canonical isomorphism $G \to \hat{\hat{G}}$.

This constant is necessary, because the measure is only defined up to a multiplicative constant.

One of the most important results of this investigation is the following result:

Theorem (Poisson summation formula)**.** Let $f \in \mathcal{S}(\mathbb{R}^n)$. Then

$$\sum_{a \in \mathbb{Z}^n} f(a) = \sum_{b \in \mathbb{Z}^n} \hat{f}(b).$$

Proof. Let

$$g(x) = \sum_{a \in \mathbb{Z}^n} f(x + a).$$

This is now a function that is invariant under translation of $\mathbb{Z}^n$. It is easy to check this is a well-defined C^∞ function on $\mathbb{R}^n/\mathbb{Z}^n$, and so has a Fourier series. We write

$$g(x) = \sum_{b \in \mathbb{Z}^n} c_b(g) e^{2\pi i b \cdot x},$$

with

$$c_b(g) = \int_{\mathbb{R}^n/\mathbb{Z}^n} e^{-2\pi i b \cdot x} g(x) \, \mathrm{d}x = \sum_{a \in \mathbb{Z}^n} \int_{[0,1]^n} e^{2\pi i b \cdot x} f(x + a) \, \mathrm{d}x.$$

We can then do a change of variables $x \mapsto x - a$, which does not change the exponential term, and get that

$$c_b(g) = \int_{\mathbb{R}^n} e^{-2\pi i b\cdot x} f(x)\, \mathrm{d}x = \hat{f}(b).$$

Finally, we have

$$\sum_{a\in\mathbb{Z}^n} f(a) = g(0) = \sum_{b\in\mathbb{Z}^n} c_b(x) = \sum_{b\in\mathbb{Z}^n} \hat{f}(b).$$

□

1.3 Mellin transform and Γ-function

We unfortunately have a bit more analysis to do, which we will use a lot later on. This is the *Mellin transform*.

Definition (Mellin transform)**.** Let $f : \mathbb{R}_{>0} \to \mathbb{C}$ be a function. We define

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

whenever this converges.

We want to think of this as an analytic function of s. The following lemma tells us when we can actually do so

Lemma. Suppose $f : \mathbb{R}_{>0} \to \mathbb{C}$ is such that

- $y^N f(y) \to 0$ as $y \to \infty$ for all $N \in \mathbb{Z}$
- there exists m such that $|y^m y(f)|$ is bounded as $y \to 0$

Then $M(f, s)$ converges and is an analytic function of s for $\mathrm{Re}(s) > m$.

The conditions say f is rapidly decreasing at ∞ and has moderate growth at 0.

Proof. We know that for any $0 < r < R < \infty$, the integral

$$\int_r^R y^s f(y)\, \frac{\mathrm{d}y}{y}$$

is analytic for all s since f is continuous.

By assumption, we know $\int_R^\infty \to 0$ as $R \to \infty$ uniformly on compact subsets of $\mathbb{C}$. So we know

$$\int_r^\infty y^s f(y)\, \frac{\mathrm{d}y}{y}$$

converges uniformly on compact subsets of $\mathbb{C}$.

On the other hand, the integral $\int_0^r$ as $r \to 0$ converges uniformly on compact subsets of $\{s \in \mathbb{C} : \mathrm{Re}(s) > m\}$ by the other assumption. So the result follows. □

This transform might seem a bit strange, but we can think of this as an analytic continuation of the Fourier transform.

Example. Suppose we are in the rather good situation that

$$\int_0^\infty |f| \, \frac{\mathrm{d}y}{y} < \infty.$$

In practice, this will hardly ever be the case, but this is a good place to start exploring. In this case, the integral actually converges on $i\mathbb{R}$, and equals the Fourier transform of $f \in L^1(G) = L^1(\mathbb{R}^\times_{>0})$. Indeed, we find

$$\hat{G} = \{y \mapsto y^{i\sigma} : \sigma \in \mathbb{R}\} \cong \mathbb{R},$$

and $\frac{\mathrm{d}y}{y}$ is just the invariant measure on G. So the formula for the Mellin transform is exactly the formula for the Fourier transform, and we can view the Mellin transform as an analytic continuation of the Fourier transform.

We now move on to explore properties of the Mellin transform. When we make a change of variables $y \leftrightarrow \alpha y$, by inspection of the formula, we find

Proposition.

$$M(f(\alpha y), s) = \alpha^{-s} M(f, s)$$

for $\alpha > 0$.

The following is a very important example of the Mellin transform:

Definition (Γ function)**.** The Γ *function* is the Mellin transform of

$$f(y) = e^{-y}.$$

Explicitly, we have

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

By general theory, we know this is analytic for $\mathrm{Re}(s) > 0$.
If we just integrate by parts, we find

$$\Gamma(s) = \int_0^\infty e^{-y} y^{s-1} \, \mathrm{d}y = \left[e^{-y} \frac{y^s}{s}\right]_0^\infty + \frac{1}{s} \int_0^\infty e^{-y} y^s \, \mathrm{d}y = \frac{1}{s}\Gamma(s+1).$$

So we find that

Proposition.

$$s\Gamma(s) = \Gamma(s+1).$$

Moreover, we can compute

$$\Gamma(1) = \int_0^\infty e^{-y} \, \mathrm{d}y = 1.$$

So we get

Proposition. For an integer $n \geq 1$, we have

$$\Gamma(n) = (n-1)!.$$

In general, iterating the above formula, we find

$$\Gamma(s) = \frac{1}{s(s+1)\cdots(s+N-1)}\Gamma(s+N).$$

Note that the right hand side makes sense for $\mathrm{Re}(s) > -N$ (except at non-positive integer points). So this allows us to extend $\Gamma(s)$ to a meromorphic function on $\{\mathrm{Re}(s) > -N\}$, with simple poles at $0, 1, \cdots, 1-N$ of residues

$$\operatorname*{res}_{s=1-N} \Gamma(s) = \frac{(-1)^{N-1}}{(N-1)!}.$$

Of course, since N was arbitrary, we know $\Gamma(s)$ extends to a meromorphic function on $\mathbb{C} \setminus \mathbb{Z}_{\leq 0}$.

Here are two facts about the Γ function that we are not going to prove, because, even if the current experience might suggest otherwise, this is not an analysis course.

Proposition.

(i) The *Weierstrass product*: We have

$$\Gamma(s)^{-1} = e^{\gamma s} s \prod_{n\geq 1} \left(1 + \frac{s}{n}\right) e^{-s/n}$$

for all $s \in \mathbb{C}$. In particular, $\Gamma(s)$ is never zero. Here γ is the *Euler-Mascheroni constant*, given by

$$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right).$$

(ii) *Duplication and reflection formulae*:

$$\pi^{\frac{1}{2}}\Gamma(2s) = 2^{2s-1}\Gamma(s)\Gamma\left(s + \frac{1}{2}\right)$$

and

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi z}.$$

The main reason why we care about the Mellin transform in this course is that a lot of *Dirichlet series* are actually Mellin transforms of some functions. Suppose we have a Dirichlet series

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

where the a_n grow not too quickly. Then we can write

$$\begin{aligned}(2\pi)^{-s}\Gamma(s)\sum_{n=1}^{\infty}\frac{a_n}{n^s} &= \sum_{n=1}^{\infty} a_n (2\pi n)^{-s} M(e^{-y}, s)\\ &= \sum_{n=1}^{\infty} M(e^{-2\pi n y}, s)\\ &= M(f, s),\end{aligned}$$

where we set

$$f(y) = \sum_{n \geq 1} a_n e^{-2\pi n y}.$$

Since we know about the analytic properties of the Γ function, by understanding $M(f, s)$, we can deduce useful properties about the Dirichlet series itself.

2 Riemann ζ-function

We ended the previous chapter by briefly mentioning Dirichlet series. The first and simplest example one can write down is the Riemann ζ-function.

Definition (Riemann ζ-function)**.** The *Riemann ζ-function* is defined by

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

for $\mathrm{Re}(s) > 1$.

This ζ-function is related to prime numbers by the following famous formula:

Proposition (Euler product formula)**.** We have

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

Proof. Euler's proof was purely formal, without worrying about convergence. We simply note that

$$\prod_{p \text{ prime}} \frac{1}{1 - p^{-s}} = \prod_{p} (1 + p^{-s} + (p^2)^{-s} + \cdots) = \sum_{n \geq 1} n^{-s},$$

where the last equality follows by unique factorization in $\mathbb{Z}$. However, to prove this properly, we need to be a bit more careful and make sure things converge.

Saying the infinite product $\prod_p$ convergence is the same as saying $\sum p^{-s}$ converges, by basic analysis, which is okay since we know $\zeta(s)$ converges absolutely when $\mathrm{Re}(s) > 1$. Then we can look at the difference

$$\begin{aligned} \zeta(s) - \prod_{p \leq X} \frac{1}{1 - p^{-s}} &= \zeta(s) - \prod_{p \leq X} (1 + p^{-s} + p^{-2s} + \cdots) \\ &= \prod_{n \in \mathcal{N}_X} n^{-s}, \end{aligned}$$

where $\mathcal{N}_X$ is the set of all $n \geq 1$ such that at least one prime factor is $\geq X$. In particular, we know

$$\left| \zeta(s) - \prod_{p \leq X} \frac{1}{1 - p^{-s}} \right| \leq \sum_{n \geq X} |n^{-s}| \to 0$$

as $X \to \infty$. So the result follows. □

The Euler product formula is the beginning of the connection between the ζ-function and the distribution of primes. For example, as the product converges for $\mathrm{Re}(s) > 1$, we know in particular that $\zeta(s) \neq 0$ for all s when $\mathrm{Re}(s) > 1$. Whether or not $\mathrm{Re}(s)$ vanishes elsewhere is a less straightforward matter, and this involves quite a lot of number theory.

We will, however, not care about primes in this course. Instead, we look at some further analytic properties of the ζ function. To do so, we first write it as a Mellin transform.

Theorem. If $\mathrm{Re}(s) > 1$, then

$$(2\pi)^{-s}\Gamma(s)\zeta(s) = \int_0^\infty \frac{y^s}{e^{2\pi y} - 1}\frac{\mathrm{d}y}{y} = M(f, s),$$

where

$$f(y) = \frac{1}{e^{2\pi y} - 1}.$$

This is just a simple computation.

Proof. We can write

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

for $y > 0$.

As $y \to 0$, we find

$$f(y) \sim \frac{1}{2\pi y}.$$

So when $\mathrm{Re}(s) > 1$, the Mellin transform converges, and equals

$$\sum_{n\geq 1} M(e^{-2\pi n y}, s) = \sum_{n\geq 1} (2\pi n)^{-s} M(e^{-y}, s) = (2\pi)^{-s}\Gamma(s)\zeta(s).$$

□

Corollary. $\zeta(s)$ has a meromorphic continuation to $\mathbb{C}$ with a simple pole at $s = 1$ as its only singularity, and

$$\operatorname*{res}_{s=1} \zeta(s) = 1.$$

Proof. We can write

$$M(f, s) = M_0 + M_\infty = \left(\int_0^1 + \int_1^\infty\right) \frac{y^s}{e^{2\pi y} - 1}\frac{\mathrm{d}y}{y}.$$

The second integral M_∞ is convergent for all $s \in \mathbb{C}$, hence defines a holomorphic function.

For any fixed N, we can expand

$$f(y) = \sum_{n=-1}^{N-1} c_n y^n + y^N g_N(y)$$

for some $g \in C^\infty(\mathbb{R})$, as f has a simple pole at $y = 0$, and

$$c_{-1} = \frac{1}{2\pi}.$$

So for $\mathrm{Re}(s) > 1$, we have

$$\begin{aligned} M_0 &= \sum_{n=-1}^{N-1} c_n \int_0^1 y^{n+s-1}\,\mathrm{d}y + \int_0^N y^{N+s-1} g_N(y)\,\mathrm{d}y \\ &= \sum_{n=-1}^{N-1} \frac{c_n}{s+n} y^{s+n} + \int_0^1 g_N(y) y^{s+N-1}\,\mathrm{d}y. \end{aligned}$$

We now notice that this formula makes sense for $\mathrm{Re}(s) > -N$. Thus we have found a meromorphic continuation of

$$(2\pi)^{-s}\Gamma(s)\zeta(s)$$

to $\{\mathrm{Re}(s) > N\}$, with at worst simple poles at $s = 1 - N, 2 - N, \cdots, 0, 1$. Also, we know $\Gamma(s)$ has a simple pole at $s = 0, -1, -2, \cdots$. So $\zeta(s)$ is analytic at $s = 0, -1, -2, \cdots$. Since $c_{-1} = \frac{1}{2\pi}$ and $\Gamma(1) = 1$, we get

$$\operatorname*{res}_{s=1} \zeta(s) = 1.$$

□

Now we note that by the Euler product formula, if there are only finitely many primes, then $\zeta(s)$ is certainly analytic everywhere. So we deduce

Corollary. There are infinitely many primes.

Given a function ζ, it is natural to ask what values it takes. In particular, we might ask what values it takes at integers. There are many theorems and conjectures concerning the values at integers of L-functions (which are Dirichlet series like the ζ-function). These properties relate to subtle number-theoretic quantities. For example, the values of $\zeta(s)$ at negative integers are closely related to the class numbers of the cyclotomic fields $\mathbb{Q}(\zeta_p)$. These are also related to early (partial) proofs of Fermat's last theorem, and things like the Birch–Swinnerton-Dyer conjecture on elliptic curves.

We shall take a tiny step by figuring out the values of $\zeta(s)$ at negative integers. They are given by the *Bernoulli numbers*.

Definition (Bernoulli numbers)**.** The *Bernoulli numbers* are defined by a generating function

$$\sum_{n=0}^{\infty} B_n \frac{t^n}{n!} = \frac{t}{e^t - 1} = \left(1 + \frac{t}{2!} + \frac{t^2}{3!} + \cdots\right)^{-1}.$$

Clearly, all Bernoulli numbers are rational. We can directly compute

$$B_0 = 1, \quad B_1 = -\frac{1}{2}, \cdots.$$

The first thing to note about this is the following:

Proposition. $B_n = 0$ if n is odd and $n \geq 3$.

Proof. Consider

$$f(t) = \frac{t}{e^t - 1} + \frac{t}{2} = \sum_{n \geq 0, n \neq 1} B_n \frac{t^n}{n!}.$$

We find that

$$f(t) = \frac{t}{2}\frac{e^t + 1}{e^t - 1} = f(-t).$$

So all the odd coefficients must vanish. □

Corollary. We have

$$\zeta(0) = B_1 = -\frac{1}{2}, \quad \zeta(1-n) = -\frac{B_n}{n}$$

for $n > 1$. In particular, for all $n \geq 1$ integer, we know $\zeta(1-n) \in \mathbb{Q}$ and vanishes if $n > 1$ is odd.

Proof. We know

$$(2\pi)^{-s}\Gamma(s)\zeta(s)$$

has a simple pole at $s = 1 - n$, and the residue is c_{n-1}, where

$$\frac{1}{e^{2\pi y} - 1} = \sum_{n \geq -1} c_n y^n.$$

So we know

$$c_{n-1} = (2\pi)^{n-1}\frac{B_n}{n!}.$$

We also know that

$$\operatorname*{res}_{s=1-n} \Gamma(s) = \frac{(-1)^{n-1}}{(n-1)!},$$

we get that

$$\zeta(1-n) = (-1)^{n-1}\frac{B_n}{n}.$$

If $n = 1$, then this gives $-\frac{1}{2}$. If n is odd but > 1, then this vanishes. If n is even, then this is $-\frac{B_n}{n}$, as desired. □

To end our discussion on the ζ-function, we shall prove a functional equation, relating $\zeta(s)$ to $\zeta(1-s)$. To do so, we relate the ζ-function to *another* Mellin transform. We define

$$\Theta(y) = \sum_{n \in \mathbb{Z}} e^{-\pi n^2 y} = 1 + 2\sum_{n \geq 1} e^{-\pi n^2 y}.$$

This is convergent for for $y > 0$. So we can write

$$\Theta(y) = \vartheta(iy),$$

where

$$\vartheta(z) = \sum_{n \in \mathbb{Z}} e^{\pi i n^2 z},$$

which is analytic for $|e^{\pi i z}| < 1$, i.e. $\operatorname{Im}(z) > 0$. This is *Jacobi's ϑ-function*. This function is also important in algebraic geometry, representation theory, and even applied mathematics. But we will just use it for number theory. We note that

$$\Theta(y) \to 1$$

as $y \to \infty$, so we can't take its Mellin transform. What we *can* do is

Proposition.

$$M\left(\frac{\Theta(y) - 1}{2}, \frac{s}{2}\right) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s).$$

The proof is again just do it.

Proof. The left hand side is

$$\sum_{n\geq 1} M\left(e^{-\pi n^2 y}, \frac{s}{2}\right) = \sum_{n\geq 1} (\pi n^2)^{-s/2} M\left(e^{-y}, \frac{s}{2}\right) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s).$$

□

To produce a functional equation for ζ, we first do it for Θ.

Theorem (Functional equation for Θ-function)**.** If $y > 0$, then

$$\Theta\left(\frac{1}{y}\right) = y^{1/2}\Theta(y), \qquad (*)$$

where we take the positive square root. More generally, taking the branch of $\sqrt{}$ which is positive real on the positive real axis, we have

$$\vartheta\left(-\frac{1}{z}\right) = \left(\frac{z}{i}\right)^{1/2}\vartheta(z).$$

The proof is rather magical.

Proof. By analytic continuation, it suffices to prove $(*)$. Let

$$g_t(x) = e^{-\pi t x^2} = g_1(t^{1/2}x).$$

In particular,

$$g_1(x) = e^{-\pi x^2}.$$

Now recall that $\hat{g}_1 = g_1$. Moreover, the Fourier transform of $f(\alpha x)$ is $\frac{1}{\alpha}\hat{f}(y/\alpha)$. So

$$\hat{g}_t(y) = t^{-1/2}\hat{g}_1(t^{-1/2}y) = t^{-1/2}g_1(t^{-1/2}y) = t^{-1/2}e^{-\pi y^2/t}.$$

We now apply the Poisson summation formula:

$$\Theta(t) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t} = \sum_{n\in\mathbb{Z}} g_t(n) = \sum_{n\in\mathbb{Z}} \hat{g}_t(n) = t^{-1/2}\Theta(1/t). \qquad \square$$

Before we continue, we notice that most of the time, when we talk about the Γ-function, there are factors of π floating around. So we can conveniently set

Notation.

$$\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(s/2).$$
$$\Gamma_{\mathbb{C}}(s) = 2(2\pi)^{-s}\Gamma(s)$$

These are the real/complex Γ-*factors*.

We also define

Notation.

$$Z(s) \equiv \Gamma_{\mathbb{R}}(s)\zeta(s) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s).$$

The theorem is then

Theorem (Functional equation for ζ-function)**.**

$$Z(s) = Z(1-s).$$

Moreover, $Z(s)$ is meromorphic, with only poles at $s = 1$ and 0.

Proof. For $\mathrm{Re}(s) > 1$, we have

$$\begin{aligned} 2Z(s) &= M\left(\Theta(y) - 1, \frac{s}{2}\right) \\ &= \int_0^\infty [\Theta(y) - 1] y^{s/2} \frac{\mathrm{d}y}{y} \\ &= \left(\int_0^1 + \int_1^\infty\right) [\Theta(y) - 1] y^{s/2} \frac{\mathrm{d}y}{y} \end{aligned}$$

The idea is that using the functional equation for the Θ-function, we can relate the $\int_0^1$ part and the $\int_1^\infty$ part. We have

$$\begin{aligned} \int_0^1 (\Theta(y) - 1) y^{s/2} \frac{\mathrm{d}y}{y} &= \int_0^1 (\Theta(y) - y^{-1/2}) y^{s/2} \frac{\mathrm{d}y}{y} + \int_0^1 \left(y^{\frac{s-1}{2}} - y^{1/2}\right) \frac{\mathrm{d}y}{y} \\ &= \int_0^1 (y^{-1/2}\Theta(1/y) - y^{-1/2}) \frac{\mathrm{d}y}{y} + \frac{2}{s-1} - \frac{2}{s}. \end{aligned}$$

In the first term, we change variables $y \leftrightarrow 1/y$, and get

$$= \int_1^\infty y^{1/2} (\Theta(y) - 1) y^{-s/2} \frac{\mathrm{d}y}{y} + \frac{2}{s-1} - \frac{2}{s}.$$

So we find that

$$2Z(s) = \int_1^\infty (\Theta(y) - 1)(y^{s/2} + y^{\frac{1-s}{2}}) \frac{\mathrm{d}y}{y} + \frac{2}{s-1} - \frac{2}{s} = 2Z(1-s).$$

Note that what we've done by separating out the $y^{\frac{s-1}{2}} - y^{s/2}$ term is that we separated out the two poles of our function. □

Later on, we will come across more L-functions, and we will prove functional equations in the same way.

Note that we can write

$$Z(s) = \Gamma_{\mathbb{R}}(s) \prod_{p \text{ primes}} \frac{1}{1 - p^{-s}},$$

and the term $\Gamma_{\mathbb{R}}(s)$ should be thought of as the Euler factor for $p = \infty$, i.e. the Archimedean valuation on $\mathbb{Q}$.

3 Dirichlet L-functions

We now move on to study a significant generalization of ζ-functions, namely Dirichlet L-functions. While these are generalizations of the ζ-function, it turns out the ζ function is a very particular kind of L-function. For example, most L-functions are actually analytic on all of $\mathbb{C}$, except for (finite multiples of) the ζ-function.

Recall that a Dirichlet series is a series of the form

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

A Dirichlet L-function is a Dirichlet series whose coefficients come from *Dirichlet characters*.

Definition (Dirichlet characters)**.** Let $N \geq 1$. A *Dirichlet character mod N* is a character $\chi : (\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}^\times$.

As before, we write $\widehat{(\mathbb{Z}/N\mathbb{Z})^\times}$ for the group of characters.

Note that in the special case $N = 1$, we have

$$\mathbb{Z}/N\mathbb{Z} = \{0 = 1\} = (\mathbb{Z}/N\mathbb{Z})^\times,$$

and so $\widehat{(\mathbb{Z}/N\mathbb{Z})^\times} \cong \{1\}$, and the only Dirichlet character is identically 1.

Not all characters are equal. Some are less exciting than others. Suppose χ is a character mod N, and we have some integer $d > 1$. Then we have the reduction mod N map

$$(\mathbb{Z}/Nd\mathbb{Z})^\times \twoheadrightarrow (\mathbb{Z}/N\mathbb{Z})^\times,$$

and we can compose χ with this to get a character mod Nd. This is a rather boring character, because the value of $x \in (\mathbb{Z}/Nd\mathbb{Z})^\times$ only depends on the value of x mod N.

Definition (Primitive character)**.** We say a character $\chi \in \widehat{(\mathbb{Z}/n\mathbb{Z})^\times}$ is *primitive* if there is no $M < N$ with $M \mid N$ with $\chi' \in \widehat{(\mathbb{Z}/M\mathbb{Z})^\times}$ such that

$$\chi = \chi' \circ (\text{reduction mod } M).$$

Similarly we define

Definition (Equivalent characters)**.** We say characters $\chi_1 \in \widehat{(\mathbb{Z}/N_1\mathbb{Z})^\times}$ and $\chi_2 \in \widehat{(\mathbb{Z}/N_2\mathbb{Z})^\times}$ are *equivalent* if for all $x \in \mathbb{Z}$ such that $(x, N_1N_2) = 1$, we have

$$\chi_1(x \bmod N_1) = \chi_2(x \bmod N_2).$$

It is clear that if we produce a new character from an old one via reduction mod Nd, then they are equivalent.

One can show the following:

Proposition. If $\chi \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}$, then there exists a unique $M \mid N$ and a *primitive* $\chi_* \in \widehat{(\mathbb{Z}/M\mathbb{Z})^\times}$ that is equivalent to χ.

Definition (Conductor)**.** The *conductor* of a character χ is the unique $M \mid N$ such that there is a *primitive* $\chi_* \in \widehat{(\mathbb{Z}/M\mathbb{Z})^\times}$ that is equivalent to χ.

Example. Take

$$\chi = \chi_0 \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times},$$

given by $\chi_0(x) \equiv 1$. If $N > 1$, then χ_0 is not primitive, and the associated primitive character is the trivial character modulo $M = 1$. So the conductor is 1.

Using these Dirichlet characters, we can define *Dirichlet L-series*:

Definition (Dirichlet L-series)**.** Let $\chi \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}$ be a Dirichlet character. The *Dirichlet L-series* of χ is

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

Since $|\chi(n)| = 1$, we again know this is absolutely convergent for $\mathrm{Re}(s) > 1$.

As $\chi(mn) = \chi(m)\chi(n)$ whenever $(mn, N) = 1$, we get the same Euler product as we got for the ζ-function:

Proposition.

$$L(\chi, s) = \prod_{\text{prime } p \nmid N} \frac{1}{1 - \chi(p)p^{-s}}.$$

The proof of convergence is again similar to the case of the ζ-function.

It follows that

Proposition. Suppose $M \mid N$ and $\chi_M \in \widehat{(\mathbb{Z}/M\mathbb{Z})^\times}$ and $\chi_N \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}$ are equivalent. Then

$$L(\chi_M, s) = \prod_{\substack{p \nmid M \\ p \mid N}} \frac{1}{1 - \chi_M(p)p^{-s}} L(\chi_N, s).$$

In particular,

$$\frac{L(\chi_M, s)}{L(\chi_N, s)} = \prod_{\substack{p \nmid M \\ p \mid N}} \frac{1}{1 - \chi_M(p)p^{-s}}$$

is analytic and non-zero for $\mathrm{Re}(s) > 0$.

We'll next show that $L(\chi, s)$ has a meromorphic continuation to $\mathbb{C}$, and is analytic unless $\chi = \chi_0$.

Theorem.

(i) $L(\chi, s)$ has a meromorphic continuation to $\mathbb{C}$, which is analytic except for at worst a simple pole at $s = 1$.

(ii) If $\chi \neq \chi_0$ (the trivial character), then $L(\chi, s)$ is analytic everywhere. On the other hand, $L(\chi_0, s)$ has a simple pole with residue

$$\frac{\varphi(N)}{N} = \prod_{p \mid N} \left(1 - \frac{1}{p}\right),$$

where φ is the Euler function.

Proof. More generally, let $\phi : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be any N-periodic function, and let

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

Then

$$(2\pi)^{-s}\Gamma(s)L(\phi, s) = \sum_{n=1}^{\infty} \phi(n) M(e^{-2\pi n y}, s) = M(f(y), s),$$

where

$$f(y) = \sum_{n\geq 1} \phi(n) e^{-2\pi n y}.$$

We can then write

$$f(y) = \sum_{n=1}^{N}\sum_{r=0}^{\infty} \phi(n) e^{-2\pi(n+rN)y} = \sum_{n=1}^{N} \phi(n) \frac{e^{-2\pi n y}}{1 - e^{-2\pi N y}} = \sum_{n=1}^{N} \phi(n) \frac{e^{2\pi(N-n)y}}{e^{2\pi N y} - 1}.$$

As $0 \leq N - n < N$, this is $O(e^{-2\pi y})$ as $y \to \infty$. Copying for $\zeta(s)$, we write

$$M(f, s) = \left(\int_0^1 + \int_1^{\infty}\right) f(y) y^s \, \frac{\mathrm{d}y}{y} \equiv M_0(s) + M_{\infty}(s).$$

The second term is analytic for all $s \in \mathbb{C}$, and the first term can be written as

$$M_0(s) = \sum_{n=1}^{N} \phi(n) \int_0^1 \frac{e^{2\pi(N-n)y}}{e^{2\pi N y} - 1} y^s \, \frac{\mathrm{d}y}{y}.$$

Now for any L, we can write

$$\frac{e^{2\pi(N-n)y}}{e^{2\pi N y} - 1} = \frac{1}{2\pi N y} + \sum_{r=0}^{L-1} c_{r,n} y^r + y^L g_{L,n}(y)$$

for some $g_{L,n}(y) \in C^{\infty}[0, 1]$. Hence we have

$$M_0(s) = \sum_{n=1}^{N} \phi(n) \left(\int_0^1 \frac{1}{2\pi N y} y^s \frac{\mathrm{d}y}{y} + \int_0^1 \sum_{r=0}^{L-1} c_{r,n} y^{r+s-1} \, \mathrm{d}y\right) + G(s),$$

where $G(s)$ is some function analytic for $\mathrm{Re}(s) > -L$. So we see that

$$(2\pi)^{-s}\Gamma(s)L(\phi, s) = \sum_{n=1}^{N} \phi(n) \left(\frac{1}{2\pi N(s-1)} + \frac{c_{0,n}}{s} + \cdots + \frac{c_{L-1,n}}{s + L - 1}\right) + G(s).$$

As $\Gamma(s)$ has poles at $s = 0, -1, \cdots$, this cancels with all the poles apart from the one at $s = 1$.

The first part then follows from taking

$$\phi(n) = \begin{cases} \chi(n) & (n, N) = 1 \\ 0 & (n, N) \geq 1 \end{cases}.$$

By reading off the formula, since $\Gamma(1) = 1$, we know

$$\operatorname*{res}_{s=1} L(\chi, s) = \frac{1}{N}\sum_{n=1}^{N} \phi(n).$$

If $\chi \neq \chi_0$, then this vanishes by the orthogonality of characters. Otherwise, it is $|(\mathbb{Z}/N\mathbb{Z})^\times|/N = \varphi(N)/N$. □

Note that this is consistent with the result

$$L(\chi_0, s) = \prod_{p|N}(1 - p^{-s})\zeta(s).$$

So for a non-trivial character, our L-function doesn't have a pole.

The next big theorem is that in fact $L(\chi, 1)$ is non-zero. In number theory, there are lots of theorems of this kind, about non-vanishing of L-functions at different points.

Theorem. If $\chi \neq \chi_0$, then $L(\chi, 1) \neq 0$.

Proof. The trick is the consider all characters together. We let

$$\zeta_N(s) = \prod_{\chi \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}} L(\chi, s) = \prod_{p\nmid N}\prod_{\chi}(1 - \chi(p)p^{-s})^{-1}$$

for $\operatorname{Re}(s) > 1$. Now we know $L(\chi_0, s)$ has a pole at $s = 1$, and is analytic everywhere else. So if any other $L(\chi, 1) = 0$, then $\zeta_N(s)$ is analytic on $\operatorname{Re}(s) > 0$. We will show that this cannot be the case.

We begin by finding a nice formula for the product of $(1 - \chi(p)p^{-s})^{-1}$ over all characters.

Claim. If $p \nmid N$, and T is any complex number, then

$$\prod_{\chi \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}} (1 - \chi(p)T) = (1 - T^{f_p})^{\varphi(N)/f_p},$$

where f_p is the order of p in $(\mathbb{Z}/n\mathbb{Z})^\times$.

So

$$\zeta_N(s) = \prod_{p\nmid N}(1 - p^{-f_p s})^{-\varphi(N)/f_p}.$$

To see this, we write $f = f_p$, and, for convenience, write

$$G = (\mathbb{Z}/N\mathbb{Z})^\times$$
$$H = \langle p\rangle \subseteq G.$$

We note that $\hat{G}$ naturally contains $\widehat{G/H} = \{\chi \in \hat{G} : \chi(p) = 1\}$ as a subgroup. Also, we know that

$$|\widehat{G/H}| = |G/H| = \varphi(N)/f.$$

Also, the restriction map

$$\frac{\hat{G}}{\widehat{G/H}} \to \hat{H}$$

is obviously injective, hence an isomorphism by counting orders. So

$$\prod_{\chi\in\hat{G}}(1-\chi(p)T)=\prod_{\chi\in\hat{H}}(1-\chi(p)T)^{\varphi(N)/f}=\prod_{\zeta\in\mu_f}(1-\zeta T)^{\varphi(N)/f}=(1-T^f)^{\varphi(N)/f}.$$

We now notice that when we expand the product of ζ_N, at least formally, then we get a Dirichlet series with non-negative coefficients. We now prove the following peculiar property of such Dirichlet series:

Claim. Let

$$D(s)=\sum_{n\geq 1}a_n n^{-s}$$

be a Dirichlet series with real $a_n\geq 0$, and suppose this is absolutely convergent for $\mathrm{Re}(s)>\sigma>0$. Then if $D(s)$ can be analytically continued to an analytic function $\tilde{D}$ on $\{\mathrm{Re}(s)>0\}$, then the series converges for all real $s>0$.

Let $\rho>\sigma$. Then by the analytic continuation, we have a convergent Taylor series on $\{|s-\rho|<\rho\}$

$$D(s)=\sum_{k\geq 0}\frac{1}{k!}D^{(k)}(\rho)(s-\rho)^k.$$

Moreover, since $\rho>\sigma$, we can directly differentiate the Dirichlet series to obtain the derivatives:

$$D^{(k)}(\rho)=\sum_{n\geq 1}a_n(-\log n)^k n^{-\rho}.$$

So if $0<x<\rho$, then

$$D(x)=\sum_{k\geq 0}\frac{1}{k!}(p-x)^k\left(\sum_{n\geq 1}a_n(\log n)^k n^{-\rho}\right).$$

Now note that all terms in this sum are all non-negative. So the double series has to converge absolutely as well, and thus we are free to rearrange the sum as we wish. So we find

$$\begin{aligned}D(x)&=\sum_{n\geq 1}a_n n^{-\rho}\sum_{k\geq 0}\frac{1}{k!}(\rho-x)^k(\log n)^l\\&=\sum_{n\geq 1}a_n n^{-\rho}e^{(\rho-x)\log n}\\&=\sum_{n\geq 1}a_n n^{-\rho}n^{\rho-x}\\&=\sum_{n\geq 1}a_n n^{-x},\end{aligned}$$

as desired.

Now we are almost done, as

$$\zeta_N(s)=L(\chi_0,s)\prod_{\chi\neq\chi_0}L(\chi,s).$$

We saw that $L(\chi_0, s)$ has a simple pole at $s = 1$, and the other terms are all holomorphic at $s = 1$. So if some $L(\chi, 1) = 0$, then $\zeta_N(s)$ is holomorphic for $\mathrm{Re}(s) > 0$ (and in fact everywhere). Since the Dirichlet series of η_N has ≥ 0 coefficients, by the lemma, it suffices to find some point on $\mathbb{R}_{>0}$ where the Dirichlet series for ζ_N doesn't converge.

We notice

$$\zeta_N(x) = \prod_{p \nmid N} (1 + p^{-f_p x} + p^{-2f_p x} + \cdots)^{\varphi(N)/f_p} \geq \sum_{p \nmid N} p^{-\varphi(N)x}.$$

It now suffices to show that $\sum p^{-1} = \infty$, and thus the series for $\zeta_N(x)$ is not convergent for $x = \frac{1}{\varphi(N)}$.

Claim. We have

$$\sum_{p \text{ prime}} p^{-x} \sim -\log(x - 1)$$

as $x \to 1^+$. On the other hand, if $\chi \neq \chi_0$ is a Dirichlet character mod N, then

$$\sum_{p \nmid N} \chi(p) p^{-x}$$

is bounded as $x \to 1^+$.

Of course (and crucially, as we will see), the second part is not needed for the proof, but it is still nice to know.

To see this, we note that for any χ, we have

$$\log L(\chi, x) = \sum_{p \nmid N} -\log(1 - \chi(p)p^{-x}) = \sum_{p \nmid N} \sum_{r \geq 1} \frac{\chi(p)^r p^{-rx}}{r}.$$

So

$$\begin{aligned} \left| \log L(\chi, x) - \sum_{p \nmid N} \chi(p) p^{-x} \right| &< \sum_{p \nmid N} \sum_{r \geq 2} p^{-rx} \\ &= \sum_{p \nmid N} \frac{p^{-2x}}{1 - p^{-x}} \\ &\leq \sum_{n \geq 1} \frac{n^{-2}}{1/2}, \end{aligned}$$

which is a (finite) constant for $C < \infty$. When $\chi = \chi_0, N = 1$, then

$$\left| \log \zeta(x) - \sum_{p} p^{-x} \right|$$

is bounded as $x \to 1^+$. But we know

$$\zeta(s) = \frac{1}{s - 1} + O(s).$$

So we have

$$\sum_p p^{-x} \sim \log(x-1).$$

When $\chi \neq \chi_0$, then $L(\chi, 1) \neq 0$, as we have just proved! So $\log L(\chi, x)$ is bounded as $x \to 1^+$. and so we are done. □

Note that up to a finite number of factors in the Euler product (for $p \mid N$), this $\zeta_N(s)$ equals to the Dedekind ζ-function of the number field $K = \mathbb{Q}(\sqrt[n]{1})$, given by

$$\zeta_K(s) = \sum_{\text{ideals } 0 \neq I \subseteq \mathcal{O}_K} \frac{1}{(N(I))^s}.$$

We can now use what we've got to quickly prove Dirichlet's theorem:

Theorem (Dirichlet's theorem on primes in arithmetic progressions)**.** Let $a \in \mathbb{Z}$ be such that $(a, N) = 1$. Then there exists infinitely many primes $p \equiv a \pmod N$.

Proof. We want to show that the series

$$\sum_{\substack{p \text{ prime} \\ p \equiv a \bmod N}} p^{-x}$$

is unbounded as $x \to 1^+$, and in particular must be infinite. We note that for $(x, N) = 1$, we have

$$\sum_{\chi \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}} \chi(x) = \begin{cases} \varphi(N) & x \equiv 1 \pmod N \\ 0 & \text{otherwise} \end{cases},$$

since the sum of roots of unity vanishes. We also know that χ is a character, so $\chi(a)^{-1}\chi(p) = \chi(a^{-1}p)$. So we can write

$$\sum_{\substack{p \text{ prime} \\ p \equiv a \bmod N}} p^{-x} = \frac{1}{\varphi(N)} \sum_{\chi \in (\mathbb{Z}/N\mathbb{Z})^\times} \chi(a)^{-1} \sum_{\text{all } p} \chi(p) p^{-x},$$

Now if $\chi = \chi_0$, then the sum is just

$$\sum_{p \nmid N} p^{-x} \sim -\log(x-1)$$

as $x \to 1^+$. Moreover, all the other sums are bounded as $x \to 1^+$. So

$$\sum_{p \equiv a \bmod N} p^{-x} \sim -\frac{1}{\varphi(N)} \log(x-1).$$

So the whole sum must be unbounded as $x \to 1^+$. So in particular, the sum must be infinite. □

This actually tells us something more. It says

$$\frac{\sum\limits_{p\equiv a \bmod N} p^{-x}}{\sum\limits_{\text{all } p} p^{-x}} \sim \frac{1}{\varphi(N)}.$$

as $x \to 1^+$. So in some well-defined sense (namely *analytic density*), $\frac{1}{\varphi(N)}$ of the primes are $\equiv a \pmod N$.

In fact, we can prove that

$$\lim_{X\to\infty} \frac{|\{p \le X : p \equiv a \bmod N\}|}{|\{p \le X\}|} = \frac{1}{\varphi(N)}.$$

This theorem has many generalizations. In general, let L/K be a finite Galois extension of number fields with Galois group $G = \mathrm{Gal}(L/K)$. Then for all primes $\mathfrak{p}$ of K which is unramified in L, we can define a *Frobenius conjugacy class* $[\sigma_{\mathfrak{p}}] \subseteq G$.

Theorem (Cebotarev density theorem)**.** *Cebotarev density theorem* Let L/K be a Galois extension. Then for any conjugacy class $C \subseteq \mathrm{Gal}(L/K)$, there exists infinitely many $\mathfrak{p}$ with $[\sigma_{\mathfrak{p}}] = C$.

If $L/K = \mathbb{Q}(\sqrt[n]{1})/\mathbb{Q}$, then $G \cong (\mathbb{Z}/N\mathbb{Z})^\times$, and σ_p is just the element of G given by $p \pmod N$. So if we fix $a \pmod N \in G$, then there are infinitely many p with $p \equiv a \pmod N$. So we see the Cebotarev density theorem is indeed a generalization of Dirichlet's theorem.

4 The modular group

We now move on to study the other words that appear in the title of the course, namely modular forms. Modular forms are very special functions defined on the upper half plane

$$\mathcal{H} = \{z \in \mathbb{C} : \text{Im}(z) > 0\}.$$

The main property of a modular form is that they transform nicely under Möbius transforms. In this chapter, we will first try to understand these Möbius transforms. Recall that a matrix

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{GL}_2(\mathbb{C})$$

acts on $\mathbb{C} \cup \infty$ by

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

If we further restrict to matrices in $\text{GL}_2(\mathbb{R})$, then this maps $\mathbb{C} \setminus \mathbb{R}$ to $\mathbb{C} \setminus \mathbb{R}$, and $\mathbb{R} \cup \{\infty\}$ to $\mathbb{R} \cup \{\infty\}$.

We want to understand when this actually fixes the upper half plane. This is a straightforward computation

$$\text{Im}\, \gamma(z) = \frac{1}{2i}\left(\frac{az + b}{cz + d} - \frac{a\bar{z} + b}{c\bar{z} + d}\right) = \frac{1}{2i}\frac{(ad - bc)(z - \bar{z})}{|cz + d|^2} = \det(\gamma)\frac{\text{Im}\, z}{|cz + d|^2}.$$

Thus, we know $\text{Im}(\gamma(z))$ and $\text{Im}(z)$ have the same sign iff $\det(\gamma) > 0$. We write

Definition $(\text{GL}_2(\mathbb{R})^+)$**.**

$$\text{GL}_2(\mathbb{R})^+ = \{\gamma \in \text{GL}_2(\mathbb{R}) : \det \gamma > 0\}.$$

This is the group of Möbius transforms that map $\mathcal{H}$ to $\mathcal{H}$.

However, note that the action of $\text{GL}_2(\mathbb{R})^+$ on $\mathcal{H}$ is not faithful. The kernel is given by the subgroup

$$\mathbb{R}^\times \cdot I = \mathbb{R}^\times \cdot \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus, we are naturally led to define

Definition $(\text{PGL}_2(\mathbb{R})^+)$**.**

$$\text{PGL}_2(\mathbb{R})^+ = \frac{\text{GL}_2(\mathbb{R})^+}{\mathbb{R}^\times \cdot I}.$$

There is a slightly better way of expressing this. Now note that we can obtain any matrix in $\text{GL}_2(\mathbb{R}^+)$, by multiplying an element of $\text{SL}_2(\mathbb{R})$ with a unit in $\mathbb{R}$. So we have

$$\text{PGL}_2(\mathbb{R})^+ \cong \text{SL}_2(\mathbb{R})/\{\pm I\} \equiv \text{PSL}_2(\mathbb{R}).$$

What we have is thus a faithful action of $\text{PSL}_2(\mathbb{R})$ on the upper half plane $\mathcal{H}$. From IA Groups, we know this action is transitive, and the stabilizer of $i = \sqrt{-1}$ is $\text{SO}(2)/\{\pm I\}$.

In fact, this group $\text{PSL}_2(\mathbb{R})$ is the group of all holomorphic automorphisms of $\mathcal{H}$, and the subgroup $\text{SO}(2) \subseteq \text{SL}_2(\mathbb{R})$ is a maximal compact subgroup.

Theorem. The group $\mathrm{SL}_2(\mathbb{R})$ admits the *Iwasawa decomposition*

$$\mathrm{SL}_2(\mathbb{R}) = KAN = NAK,$$

where

$$K = \mathrm{SO}(2), \quad A = \left\{\begin{pmatrix} r & 0 \\ 0 & \frac{1}{r}\end{pmatrix}\right\}, \quad N = \left\{\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}\right\}$$

Note that this implies a few of our previous claims. For example, any $z = x + iy \in \mathbb{C}$ can be written as

$$z = x + iy = \begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}\begin{pmatrix} \sqrt{y} & 0 \\ 0 & \frac{1}{\sqrt{y}}\end{pmatrix} \cdot i,$$

using the fact that $K = \mathrm{SO}(2)$ fixes i, and this gives transitivity.

Proof. This is just Gram–Schmidt orthogonalization. Given $g \in \mathrm{GL}_2(\mathbb{R})$, we write

$$ge_1 = e_1', \quad ge_2 = e_2',$$

By Gram-Schmidt, we can write

$$\begin{aligned} f_1 &= \lambda_1 e_1' \\ f_2 &= \lambda_2 e_1' + \mu e_2' \end{aligned}$$

such that

$$\|f_1\| = \|f_2\| = 1, \quad (f_1, f_2) = 0.$$

So we can write

$$\begin{pmatrix} f_1 & f_2\end{pmatrix} = \begin{pmatrix} e_1' & e_2'\end{pmatrix}\begin{pmatrix} \lambda_1 & \lambda_2 \\ 0 & \mu\end{pmatrix}$$

Now the left-hand matrix is orthogonal, and by decomposing the inverse of $\begin{pmatrix} \lambda_1 & \lambda_2 \\ 0 & \mu\end{pmatrix}$, we can write $g = \begin{pmatrix} e_1' & e_2'\end{pmatrix}$ as a product in KAN. □

In general, we will be interested in subgroups $\Gamma \leq \mathrm{SL}_2(\mathbb{R})$, and their images $\bar{\Gamma}$ in $\Gamma \in \mathrm{PSL}_2(\mathbb{R})$, i.e.

$$\bar{\Gamma} = \frac{\Gamma}{\Gamma \cap \{\pm I\}}.$$

We are mainly interested in the case $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, or a subgroup of finite index.

Definition (Modular group)**.** The *modular group* is

$$\mathrm{PSL}_2(\mathbb{Z}) = \frac{\mathrm{SL}_2(\mathbb{Z})}{\{\pm I\}}.$$

There are two particularly interesting elements of the modular group, given by

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

Then we have $T(z) = z + 1$ and $S(z) = -\frac{1}{z}$. One immediately sees that T has infinite order and $S^2 = 1$ (in $\mathrm{PSL}_2(\mathbb{Z})$). We can also compute

$$TS = \pm\begin{pmatrix} 1 & -1 \\ 1 & 0\end{pmatrix}$$

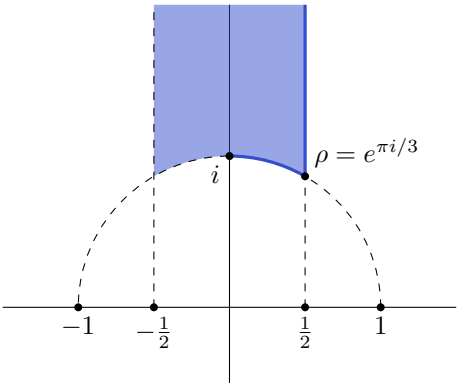
and

$$(TS)^3 = 1.$$

The following theorem essentially summarizes the basic properties of the modular group we need to know about:

Theorem. Let

$$\mathcal{D} = \left\{z \in \mathcal{H} : -\frac{1}{2} \leq \operatorname{Re} z \leq \frac{1}{2}, |z| > 1\right\} \cup \{z \in \mathcal{H} : |z| = 1, \operatorname{Re}(z) \geq 0\}.$$



Then $\mathcal{D}$ is a *fundamental domain* for the action of $\bar{\Gamma}$ on $\mathcal{H}$, i.e. every orbit contains exactly one element of $\mathcal{D}$.

The stabilizer of $z \in \mathcal{D}$ in Γ is trivial if $z \neq i, \rho$, and the stabilizers of i and ρ are

$$\bar{\Gamma}_i = \langle S \rangle \cong \frac{\mathbb{Z}}{2Z}, \quad \bar{\Gamma}_\rho = \langle TS \rangle \cong \frac{\mathbb{Z}}{3\mathbb{Z}}.$$

Finally, we have $\bar{\Gamma} = \langle S, T \rangle = \langle S, TS \rangle$.

In fact, we have

$$\bar{\Gamma} = \langle S, T \mid S^2 = (TS)^3 = e \rangle,$$

but we will neither prove nor need this.

The proof is rather technical, and involves some significant case work.

Proof. Let $\bar{\Gamma}^* = \langle S, T \rangle \subseteq \bar{\Gamma}$. We will show that if $z \in \mathcal{H}$, then there exists $\gamma \in \bar{\Gamma}^*$ such that $\gamma(z) \in \mathcal{D}$.

Since $z \notin \mathbb{R}$, we know $\mathbb{Z} + \mathbb{Z}z = \{cz + d : c, d \in \mathbb{Z}\}$ is a discrete subgroup of $\mathbb{C}$. So we know

$$\{|cz + d| : c, d \in \mathbb{Z}\}$$

is a discrete subset of $\mathbb{R}$, and is in particular bounded away from 0. Thus, we know

$$\left\{\operatorname{Im} \gamma(z) = \frac{\operatorname{Im}(z)}{|cz + d|^2} : \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \bar{\Gamma}^*\right\}$$

is a discrete subset of $\mathbb{R}_{>0}$ and is bounded above. Thus there is some $\gamma \in \bar{\Gamma}^*$ with $\operatorname{Im} \gamma(z)$ maximal. Replacing γ by $T^n\gamma$ for suitable n, we may assume $|\operatorname{Re} \gamma(z)| \leq \frac{1}{2}$.

We consider the different possible cases.

– If $|\gamma(z)| < 1$, then

$$\operatorname{Im} S\gamma(z) = \operatorname{Im} \frac{-1}{\gamma(z)} = \frac{\operatorname{Im}\gamma(z)}{|\gamma(z)|^2} > \operatorname{Im}\gamma(z),$$

which is impossible. So we know $|\gamma(z)| \geq 1$. So we know $\gamma(z)$ lives in the closure of $\mathcal{D}$.

– If $\operatorname{Re}(\gamma(z)) = -\frac{1}{2}$, then $T\gamma(z)$ has real part $+\frac{1}{2}$, and so $T(\gamma(z)) \in \mathcal{D}$.

– If $-\frac{1}{2} < \operatorname{Re}(z) < 0$ and $|\gamma(z)| = 1$, then $|S\gamma(z)| = 1$ and $0 < \operatorname{Re} S\gamma(z) < \frac{1}{2}$, i.e. $S\gamma(z) \in \mathcal{D}$.

So we can move it to somewhere in $\mathcal{D}$.

We shall next show that if $z, z' \in \mathcal{D}$, and $z' = \gamma(z)$ for $\gamma \in \bar{\Gamma}$, then $z = z'$. Moreover, either

– $\gamma = 1$; or

– $z = i$ and $\gamma = S$; or

– $z = \rho$ and $\gamma = TS$ or $(TS)^2$.

It is clear that this proves everything.

To show this, we wlog

$$\operatorname{Im}(z') = \frac{\operatorname{Im} z}{|cz + d|^2} \geq \operatorname{Im} z$$

where

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

and we also wlog $c \geq 0$.

Therefore we know that $|cz + d| \leq 1$. In particular, we know

$$1 \geq \operatorname{Im}(cz + d) = c \operatorname{Im}(z) \geq c\frac{\sqrt{3}}{2}$$

since $z \in \mathcal{D}$. So $c = 0$ or 1.

– If $c = 0$, then

$$\gamma = \pm \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}$$

for some $m \in \mathbb{Z}$, and this $z' = z + m$. But this is clearly impossible. So we must have $m = 0$, $z = z'$, $\gamma = 1 \in \mathrm{PSL}_2(\mathbb{Z})$.

– If $c = 1$, then we know $|z + d| \leq 1$. So z is at distance 1 from an integer. As $z \in \mathcal{D}$, the only possibilities are $d = 0$ or -1.

  ◦ If $d = 0$, then we know $|z| = 1$. So

$$\gamma = \begin{pmatrix} a & -1 \\ 1 & 0 \end{pmatrix}$$

for some $a \in \mathbb{Z}$. Then $z' = a - \frac{1}{z}$. Then

  - either $a = 0$, which forces $z = i$, $\gamma = S$; or
  - $a = 1$, and $z' = 1 - \frac{1}{z}$, which implies $z = z' = \rho$ and $\gamma = TS$.
- If $d = -1$, then by looking at the picture, we see that $z = \rho$. Then

$$|cz + d| = |z - 1| = 1,$$

and so

$$\operatorname{Im} z' = \operatorname{Im} z = \frac{\sqrt{3}}{2}.$$

So we have $z' = \rho$ as well. So

$$\frac{a\rho + b}{\rho - 1} = \rho,$$

which implies

$$\rho^2 - (a + 1)\rho - b = 0$$

So $\rho = -1, a = 0$, and $\gamma = (TS)^2$.

□

Note that this proof is the same as the proof of reduction theory for binary positive definite binary quadratic forms.

What does the quotient $\bar{\Gamma} \setminus \mathbb{N}$ look like? Each point in the quotient can be identified with an element in $\mathcal{D}$. Moreover, S and T identify the portions of the boundary of $\mathcal{D}$. Thinking hard enough, we see that the quotient space is homeomorphic to a disk.

An important consequence of this is that the quotient $\Gamma \setminus \mathcal{H}$ has *finite invariant measure.*

Proposition. The measure

$$\mathrm{d}\mu = \frac{\mathrm{d}x\, \mathrm{d}y}{y^2}$$

is invariant under $\mathrm{PSL}_2(\mathbb{R})$. If $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{Z})$ is of finite index, then $\mu(\Gamma \setminus \mathcal{H}) < \infty$.

Proof. Consider the 2-form associated to μ, given by

$$\eta = \frac{\mathrm{d}x \wedge \mathrm{d}y}{y^2} = \frac{i\mathrm{d}z \wedge \mathrm{d}\bar{z}}{2(\operatorname{Im} z)^2}.$$

We now let

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

Then we have

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

Moreover, we have

$$\frac{\mathrm{d}\gamma(z)}{\mathrm{d}z} = \frac{a(cz + d) - c(az + b)}{(cz + d)^2} = \frac{1}{(cz + d)^2}.$$

Plugging these into the formula, we see that η is invariant under γ.

Now if $\bar{\Gamma} \leq \mathrm{PSL}_2(\mathbb{Z})$ has finite index, then we can write $\mathrm{PSL}_2(\mathbb{Z})$ as a union of cosets

$$\mathrm{PSL}_2(\mathbb{Z}) = \coprod_{i=1}^{n} \bar{\gamma}\gamma_i,$$

where $n = (\mathrm{PSL}_2(\mathbb{Z}) : \bar{\Gamma})$. Then a fundamental domain for $\bar{\Gamma}$ is just

$$\bigcup_{i=1}^{n} \gamma_i(\mathcal{D}),$$

and so

$$\mu(\bar{\Gamma} \setminus H) = \sum \mu(\gamma_i \mathcal{D}) = n\mu(\mathcal{D}).$$

So it suffices to show that $\mu(\mathcal{D})$ is finite, and we simply compute

$$\mu(\mathcal{D}) = \int_{\mathcal{D}} \frac{\mathrm{d}x\, \mathrm{d}y}{y^2} \leq \int_{x=-\frac{1}{2}}^{x=\frac{1}{2}} \int_{y=\sqrt{2}/2}^{y=\infty} \frac{\mathrm{d}x\, \mathrm{d}y}{y^2} < \infty.$$

□

It is an easy exercise to show that we actually have

$$\mu(\mathcal{D}) = \frac{\pi}{3}.$$

We end with a bit terminology.

Definition (Principal congruence subgroup)**.** For $N \geq 1$, the *principal congruence subgroup* of level N is

$$\Gamma(N) = \{\gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv I \pmod{N}\} = \ker(\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})).$$

Any $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ containing some $\Gamma(N)$ is called a *congruence subgroup*, and its *level* is the smallest N such that $\Gamma \supseteq \Gamma(N)$

This is a normal subgroup of finite index.

Definition ($\Gamma_0(N)$, $\Gamma_1(N)$)**.** We define

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

and

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

We similarly define $\Gamma^0(N)$ and $\Gamma^1(N)$ to be the transpose of $\Gamma_0(N)$ and $\Gamma_1(N)$ respectively.

Note that "almost all" subgroups of $\mathrm{SL}_2(\mathbb{Z})$ are *not* congruence subgroups. On the other hand, if we try to define the notion of congruence subgroups in higher dimensions, we find that all subgroups of $\mathrm{SL}_n(\mathbb{Z})$ for $n > 2$ are congruence!

5 Modular forms of level 1

5.1 Basic definitions

We can now define a modular form. Recall that we have $\mathrm{SL}_2(\mathbb{Z}) = \Gamma(1)$.

Definition (Modular form of level 1)**.** A holomorphic function $f : \mathcal{H} \to \mathbb{C}$ is a *modular form of weight $k \in \mathbb{Z}$ and level* 1 if

(i) For any
$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1),$$
we have
$$f(\gamma(z)) = (cz + d)^k f(z).$$

(ii) f is holomorphic at ∞ (to be defined precisely later).

What can we deduce about modular forms from these properties? If we take $\gamma = -I$, then we get
$$f(z) = (-1)^k f(z).$$
So if k is odd, then $f \equiv 0$. So they only exist for even weights. If we have even weights, then it suffices to consider $\bar{\Gamma} = \langle S, T\rangle$. Since
$$f(z) \mapsto (cz + d)^{-k} f(\gamma(z))$$
is a *group action* of $\Gamma(1)$ on functions on $\mathcal{H}$, it suffices to check that f is invariant under the generators S and T. Thus, (i) is equivalent to
$$f(z + 1) = f(z), \quad f(-1/z) = z^k f(z).$$
How do we interpret (ii)? We know f is $\mathbb{Z}$-periodic. If we write $q = e^{2\pi i z}$, then we have $z \in \mathcal{H}$ iff $0 < |q| < 1$, and moreover, if two different z give the same q, then the values of f on the two z agree. In other words, $f(z)$ only depends on q, and thus there exists a holomorphic function $\tilde{f}(q)$ on $\{0 < |q| < 1\}$ such that
$$\tilde{f}(e^{2\pi i z}) = f(z).$$
Explicitly, we can write
$$\tilde{f}(q) = f\left(\frac{1}{2\pi i} \log q\right).$$
By definition, $\tilde{f}$ is a holomorphic function on a punctured disk. So we have a Laurent expansion
$$\tilde{f}(q) = \sum_{n=-\infty}^{\infty} a_n(f) q^n,$$
called the *Fourier expansion* or *q-expansion* of f. We say f is meromorphic (resp. holomorphic) at ∞ if $\tilde{f}$ is meromorphic (resp. holomorphic) at $q = 0$.

In other words, it is meromorphic at ∞ if $a_n(f) = 0$ for n sufficiently negative, and holomorphic if $a_n(f) = 0$ for all $n \geq 0$. The latter just says $f(z)$ is bounded as $\mathrm{Im}(z) \to \infty$.

The following definition is also convenient:

Definition (Cusp form)**.** A modular form f is a *cusp form* if the constant term $a_0(f)$ is 0.

We will later see that "almost all" modular forms are cusp forms.

In this case, we have

$$\tilde{f} = \sum_{n\geq 1} a_n(f) q^n.$$

From now on, we will drop the $\tilde{}$, which should not cause confusion.

Definition (Weak modular form)**.** A *weak modular form* is a holomorphic form on $\mathcal{H}$ satisfying (i) which is *meromorphic* at ∞.

We will use these occasionally.

The transformation rule for modular forms seem rather strong. So, are there actually modular forms? It turns out that there are quite a lot of modular forms, and remarkably, there is a relatively easy way of listing out all the modular forms.

The main class (and in fact, as we will later see, a generating class) of modular forms is due to Eisenstein. This has its origin in the theory of elliptic functions, but we will not go into that.

Definition (Eisenstein series)**.** Let $k \geq 4$ be even. We define

$$G_k(z) = \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} \frac{1}{(mz+n)^k} = \sum_{(m,n)\in\mathbb{Z}^2}{}' \frac{1}{(mz+n)^k}.$$

Here the $\sum'$ denotes that we are omitting 0, and in general, it means we don't sum over things we obviously don't want to sum over.

When we just write down this series, it is not clear that it is a modular form, or even that it converges. This is given by the following theorem:

Theorem. G_k is a modular form of weight k and level 1. Moreover, its q-expansion is

$$G_k(z) = 2\zeta(k)\left(1 - \frac{2k}{B_k}\sum_{n\geq 1}\sigma_{k-1}(n)q^n\right), \tag{1}$$

where

$$\sigma_r(n) = \sum_{1\leq d|n} d^r.$$

Convergence of the series follows from the following more general result. Note that since $z \notin \mathbb{R}$, we know $\{1, z\}$ is an $\mathbb{R}$-basis for $\mathbb{C}$.

Proposition. Let $(e_1, \cdots, e_d)$ be some basis for $\mathbb{R}^d$. Then if $r \in \mathbb{R}$, the series

$$\sum_{\mathbf{m}\in\mathbb{Z}^d}{}' \|m_1 e_1 + \cdots + m_d e_d\|^{-r}$$

converges iff $r > d$.

Proof. The function

$$(x_i) \in \mathbb{R}^d \mapsto \left\| \sum_{i=1} x_i e_i \right\|$$

is a norm on $\mathbb{R}^d$. As any 2 norms on $\mathbb{R}^d$ are equivalent, we know this is equivalent to the sup norm $\|\cdot\|_\infty$. So the series converges iff the corresponding series

$$\sum_{\mathbf{m}\in\mathbb{Z}^d}{}' \|\mathbf{m}\|_\infty^{-r}$$

converges. But if $1 \leq N \leq Z$, then the number of $\mathbf{m} \in \mathbb{Z}^d$ such that $\|\mathbf{m}\|_\infty = N$ is $(2N+1)^d - (2N-1)^d \sim 2^d d N^{d-1}$. So the series converges iff

$$\sum_{N\geq 1} N^{-r} N^{d-1}$$

converges, which is true iff $r > d$. □

Proof of theorem. Then convergence of the Eisenstein series by applying this to $\mathbb{R}^2 \cong \mathbb{C}$. So the series is absolutely convergent. Therefore we can simply compute

$$G_k(z+1) = \sum_{m,n}{}' \frac{1}{(mz + (m+n))^k} = G_k(z).$$

Also we can compute

$$G_k\left(-\frac{1}{z}\right) = \sum_{m,n}{}' = \frac{z^k}{(-m+nz)^k} = z^k G_k(z).$$

So G_k satisfies the invariance property. To show that G_k is holomorphic, and holomorphic at infinity, we'll derive the q-expansion (1). □

Lemma.

$$\sum_{n=\infty}^{\infty} \frac{1}{(n+w)^k} = \frac{(-2\pi i)^k}{(k-1)!} \sum_{d=1}^{\infty} d^{k-1} e^{2\pi i d w}$$

for any $w \in \mathcal{H}$ and $k \geq 2$.

There are (at least) two ways to prove this. One of this is to use the series for the cotangent, but here we will use Poisson summation.

Proof. Let

$$f(x) = \frac{1}{(x+w)^k}.$$

We compute

$$\hat{f}(y) = \int_{-\infty}^{\infty} \frac{e^{-2\pi i x y}}{(x+w)^k}\, \mathrm{d}x.$$

We replace this with a contour integral. We see that this has a pole at $-w$. If $y > 0$, then we close the contour downwards, and we have

$$\hat{f}(y) = -2\pi i \operatorname{Res}_{z=-w} \frac{e^{-2\pi i y z}}{(z+w)^k} = -2\pi i \frac{(-2\pi i y)^{k-1}}{(k-1)!} e^{2\pi i y w}.$$

If $y \leq 0$, then we close in the upper half plane, and since there is no pole, we have $\hat{f}(y) = 0$. So we have

$$\sum_{n=-\infty}^{\infty} \frac{1}{(n+w)^k} = \sum_{n\in\mathbb{Z}} f(n) = \sum_{d\in\mathbb{Z}} \hat{f}(d) = \frac{(-2\pi i)^k}{(k-1)!} \sum_{d\geq 1} d^{k-1} e^{2\pi i d w}$$

by Poisson summation formula. □

Note that when we proved the Poisson summation formula, we required f to decrease very rapidly at infinity, and our f does not satisfy that condition. However, we can go back and check that the proof still works in this case.

Now we get back to the Eisenstein series. Note that since k is even, we can drop certain annoying signs. We have

$$\begin{aligned} G_k(z) &= 2\sum_{n\geq 1} \frac{1}{n^k} + 2\sum_{m\geq 1}\sum_{n\in\mathbb{Z}} \frac{1}{(n+mz)^k} \\ &= 2\zeta(k) + 2\sum_{m\geq 1} \frac{(2\pi i)^k}{(k-1)!} \sum_{d\geq 1} d^{k-1} q^{dm}. \\ &= 2\zeta(k) + 2\frac{(2\pi i)^k}{(k-1)!} \sum_{n\geq 1} \sigma_{k-1}(n) q^n. \end{aligned}$$

Then the result follows from the fact that

$$\zeta(k) = -\frac{1}{2}(2\pi i)^k \frac{B_k}{k!}.$$

So we see that G_k is holomorphic in $\mathcal{H}$, and is also holomorphic at ∞.

It is convenient to introduce a *normalized* Eisenstein series

Definition (Normalized Eisenstein series)**.** We define

$$\begin{aligned} E_k(z) &= (2\zeta(k))^{-1} G_k(z) \\ &= 1 - \frac{2k}{B_k} \sum_{n\geq 1} \sigma_{k-1}(n) q^n \\ &= \frac{1}{2} \sum_{\substack{(m,n)=1 \\ m,n\in\mathbb{Z}}} \frac{1}{(mz+n)^k}. \end{aligned}$$

The last line follows by taking out any common factor of m, n in the series defining G_k.

Thus, to figure out the (normalized) Eisenstein series, we only need to know the Bernoulli numbers.

Example. We have

$$\begin{aligned} B_2 &= \frac{1}{6}, & B_4 &= \frac{-1}{30}, & B_6 &= \frac{1}{42}, & B_8 &= \frac{-1}{30} \\ B_{10} &= \frac{5}{66}, & B_{12} &= \frac{-631}{2730}, & B_{14} &= \frac{7}{6}. \end{aligned}$$

Using these, we find

$$\begin{aligned} E_4 &= 1 + 240 \sum \sigma_3(n) q^n \\ E_6 &= 1 - 504 \sum \sigma_5(n) q^n \\ E_8 &= 1 + 480 \sum \sigma_7(n) q^n \\ E_{10} &= 1 - 264 \sum \sigma_9(n) q^n \\ E_{12} &= 1 + \frac{65520}{691} \sum \sigma_{11}(n) q^n \\ E_{14} &= 1 - 24 \sum \sigma_{13}(n) q^n. \end{aligned}$$

We notice that there is a simple pattern for $k \leq 14$, except for $k = 12$.

For more general analysis of modular forms, it is convenient to consider the following notation:

Definition (Slash operator)**.** Let

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \gamma \in \mathrm{GL}_2(\mathbb{R})^+, \quad z \in \mathcal{H},$$

and $f : \mathcal{H} \to \mathbb{C}$ any function. We write

$$j(\gamma, z) = cz + d.$$

We define the *slash operator* to be

$$(f|_k \gamma)(z) = (\det \gamma)^{k/2} j(\gamma, z)^{-k} f(\gamma(z)).$$

Note that some people leave out the $\det \gamma^{k/2}$ factor, but if we have it, then whenever $\gamma = Ia$, then

$$f|_k \gamma = \mathrm{sgn}(a)^k f,$$

which is annoying. In this notation, then condition (i) for f to be a modular form is just

$$f|_k \gamma = f$$

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

To prove things about our j operator, it is convenient to note that

$$\gamma \begin{pmatrix} z \\ 1 \end{pmatrix} = j(\gamma, z) \begin{pmatrix} \gamma(z) \\ 1 \end{pmatrix}. \qquad (*)$$

Proposition.

(i) $j(\gamma\delta, z) = j(\gamma, \delta(z)) j(\delta, z)$ (in fancy language, we say j is a 1-cocycle).

(ii) $j(\gamma^{-1}, z) = j(\gamma, \gamma^{-1}(z))^{-1}$.

(iii) $\gamma : \varphi \mapsto f|_k \gamma$ is a (right) action of $G = \mathrm{GL}_2(\mathbb{R})^+$ on functions on $\mathcal{H}$. In other words,

$$f|_k \gamma|_k \delta = f|_k (\gamma\delta).$$

Note that this implies that if if $\Gamma \leq \mathrm{GL}_2(\mathbb{R})^+$ and $\Gamma = \langle \gamma_1, \cdots, \gamma_m \rangle$ then

$$f \underset{k}{|} \gamma = f \iff f \underset{k}{|} \gamma_i = f \text{ for all } i = 1, \cdots, m.$$

The proof is just a computation.

Proof.

(i) We have

$$j(\gamma\delta, z) \begin{pmatrix} \gamma\delta(z) \\ 1 \end{pmatrix} = \gamma\delta \begin{pmatrix} z \\ 1 \end{pmatrix} = j(\delta, z)\gamma \begin{pmatrix} \delta(z) \\ 1 \end{pmatrix} = j(\delta, z) j(\gamma, \delta(z)) \begin{pmatrix} z \\ 1 \end{pmatrix}$$

(ii) Take $\delta = \gamma^{-1}$.

(iii) We have

$$\begin{aligned} ((f \underset{k}{|} \gamma) \underset{k}{|} \delta)(z) &= (\det \delta)^{k/2} j(\delta, z)^{-k} (f \underset{k}{|} \gamma)(\delta(z)) \\ &= (\det \delta)^{k/2} j(\delta, z)^{-k} (\det \gamma)^{k/2} j(\gamma, \delta(z))^{-k} f(\gamma\delta(z)) \\ &= (\det \gamma\delta)^{k/2} j(\gamma\delta, z)^{-k} f(\gamma\delta(z)) \\ &= (f \underset{k}{|} \gamma\delta)(z). \end{aligned}$$

□

Back to the Eisenstein series. G_k arise naturally in elliptic functions, which are coefficients in the series expansion of Weierstrass $\wp$ function.

There is another group-theoretic interpretation, which generalizes in many ways. Consider

$$\Gamma(1)_\infty = \left\{ \pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} : n \in \mathbb{Z} \right\} \leq \Gamma(1) = \mathrm{SL}_2(\mathbb{Z}),$$

which is the stabilizer of ∞. If

$$\delta = \pm \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \in \Gamma(1)_\infty,$$

then we have

$$j(\delta\gamma, z) = j(\delta, \gamma(z)) j(\gamma, z) = \pm j(\gamma, z).$$

So $j(\gamma, z)^2$ depends only on the coset $\Gamma(1)_\infty \gamma$. We can also check that if

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \gamma = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \Gamma(1),$$

then $\Gamma(1)_\infty \gamma = \Gamma(1)_\infty \gamma'$ iff $(c, d) = \pm(c', d')$.

Moreover, $\gcd(c, d) = 1$ iff there exists a, b such that

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = 1.$$

We therefore have

$$E_k(z) = \sum_{\gamma \in \Gamma(1)_\infty \backslash \Gamma(1)} j(\gamma, z)^{-k},$$

where we sum over (any) coset representatives of $\Gamma(1)_\infty$.

We can generalize this in two ways. We can either replace j with some other appropriate function, or change the groups.

5.2 The space of modular forms

In this section, we are going to find out *all* modular forms! For $k \in \mathbb{Z}$, we write $M_k = M_k(\Gamma(1))$ for the set of modular forms of weight k (and level 1). We have $S_k \subseteq S_k(\Gamma(1))$ containing the cusp forms. These are $\mathbb{C}$-vector spaces, and are zero for odd k.

Moreover, from the definition, we have a natural product

$$M_k \cdot M_\ell \subseteq M_{k+\ell}.$$

Likewise, we have

$$S_k \cdot M_\ell \subseteq S_{k+\ell}.$$

We let

$$M_* = \bigoplus_{k\in\mathbb{Z}} M_k, \quad S_* = \bigoplus_{k\in\mathbb{Z}} S_k.$$

Then M_* is a graded ring and S_* is a graded ideal. By definition, we have

$$S_k = \ker(a_0 : M_k \to \mathbb{C}).$$

To figure out what all the modular forms are, we use the following constraints on the zeroes of a modular form:

Proposition. Let f be a weak modular form (i.e. it can be meromorphic at ∞) of weight k and level 1. If f is not identically zero, then

$$\left(\sum_{z_0\in\mathcal{D}\setminus\{i,\rho\}} \operatorname{ord}_{z_0}(f)\right) + \frac{1}{2}\operatorname{ord}_i(f) + \frac{1}{3}\operatorname{ord}_\rho f + \operatorname{ord}_\infty(f) = \frac{k}{12},$$

where $\operatorname{ord}_\infty f$ is the least $r \in \mathbb{Z}$ such that $a_r(f) \neq 0$.

Note that if $\gamma \in \Gamma(1)$, then $j(\gamma, z) = cz + d$ is never 0 for $z \in \mathcal{H}$. So it follows that $\operatorname{ord}_z f = \operatorname{ord}_{\gamma(z)} f$.

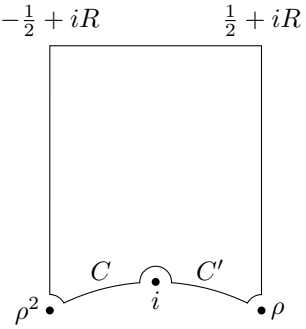
We will prove this using the argument principle.

Proof. Note that the function $\tilde{f}(q)$ is non-zero for $0 < |q| < \varepsilon$ for some small ε by the principle of isolated zeroes. Setting

$$\varepsilon = e^{-2\pi R},$$

we know $f(z) \neq 0$ if $\operatorname{Im} z \geq R$.

In particular, the number of zeroes of f in $\mathcal{D}$ is finite. We consider the integral along the following contour, counterclockwise.



We assume f has no zeroes along the contour. Otherwise, we need to go around the poles, which is a rather standard complex analytic maneuver we will not go through.

For ε sufficiently small, we have

$$\int_\Gamma \frac{f'(z)}{f(z)}\,\mathrm{d}z = 2\pi i \sum_{z_0 \in \mathcal{D}\setminus\{i,\rho\}} \operatorname{ord}_{z_0} f$$

by the argument principle. Now the top integral is

$$\int_{\frac{1}{2}+iR}^{-\frac{1}{2}iR} \frac{f'}{f}\,\mathrm{d}z = -\int_{|q|=\varepsilon} \frac{\frac{\mathrm{d}\tilde{f}}{\mathrm{d}q}}{\tilde{f}(q)}\,\mathrm{d}q = -2\pi i \operatorname{ord}_\infty f.$$

As $\frac{f'}{f}$ has at worst a simple pole at $z = i$, the residue is $\operatorname{ord}_i f$. Since we are integrating along only half the circle, as $\varepsilon \to 0$, we pick up

$$-\pi i \operatorname{res} = -\pi i \operatorname{ord}_i f.$$

Similarly, we get $-\frac{2}{3}\pi i \operatorname{ord}_\rho f$ coming from ρ and ρ^2.

So it remains to integrate along the bottom circular arcs. Now note that $S : z \mapsto -\frac{1}{z}$ maps C to C' with opposite orientation, and

$$\frac{\mathrm{d}f(Sz)}{f(Sz)} = k\frac{\mathrm{d}z}{z} + \frac{\mathrm{d}f(z)}{f(z)}$$

as

$$f(Sz) = z^k f(z).$$

So we have

$$\begin{aligned}\int_C + \int_{C'} \frac{f'}{f}\,\mathrm{d}z &= \int_{C'} \frac{f'}{f}\,\mathrm{d}z - \left(\frac{k}{z}\,\mathrm{d}z + \frac{f'}{f}\,\mathrm{d}z\right) - -k\int_{C'} \frac{\mathrm{d}z}{z}\\ &\to k\int_\rho^i \frac{\mathrm{d}z}{z}\\ &= \frac{\pi i k}{6}.\end{aligned}$$

So taking the limit $\varepsilon \to 0$ gives the right result. □

Corollary. If $k < 0$, then $M_k = \{0\}$.

Corollary. If $k = 0$, then $M_0 = \mathbb{C}$, the constants, and $S_0 = \{0\}$.

Proof. If $f \in M_0$, then $g = f - f(1)$. If f is not constant, then $\operatorname{ord}_i g \geq 1$, so the LHS is > 0, but the RHS is $= 0$. So $f \in \mathbb{C}$.

Of course, $a_0(f) = f$. So $S_0 = \{0\}$. □

Corollary.

$$\dim M_k \leq 1 + \frac{k}{12}.$$

In particular, they are finite dimensional.

Proof. We let $f_0, \cdots, f_d$ be $d+1$ elements of M_k, and we choose distinct points $z_1, \cdots, z_d \in \mathcal{D} \setminus \{i, \rho\}$. Then there exists $\lambda_0, \cdots, \lambda_d \in \mathbb{C}$, not all 0, such that

$$f = \sum_{i=0}^{d} \lambda_i f_i$$

vanishes at all these points. Now if $d > \frac{k}{12}$, then LHS is $> \frac{k}{12}$. So $f \equiv 0$. So (f_i) are linearly dependent, i.e. $\dim M_k < d+1$. □

Corollary. $M_2 = \{0\}$ and $M_k = \mathbb{C}E_k$ for $4 \leq k \leq 10$ (k even). We also have $E_8 = E_4^2$ and $E_{10} = E_4 E_6$.

Proof. Only $M_2 = \{0\}$ requires proof. If $0 \neq f \in M_2$, then this implies

$$a + \frac{b}{2} + \frac{c}{3} = \frac{1}{6}$$

for integers $a, b, c \geq 0$, which is not possible.

Alternatively, if $f \in M_2$, then $f^2 \in M_4$ and $f^3 \in M_6$. This implies $E_4^3 = E_6^2$, which is not the case as we will soon see.

Note that we know $E_8 = E_4^2$, and is not just a multiple of it, by checking the leading coefficient (namely 1). □

Corollary. The cusp form of weight 12 is

$$E_4^3 - E_6^2 = (1 + 240q + \cdots)^3 - (1 - 504q + \cdots)^2 = 1728q + \cdots.$$

Note that $1728 = 12^3$.

Definition (Δ and τ)**.**

$$\Delta = \frac{E_4^3 - E_6^2}{1728} = \sum_{n \geq 1} \tau(n) q^n \in S_{12}.$$

This function τ is very interesting, and is called *Ramanujan's τ-function*. It has nice arithmetic properties we'll talk about soon.

The following is a crucial property of Δ:

Proposition. $\Delta(z) \neq 0$ for all $z \in \mathcal{H}$.

Proof. We have

$$\sum_{z_0 \neq i, \rho} \operatorname{ord}_{z_0} \Delta + \frac{1}{2} \operatorname{ord}_i \Delta + \frac{1}{3} \operatorname{ord}_\rho \Delta + \operatorname{ord}_\infty \Delta = \frac{k}{12} = 1.$$

Since $\operatorname{ord}_\rho \Delta = 1$, it follows that there can't be any other zeroes. □

It follows from this that

Proposition. The map $f \mapsto \Delta f$ is an isomorphism $M_{k-12}(\Gamma(1)) \to S_k(\Gamma(1))$ for all $k > 12$.

Proof. Since $\Delta \in S_{12}$, it follows that if $f \in M_{k-1}$, then $\Delta f \in S_k$. So the map is well-defined, and we certainly get an injection $M_{k-12} \to S_k$. Now if $g \in S_k$, since $\operatorname{ord}_\infty \Delta = 1 \leq \operatorname{ord}_\infty g$ and $\Delta \neq \mathcal{H}$. So $\frac{g}{\Delta}$ is a modular form of weight $k - 12$. □

Thus, we find that

Theorem.

(i) We have

$$\dim M_k(\Gamma(1)) = \begin{cases} 0 & k < 0 \text{ or } k \text{ odd} \\ \lfloor \frac{k}{12} \rfloor & k > 0, k \equiv 2 \pmod{12} \\ 1 + \lfloor \frac{k}{12} \rfloor & \text{otherwise} \end{cases}$$

(ii) If $k > 4$ and even, then

$$M_k = S_k \oplus \mathbb{C} E_k.$$

(iii) Every element of M_k is a polynomial in E_4 and E_6.

(iv) Let

$$b = \begin{cases} 0 & k \equiv 0 \pmod 4 \\ 1 & k \equiv 2 \pmod 4 \end{cases}.$$

Then

$$\{h_j = \Delta^j E_6^b E_4^{(k-12j-6b)/4} : 0 \leq j < \dim M_k\}.$$

is a basis for M_k, and

$$\{h_j : 1 \leq j < \dim M_k\}$$

is a basis for S_k.

Proof.

(ii) S_k is the kernel of the homomorphism $M_k \to \mathbb{C}$ sending $f \mapsto a_0(f)$. So the complement of S_k has dimension at most 1, and we know E_k is an element of it. So we are done.

(i) For $k < 12$, this agrees with what we have already proved. By the proposition, we have

$$\dim M_{k-12} = \dim S_k.$$

So we are done by induction and (ii).

(iii) This is true for $k < 12$. If $k \geq 12$ is even, then we can find $a, b \geq 0$ with $4a + 6b = k$. Then $E_4^a E_6^b \in M_k$, and is not a cusp form. So

$$M_k = \mathbb{C} E_4^a E_6^b \oplus \Delta M_{k-12}.$$

But Δ is a polynomial in E_4, E_6, So we are done by induction on k.

(iv) By (i), we know $k - 12j - 6k \geq 0$ for $j < \dim M_k$, and is a multiple of 4. So $h_j \in M_k$. Next note that the q-expansion of h_j begins with q^j. So they are all linearly independent. □

So we have completely determined all modular forms, and this is the end of the course.

5.3 Arithmetic of Δ

Recall that we had

$$\Delta = \sum \tau(n) q^n,$$

and we knew

$$\tau(1) = 1, \quad \tau(n) \in \mathbb{Q}.$$

In fact, more is true.

Proposition.

(i) $\tau(n) \in \mathbb{Z}$ for all $n \geq 1$.

(ii) $\tau(n) = \sigma_{11}(n) \pmod{691}$

The function τ satisfies many more equations, some of which are on the second example sheet.

Proof.

(i) We have

$$1728\Delta = (1 + 240A_3(q))^3 - (1 - 504A_5(q))^2,$$

where

$$A_r = \sum_{n \geq 1} \sigma_r(n) q^n.$$

We can write this as

$$1728\Delta = 3 \cdot 240A_3 + 3 \cdot 240^2 A_3^2 + 240^3 A_3^3 + 2 \cdot 504A_5 - 504^2 A_5^2.$$

Now recall the deep fact that $1728 = 12^3$ and $504 = 21 \cdot 24$.

Modulo 1728, this is equal to

$$720A_3 + 1008A_5.$$

So it suffices to show that

$$5\sigma_3 + 7\sigma_5(n) \equiv 0 \pmod{12}.$$

In other words, we need

$$5d^3 + 7d^5 \equiv 0 \pmod{12},$$

and we can just check this manually for all d.

(ii) Consider

$$E_4^3 = 1 + \sum_{n \geq 1} b_n q^n$$

with $b_n \in \mathbb{Z}$. We also have

$$E_{12} = 1 + \frac{65520}{691} \sum_{n \geq 1} \sigma_{11}(n) q^n.$$

Also, we know

$$E_{12} - E_4^3 \in S_{12}.$$

So it is equal to $\lambda\Delta$ for some $\lambda \in \mathbb{Q}$. So we find that for all $n \geq 1$, we have

$$\frac{665520}{691}\sigma_{11}(n) - b_n = \lambda\tau(n).$$

In other words,

$$65520\sigma_{11}(n) - 691b_n = \mu\tau(n)$$

for some $\tau \in \mathbb{Q}$.

Putting $n = 1$, we know $\tau(1) = 1$, $\sigma_{11}(1) = 1$, and $b_1 \in \mathbb{Z}$. So $\mu \in \mathbb{Z}$ and $\mu \equiv 65520 \pmod{691}$. So for all $n \geq 1$, we have

$$65520\sigma_{11}(n) \equiv 65520\tau(n) \pmod{691}.$$

Since 691 and 65520 are coprime, we are done.

□

This proof is elementary, once we had the structure theorem, but doesn't really explain *why* the congruence is true.

The function $\tau(n)$ was studied extensively by Ramanujan. He proved the 691 congruence (and many others), and (experimentally) observed that if $(m, n) = 1$, then

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

Also, he observed that for any prime p, we have

$$|\tau(p)| < 2p^{11/2},$$

which was a rather curious thing to notice. Both of these things are true, and we will soon prove that the first is true. The second is also true, but it uses deep algebraic geometry. It was proved by Deligne in 1972, and he got a fields medal for proving this. So it's pretty hard.

We will also prove a theorem of Jacobi:

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

The numbers $\tau(p)$ are related to *Galois representations*.

Rationality and integrality

So far, we have many series that have rational coefficients in them. Given any subring $R \subseteq \mathbb{C}$, we let $M_k(R) = M_k(\Gamma(1), R)$ be the set of all $f \in M_k$ such that all $a_n(f) \in R$. Likewise, we define $S_k(R)$. For future convenience, we will prove a short lemma about them.

Lemma.

(i) Suppose $\dim M_k = d + 1 \geq 1$. Then there exists a basis $\{g_j : 0 \leq j \leq d\}$ for M_k such that

- $g_j \in M_k(\mathbb{Z})$ for all $j \in \{0, \cdots, d\}$.
- $a_n(g_j) = \delta_{nj}$ for all $j, n \in \{0, \cdots, d\}$.

(ii) For any R, $M_k(R) \cong R^{d+1}$ generated by $\{g_j\}$.

Proof.

(i) We take our previous basis $h_j = \Delta^j E_6^b E_4^{(k-12j-6b)/4} \in M_k(\mathbb{Z})$. Then we have $a_n(h_n) = 1$, and $a_j(h_n) = 0$ for all $j < n$. Then we just row reduce.

(ii) The isomorphism is given by

$$\begin{aligned} M_k(R) &\longleftrightarrow R^{d+1} \\ f &\longmapsto (a_n(f)) \\ \sum_{j=0}^{d} c_j g_j &\longleftarrow\!\shortmid (c_n) \end{aligned}$$

□

6 Hecke operators

6.1 Hecke operators and algebras

Recall that for $f : \mathcal{H} \to \mathbb{C}$, $\gamma \in \mathrm{GL}_2(\mathbb{R})^+$ and $k \in \mathbb{Z}$, we defined

$$(f\underset{k}{|}\gamma)(z) = (\det \gamma)^{k/2} j(\gamma, z)^k f(\gamma(z)),$$

where

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad j(\gamma, z) = cz + d.$$

We then defined

$$M_k = \{f : f\underset{k}{|}\gamma = f \text{ for all } \gamma \in \Gamma(1) + \text{holomorphicity condition}\}.$$

We showed that these are finite-dimensional, and we found a basis. But there is more to be said about modular forms. Just because we know polynomials have a basis $1, x, x^2, \cdots$ does not mean there isn't anything else to say about polynomials!

In this chapter, we will figure that M_k has the structure of a module for the *Hecke algebra*. This structure underlies the connection with arithmetic, i.e. Galois representations etc.

How might we try to get some extra structure on M_k? We might try to see what happens if we let something else in $\mathrm{GL}_2(\mathbb{R})^+$ act on f. Unfortunately, in general, if f is a modular form and $\gamma \in \mathrm{GL}_2(\mathbb{R})^+$, then $g = f\underset{k}{|}\gamma$ is not a modular form. Indeed, given a $\delta \in \Gamma(1)$, then it acts on g by

$$g\underset{k}{|}\delta = f\underset{k}{|}\gamma\delta = (f\underset{k}{|}\gamma\delta\gamma^{-1})\gamma$$

and usually $\gamma\delta\gamma^{-1} \notin \Gamma(1)$. In fact the normalizer of $\Gamma(1)$ in $\mathrm{GL}_2(\mathbb{R})^+$ is generated by $\Gamma(1)$ and aI for $a \in \mathbb{R}^*$.

It turns out we need to act in a smarter way. To do so, we have to develop quite a lot of rather elementary group theory.

Consider a group G, and $\Gamma \leq G$. The idea is to use the *double cosets* of Γ defined by

$$\Gamma g \Gamma = \{\gamma g \gamma' : \gamma, \gamma' \in \Gamma\}.$$

One alternative way to view this is to consider the right multiplication action of G, hence Γ on the right cosets Γg. Then the double coset $\Gamma g \Gamma$ is the union of the orbits of Γg under the action of Γ. We can write this as

$$\Gamma g \Gamma = \coprod_{i \in I} \Gamma g_i$$

for some $g_i \in g\Gamma \subseteq G$ and index set I.

In our applications, we will want this disjoint union to be finite. By the orbit-stabilizer theorem, the size of this orbit is the index of the stabilizer of Γg in Γ. It is not hard to see that the stabilizer is given by $\Gamma \cap g^{-1}\Gamma g$. Thus, we are led to consider the following hypothesis:

Hypothesis (H): For all $g \in G$, $(\Gamma : \Gamma \cap g^{-1}\Gamma g) < \infty$.

Then (G, Γ) satisfies (H) iff for any g, the double coset $\Gamma g \Gamma$ is the union of finitely many cosets.

The important example is the following:

Theorem. Let $G = \mathrm{GL}_2(\mathbb{Q})$, and $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ a subgroup of finite index. Then (G, Γ) satisfies (H).

Proof. We first consider the case $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. We first suppose

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

and $\det g = \pm N$, $N \geq 1$. We claim that

$$g^{-1}\Gamma g \cap \Gamma \supseteq \Gamma(N),$$

from which it follows that

$$(\Gamma : \Gamma \cap g^{-1}\Gamma g) < \infty.$$

So given $\gamma \in \Gamma(N)$, we need to show that $g\gamma g^{-1} \in \Gamma$, i.e. it has integer coefficients. We consider

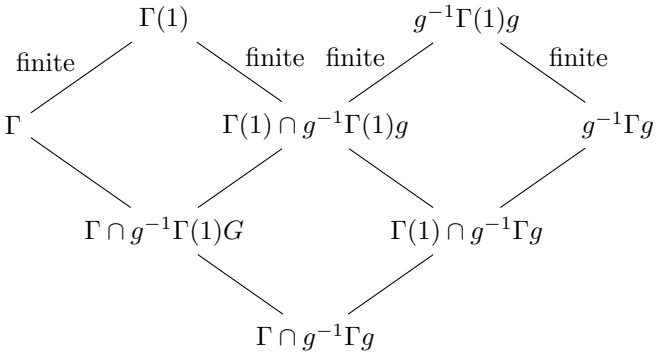
$$\pm N \cdot g\gamma g^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \gamma \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \equiv \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \equiv NI \equiv 0 \pmod{N}.$$

So we know that $g\gamma g^{-1}$ must have integer entries. Now in general, if $g' \in \mathrm{GL}_2(\mathbb{Q})$, then we can write

$$g' = \frac{1}{M} g$$

for g with integer entries, and we know conjugating by g and g' give the same result. So (G, Γ) satisfies (H).

The general result follows by a butterfly. Recall that if $(G : H) < \infty$ and $(G : H') < \infty$, then $(G : H \cap H') < \infty$. Now if $\Gamma \subseteq \Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$ is of finite index, then we can draw the diagram



Each group is the intersection of the two above, and so all inclusions are of finite index. □

Note that the same proof works for $\mathrm{GL}_N(\mathbb{Q})$ for any N.

Before we delve into concreteness, we talk a bit more about double cosets. Recall that cosets partition the group into pieces of equal size. Is this true for double cosets as well? We can characterize double cosets as orbits of $\Gamma \times \Gamma$ acting on G by

$$(\gamma, \delta) \cdot g = \gamma g \delta^{-1}.$$

So G is indeed the disjoint union of the double cosets of Γ.

However, it is not necessarily the case that all double cosets have the same size. For example $|\Gamma e \Gamma| = |\Gamma|$, but for a general g, $|\Gamma g \Gamma|$ can be the union of many cosets of Γ.

Our aim is to define a ring $\mathcal{H}(G, \Gamma)$ generated by double cosets called the *Hecke algebra*. As an abelian group, it is the free abelian group on symbols $[\Gamma g \Gamma]$ for each double coset $[\Gamma g \Gamma]$. It turns out instead of trying to define a multiplication for the Hecke algebra directly, we instead try to define an action of this on interesting objects, and then there is a unique way of giving $\mathcal{H}(G, \Gamma)$ a multiplicative structure such that this is a genuine action.

Given a group G, a *G-module* is an abelian group with a $\mathbb{Z}$-linear G-action. In other words, it is a module of the group ring $\mathbb{Z}G$. We will work with right modules, instead of the usual left modules.

Given such a module and a subgroup $\Gamma \leq G$, we will write

$$M^\Gamma = \{m \in M : m\gamma = m \text{ for all } \gamma \in \Gamma\}.$$

Notation. For $g \in G$ and $m \in M^\Gamma$, we let

$$m|[\Gamma g \Gamma] = \sum_{i=1}^{n} m g_i, \tag{$*$}$$

where

$$\Gamma g \Gamma = \coprod_{i=1}^{n} \Gamma g_i.$$

The following properties are immediate, but also crucial.

Proposition.

(i) $m|[\Gamma g \Gamma]$ depends only on $\Gamma g \Gamma$.

(ii) $m|[\Gamma g \Gamma] \in M^\Gamma$.

Proof.

(i) If $g_i' = \gamma_i g_i$ for $\gamma_i \in \Gamma$, then

$$\sum m g_i' = \sum m \gamma_i g_i = \sum m g_i$$

as $m \in M^\Gamma$.

(ii) Just write it out, using the fact that $\{\Gamma g_i\}$ is invariant under Γ.

□

Theorem. There is a product on $\mathcal{H}(G,\Gamma)$ making it into an associative ring, the *Hecke algebra* of (G,Γ), with unit $[\Gamma e\Gamma] = [\Gamma]$, such that for every G-module M, we have M^Γ is a right $\mathcal{H}(G,\Gamma)$-module by the operation $(*)$.

In the proof, and later on, we will use the following observation: Let $\mathbb{Z}[\Gamma\backslash G]$ be the free abelian group on cosets $[\Gamma g]$. This has an obvious right G-action by multiplication. We know a double coset is just an orbit of Γ acting on a single coset. So there is an isomorphism between

$$\Theta : \mathcal{H}(G,\Gamma) \to \mathbb{Z}[\Gamma\backslash G]^\Gamma.$$

given by

$$[\Gamma g\Gamma] \mapsto \sum [\Gamma g_i],$$

where

$$\Gamma g\Gamma = \coprod \Gamma g_i.$$

Proof. Take $M = \mathbb{Z}[\Gamma\backslash G]$, and let

$$\Gamma g\Gamma = \coprod \Gamma g_i$$
$$\Gamma h\Gamma = \coprod \Gamma h_j.$$

Then

$$\sum_i [\Gamma g_i] \in M^\Gamma,$$

and we have

$$\sum_i [\Gamma g_i]|[\Gamma h\Gamma] = \sum_{i,j} [\Gamma g_i h_j] \in M^\Gamma,$$

and this is well-defined. This gives us a well-defined product on $\mathcal{H}(G,\Gamma)$. Explicitly, we have

$$[\Gamma g\Gamma]\cdot[\Gamma h\Gamma] = \Theta^{-1}\left(\sum_{i,j}[\Gamma g_i h_j]\right).$$

It should be clear that this is associative, as multiplication in G is associative, and $[\Gamma] = [\Gamma e\Gamma]$ is a unit.

Now if M is *any* right G-module, and $m \in M^\Gamma$, we have

$$m|[\Gamma g\Gamma]|[\Gamma h\Gamma] = \left(\sum mg_i\right)|[\Gamma h\Gamma] = \sum mg_ih_j = m([\Gamma g\Gamma]\cdot[\Gamma h\Gamma]).$$

So M^Γ is a right $\mathcal{H}(G,\Gamma)$-module. □

Now in our construction of the product, we need to apply the map Θ^{-1}. It would be nice to have an explicit formula for the product in terms of double cosets. To do so, we choose representatives $S \subseteq G$ such that

$$G = \coprod_{g\in S} \Gamma g\Gamma.$$

Proposition. We write

$$\Gamma g \Gamma = \coprod_{i=1}^{r} \Gamma g_i$$
$$\Gamma h \Gamma = \coprod_{j=1}^{s} \Gamma h_j.$$

Then

$$[\Gamma g \Gamma] \cdot [\Gamma h \Gamma] = \sum_{k \in S} \sigma(k) [\Gamma k \Gamma],$$

where $\sigma(k)$ is the number of pairs (i, j) such that $\Gamma g_i h_j = \Gamma k$.

Proof. This is just a simple counting exercise. □

Of course, we could have taken this as the definition of the product, but we have to prove that this is independent of the choice of representatives g_i and h_j, and of S, and that it is associative, which is annoying.

6.2 Hecke operators on modular forms

We are now done with group theory. For the rest of the chapter, we take $G = \mathrm{GL}_2(\mathbb{Q})^+$ and $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. We are going to compute the Hecke algebra in this case.

The first thing to do is to identify what the single and double cosets are. Let's first look at the case where the representative lives in $\mathrm{GL}_2(\mathbb{Z})^+$. We let

$$\gamma \in \mathrm{GL}_2(\mathbb{Z})^+$$

with

$$\det \gamma = n > 0.$$

The rows of γ generate a subgroup $\Lambda \subseteq \mathbb{Z}^2$. If the rows of γ' also generate the same subgroup Λ, then there exists $\delta \in \mathrm{GL}_2(\mathbb{Z})$ with $\det \delta = \pm 1$ such that

$$\gamma' = \delta \gamma.$$

So we have $\deg \gamma' = \pm n$, and if $\det \gamma' = +n$, then $\delta \in \mathrm{SL}_2(\mathbb{Z}) = \Gamma$. This gives a bijection

$$\left\{ \begin{array}{c} \text{cosets } \Gamma\gamma \text{ such that} \\ \gamma \in \mathrm{Mat}_2(\mathbb{Z}),\ \det \gamma = n \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{c} \text{subgroups } \Lambda \subseteq \mathbb{Z}^2 \\ \text{of index } n \end{array} \right\}$$

What we next want to do is to pick representatives of these subgroups¡ hence the cosets. Consider an arbitrary subgroup $\Lambda \subseteq \mathbb{Z}^2 = \mathbb{Z}e_1 \oplus \mathbb{Z}e_2$. We let

$$\Lambda \cap \mathbb{Z}e_2 = \mathbb{Z} \cdot de_2$$

for some $d \geq 1$. Then we have

$$\Lambda = \langle ae_1 + be_2, de_2 \rangle$$

for some $a \geq 1$, $b \in \mathbb{Z}$ such that $0 \leq b < d$. Under these restrictions, a and b are unique. Moreover, $ad = n$. So we can define

$$\Pi_n = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathrm{Mat}_2(\mathbb{Z}) : a, d \geq 1, ad = n, 0 \leq b < d \right\}.$$

Then

$$\left\{ \gamma \in \mathrm{Mat}_2(\mathbb{Z}) : \det \gamma = n \right\} = \coprod_{\gamma \in \Pi_n} \Gamma\gamma.$$

These are the single cosets. How about the double cosets? The left hand side above is invariant under Γ on the left and right, and is so a union of double cosets.

Proposition.

(i) Let $\gamma \in \mathrm{Mat}_2(\mathbb{Z})$ and $\det \gamma = n \geq 1$. Then

$$\Gamma\gamma\Gamma = \Gamma \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix} \Gamma$$

for unique $n_1, n_2 \geq 1$ and $n_2 \mid n_1$, $n_1 n_2 = n$.

(ii)

$$\left\{ \gamma \in \mathrm{Mat}_2(\mathbb{Z}) : \det \gamma = n \right\} = \coprod \Gamma \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix} \Gamma,$$

where we sum over all $1 \leq n_2 \mid n_1$ such that $n = n_1 n_2$.

(iii) Let γ, n_1, n_2 be as above. if $d \geq 1$, then

$$\Gamma(d^{-1}\gamma)\Gamma = \Gamma \begin{pmatrix} n_1/d & 0 \\ 0 & n_2/d \end{pmatrix} \Gamma,$$

Proof. This is the Smith normal form theorem, or, alternatively, the fact that we can row and column reduce. □

Corollary. The set

$$\left\{ \left[\Gamma \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \Gamma \right] : r_1, r_2 \in \mathbb{Q}_{>0}, \frac{r_1}{r_2} \in \mathbb{Z} \right\}$$

is a basis for $\mathcal{H}(G, \Gamma)$ over $\mathbb{Z}$.

So we have found a basis. The next goal is to find a *generating set*. To do so, we define the following matrices:

For $1 \leq n_2 \mid n_1$, we define

$$T(n_1, n_2) = \left[\Gamma \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix} \Gamma \right]$$

For $n \geq 1$, we write

$$R(n) = \left[\Gamma \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} \Gamma \right] = \left[\Gamma \begin{pmatrix} n & 0 \\ 0 & n \end{pmatrix} \right] = T(n, n)$$

Finally, we define

$$T(n) = \sum_{\substack{1 \le n_2 | n_1 \\ n_1 n_2 = n}} T(n_1, n_2)$$

In particular, we have

$$T(1,1) = R(1) = 1 = T(1),$$

and if n is square-free, then

$$T(n) = T(n,1).$$

Theorem.

(i) $R(mn) = R(m)R(n)$ and $R(m)T(n) = T(n)R(m)$ for all $m, n \ge 1$.

(ii) $T(m)T(n) = T(mn)$ whenever $(m,n) = 1$.

(iii) $T(p)T(p^r) = T(p^{r+1}) + pR(p)T(p^{r-1})$ of $r \ge 1$.

Before we prove this theorem, we see how it helps us find a nice generating set for the Hecke algebra.

Corollary. $\mathcal{H}(G, \Gamma)$ is *commutative*, and is generated by $\{T(p), R(p), R(p)^{-1} : p \text{ prime}\}$.

This is rather surprising, because the group we started with was very non-commutative.

Proof. We know that $T(n_1, n_2)$, $R(p)$ and $R(p)^{-1}$ generate $\mathcal{H}(G, \Gamma)$, because

$$\left[\Gamma \begin{pmatrix} p & 0 \\ 0 & p \end{pmatrix} \Gamma\right] \left[\Gamma \begin{pmatrix} n_1 & 0 \\ 0 & n_2 \end{pmatrix} \Gamma\right] = \left[\Gamma \begin{pmatrix} pn_1 & 0 \\ 0 & pn_2 \end{pmatrix} \Gamma\right]$$

In particular, when $n_2 \mid n_1$, we can write

$$T(n_1, n_2) = R(n_2) T\left(\frac{n_1}{n_2}, 1\right).$$

So it suffices to show that we can produce any $T(n,1)$ from the $T(m)$ and $R(m)$. We proceed inductively. The result is immediate when n is square-free, because $T(n,1) = T(n)$. Otherwise,

$$\begin{aligned} T(n) &= \sum_{\substack{1 \le n_2 | n_1 \\ n_1 n_2 = n}} T(n_1, n_2) \\ &= \sum_{\substack{1 \le n_2 | n_1 \\ n_1 n_2 = n}} R(n_2) T\left(\frac{n_1}{n_2}, 1\right) \\ &= T(n,1) + \sum_{\substack{1 < n_2 | n_1 \\ n_1 n_2 = n}} R(n_2) T\left(\frac{n_1}{n_2}, 1\right). \end{aligned}$$

So $\{T(p), R(p), R(p)^{-1}\}$ does generate $\mathcal{H}(G, \Gamma)$, and by the theorem, we know these generators commute. So $\mathcal{H}(G, \Gamma)$ is commutative. □

We now prove the theorem.

Proof of theorem.

(i) We have

$$\left[\Gamma\begin{pmatrix}a & 0\\ 0 & a\end{pmatrix}\Gamma\right][\Gamma\gamma\Gamma] = \left[\Gamma\begin{pmatrix}a & 0\\ 0 & a\end{pmatrix}\gamma\Gamma\right] = [\Gamma\gamma\Gamma]\left[\Gamma\begin{pmatrix}a & 0\\ 0 & a\end{pmatrix}\Gamma\right]$$

by the formula for the product.

(ii) Recall we had the isomorphism $\Theta : \mathcal{H}(G,\Gamma) \mapsto \mathbb{Z}[\Gamma\backslash G]^{\Gamma}$, and

$$\Theta(T(n)) = \sum_{\gamma\in\Pi_n}[\Gamma\gamma]$$

for some Π_n. Moreover, $\{\gamma\mathbb{Z}^2 \mid \gamma\in\Pi_n\}$ is exactly the subgroups of $\mathbb{Z}^2$ of index n.

On the other hand,

$$\Theta(T(m)T(n)) = \sum_{\delta\in\Pi_m,\gamma\in\Pi_n}[\Gamma\delta\gamma],$$

and

$$\{\delta\gamma\mathbb{Z}^2 \mid \delta\in\Pi_m\} = \{\text{subgroups of } \gamma\mathbb{Z}^2 \text{ of index } n\}.$$

Since n and m are coprime, every subgroup $\Lambda\subseteq\mathbb{Z}^2$ of index mn is contained in a unique subgroup of index n. So the above sum gives exactly $\Theta(T(mn))$.

(iii) We have

$$\Theta(T(p^r)T(p)) = \sum_{\delta\in\Pi_{p^r},\gamma\in\Pi_p}[\Gamma\delta\gamma],$$

and for fixed $\gamma\in\Pi_p$, we know $\{\delta\gamma\mathbb{Z}^2 : \delta\in\Pi_{p^r}\}$ are the index p^r subgroups of $\mathbb{Z}^2$.

On the other hand, we have

$$\Theta(T(p^{r+1})) = \sum_{\varepsilon\in\Pi_{p^{r+1}}}[\Gamma\varepsilon],$$

where $\{\varepsilon\mathbb{Z}^2\}$ are the subgroups of $\mathbb{Z}^2$ of index p^{r+1}.

Every $\Lambda = \varepsilon\mathbb{Z}^2$ of index p^{r+1} is a subgroup of some index p subgroup $\Lambda'\in\mathbb{Z}^2$ of index p^r. If $\Lambda\not\subseteq p\mathbb{Z}^2$, then Λ' is unique, and $\Lambda' = \Lambda + p\mathbb{Z}^2$. On the other hand, if $\Lambda\subseteq p\mathbb{Z}^2$, i.e.

$$\varepsilon = \begin{pmatrix}p & 0\\ 0 & p\end{pmatrix}\varepsilon'$$

for some ε' of determinant p^{r-1}, then there are $(p+1)$ such Λ' corresponding to the $(p+1)$ order p subgroups of $\mathbb{Z}^2/p\mathbb{Z}^2$.

So we have

$$\begin{aligned}\Theta(T(p^r)T(p)) &= \sum_{\varepsilon\in\Pi_{p^{r+1}}\setminus(pI\Gamma_{p^{r-1}})}[\Gamma\varepsilon] + (p+1)\sum_{\varepsilon'\in\Pi_{p^{r-1}}}[\Gamma pI\varepsilon'] \\ &= \sum_{\varepsilon\in\Pi_{p^{r+1}}}[\Gamma\varepsilon] + p\sum_{\varepsilon'\in\Pi_{p^{r-1}}}[\Gamma pI\varepsilon'] \\ &= T(p^{r+1}) + pR(p)T(p^{r-1}).\end{aligned}$$

□

What's underlying this is really just the structure theorem of finitely generated abelian groups. We can replace GL_2 with GL_N, and we can prove some analogous formulae, only much uglier. We can also do this with $\mathbb{Z}$ replaced with any principal ideal domain.

Given all these discussion of the Hecke algebra, we let them act on modular forms! We write

$$V_k = \{\text{all functions } f : \mathcal{H} \to \mathbb{C}\}.$$

This has a right $G = \mathrm{GL}_2(\mathbb{Q})^+$ action on the right by

$$g : f \mapsto f\,|_k\, g.$$

Then we have $M_k \subseteq V_k^\Gamma$. For $f \in V_k^\Gamma$, and $g \in G$, we again write

$$\Gamma g\Gamma = \coprod \Gamma g_i,$$

and then we have

$$f\,|_k\,[\Gamma g\Gamma] = \sum f\,|_k\, g_i \in V_k^\Gamma.$$

Recall when we defined the slash operator, we included a determinant in there. This gives us

$$f\,|_k\, R(n) = f$$

for all $n \geq 1$, so the $R(n)$ act trivially. We also define

$$T_n = T_n^k : V_k^\Gamma \to V_k^\Gamma$$

by

$$T_n f = n^{k/2-1} f\,|_k\, T(n).$$

Since $\mathcal{H}(G,\Gamma)$ is commutative, there is no confusion by writing T_n on the left instead of the right.

Proposition.

(i) $T_{mn}^k T_m^k T_n^k$ if $(m,n) = 1$, and

$$T_{p^{r+1}}^k = T_{p^r}^k T_p^k - p^{k-1} T_{p^{r-1}}^k.$$

(ii) If $f \in M_k$, then $T_n f \in M_k$. Similarly, if $f \in S_k$, then $T_n f \in S_k$.

(iii) We have
$$a_n(T_m f) = \sum_{1 \le d | (m,n)} d^{k-1} a_{mn/d^2}(f).$$
In particular,
$$a_0(T_m f) = \sigma_{k-1}(m) a_0(f).$$

Proof.

(i) This follows from the analogous relations for $T(n)$, plus $f|R(n) = f$.

(ii) This follows from (iii), since T_n clearly maps holomorphic f to holomorphic f.

(iii) If $r \in \mathbb{Z}$, then
$$q^r \underset{k}{|} T(m) = m^{k/2} \sum_{e|m, 0 \le b < e} e^{-k} \exp\left(2\pi i \frac{mzr}{e^2} + 2\pi i \frac{br}{e}\right),$$
where we use the fact that the elements of Π_m are those of the form
$$\Pi_m = \left\{ \begin{pmatrix} a & b \\ 0 & e \end{pmatrix} : ae = m, 0 \le b < e \right\}.$$
Now for each fixed e, the sum over b vanishes when $\frac{r}{e} \notin \mathbb{Z}$, and is e otherwise. So we find
$$q^r \underset{k}{|} T(m) = m^{k/2} \sum_{e|(n,r)} e^{1-k} q^{mr/e^2}.$$
So we have
$$\begin{aligned} T_m(f) &= \sum_{r \ge 0} a_r(f) \sum_{e|(m,r)} \left(\frac{m}{e}\right)^{k-1} q^{mr/e^2} \\ &= \sum_{1 \le d | m} e^{k-1} \sum a_{ms/d}(f) q^{ds} \\ &= \sum_{n \ge 0} \sum_{d|(m,n)} d^{k-1} a_{mn/d^2} q^n, \end{aligned}$$
where we put $n = ds$.

□

So we actually have a rather concrete formula for what the action looks like. We can use this to derive some immediate corollaries.

Corollary. Let $f \in M_k$ be such that
$$T_n(f) = \lambda f$$
for some $m > 1$ and $\lambda \in \mathbb{C}$. Then

(i) For every n with $(n, m) = 1$, we have

$$a_{mn}(f) = \lambda a_n(f).$$

If $a_0(f) \neq 0$, then $\lambda = \sigma_{k-1}(m)$.

Proof. This just follows from above, since

$$a_n(T_m f) = \lambda a_n(f),$$

and then we just plug in the formula. □

This gives a close relationship between the eigenvalues of T_m and the Fourier coefficients. In particular, if we have an f that is an eigenvector for *all* T_m, then we have the following corollary:

Corollary. Let $0 \neq f \in M_k$, and $k \geq 4$ with $T_m f = \lambda_m f$ for all $m \geq 1$. Then

(i) If $f \in S_k$, then $a_1(f) \neq 0$ and

$$f = a_1(f) \sum_{n \geq 1} \lambda_n q^n.$$

(ii) If $f \notin S_k$, then

$$f = a_0(f) E_k.$$

Proof.

(i) We apply the previous corollary with $n = 1$.

(ii) Since $a_0(f) \neq 0$, we know $a_n(f) = \sigma_{k-1}(m) a_1(f)$ by (both parts of) the corollary. So we have

$$f = a_0(f) + a_1(f) \sum_{n \geq 1} \sigma_{k-1}(n) q^n = A + BE_k.$$

But since F and E_k are modular forms, and $k \neq 0$, we know $A = 0$.

□

Definition (Hecke eigenform)**.** Let $f \in S_k \setminus \{0\}$. Then f is a *Hecke eigenform* if for all $n \geq 1$, we have

$$T_n f = \lambda_n f$$

for some $l_n \in \mathbb{C}$. It is *normalized* if $a_1(f) = 1$.

We now state a theorem, which we cannot prove yet, because there is still one missing ingredient. Instead, we will give a partial proof of it.

Theorem. There exists a basis for S_k consisting of normalized Hecke eigenforms.

So this is actually typical phenomena!

Partial proof. We know that $\{T_n\}$ are commuting operators on S_k.

Fact. There exists an inner product on S_k for which $\{T_n\}$ are self-adjoint.

Then by linear algebra, the $\{T_n\}$ are simultaneously diagonalized. □

Example. We take $k = 12$, and $\dim S_{12} = 1$. So everything in here is an eigenvector. In particular,

$$\Delta(z) = \sum_{n\geq 1} \tau(n) q^n$$

is a normalized Hecke eigenform. So $\tau(n) = \lambda_n$. Thus, from properties of the T_n, we know that

$$\begin{aligned} \tau(mn) &= \tau(m)\tau(n)\\ \tau(p^{r+1}) &= \tau(p)\tau(p^r) - p^{11}\tau(p^{r-1}) \end{aligned}$$

whenever $(m, n) = 1$ and $r \geq 1$.

We can do this similarly for $k = 16, 18, 20, 22, 26$, because $\dim S_k = 1$, with Hecke eigenform $f = E_{k-12}\Delta$.

Unfortunately, when $\dim S_k(\Gamma(1)) > 1$, there do not appear to be any "natural" eigenforms. It seems like we just have to take the space and diagonalize it by hand. For example, S_{24} has dimension 2, and the eigenvalues of the T_n live in the strange field $\mathbb{Q}(\sqrt{144169})$ (note that 144169 is a prime), and not in $\mathbb{Q}$. We don't seem to find good reasons for why this is true. It appears that the nice things that happen for small values of k happen only because there is no choice.

7 L-functions of eigenforms

Given any modular form, or in fact any function with a q expansion

$$f = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma(1)),$$

we can form a Dirichlet series

$$L(f, s) = \sum_{n\geq 1} a_n n^{-s}.$$

Our goal of this chapter is to study the behaviour of this L-function. There are a few things we want to understand. First, we want to figure out when this series converges. Next, we will come up with an Euler product formula for the L-series. Finally, we come up with an analytic continuation and then a functional equation.

Afterwards, we will use such analytic methods to understand how E_2, which we figured is not a modular form, transforms under $\Gamma(1)$, and this in turns gives us a product formula for $\Delta(z)$.

Notation. We write $|a_n| = O(n^{k/2})$ if there exists $c \in \mathbb{R}$ such that for sufficiently large n, we have $|a_n| \leq cn^{k/2}$. We will also write this as

$$|a_n| \ll n^{k/2}.$$

The latter notation might seem awful, but it is very convenient if we want to write down a chain of such "inequalities".

Proposition. Let $f \in S_k(\Gamma(1))$. Then $L(f, s)$ converges absolutely for $\mathrm{Re}(s) > \frac{k}{2} + 1$.

To prove this, it is enough to show that

Lemma. If

$$f = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma(1)),$$

then

$$|a_n| \ll n^{k/2}$$

Proof. Recall from the example sheet that if $f \in S_k$, then $y^{k/2}|f|$ is bounded on the upper half plane. So

$$|a_n(f)| = \left|\frac{1}{2\pi}\int_{|q|=r} q^{-n}\tilde{f}(q)\frac{\mathrm{d}q}{q}\right|$$

for $r \in (0, 1)$. Then for *any* y, we can write this as

$$\left|\int_0^1 e^{-2\pi i n(x+iy)} f(x + iy)\mathrm{d}x\right| \leq e^{2\pi n y} \sup_{0\leq x\leq 1} |f(x + iy)| \ll e^{2\pi n y} y^{-k/2}.$$

We now pick $y = \frac{1}{n}$, and the result follows. □

As before, we can write the L-function as an Euler product. This time it looks a bit more messy.

Proposition. Suppose f is a normalized eigenform. Then

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

This is a very important fact, and is one of the links between cusp forms and algebraic number theory.

There are two parts of the proof — a formal manipulation, and then a convergence proof. We will not do the convergence part, as it is exactly the same as for $\zeta(s)$.

Proof. We look at

$$(1 - a_p p^{-s} + p^{k-1-2s})(1 + a_p p^{-s} + a_{p^2} p^{-2s} + \cdots) \\ = 1 + \sum_{r \geq 2} (a_{p^r} + p^{k-1} a_{p^{r-2}} - a_p a_p^{r-1}) p^{-rs}.$$

Since we have an eigenform, all of those coefficients are zero. So this is just 1. Thus, we know

$$1 + a_p p^{-s} + a_{p^2} p^{-2s} + \cdots = \frac{1}{1 - a_p p^{-s} + p^{k-1-2s}}.$$

Also, we know that when $(m, n) = 1$, we have

$$a_{mn} = a_m a_n,$$

and also $a_1 = 1$. So we can write

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

□

We now obtain an analytic continuation and functional equation for our L-functions. It is similar to what we did for the ζ-function, but it is easier this time, because we don't have poles.

Theorem. If $f \in S_k$ then, $L(f, s)$ is entire, i.e. has an analytic continuation to all of C. Define

$$\Lambda(f, s) = (2\pi)^{-s} \Gamma(s) L(f, s) = M(f(iy), s).$$

Then we have

$$\Lambda(f, s) = (-1)^{k/2} \Lambda(f, k - s).$$

The proof follows from the following more general fact:

Theorem. Suppose we have a function

$$0 \neq f(z) = \sum_{n \geq 1} a_n q^n,$$

with $a_n = O(n^R)$ for some R, and there exists $N > 0$ such that

$$f \Big|_k \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right) = cf$$

for some $k \in \mathbb{Z}_{>0}$ and $c \in \mathbb{C}$. Then the function

$$L(s) = \sum_{n \geq 1} a_n n^{-s}$$

is entire. Moreover, $c^2 = (-1)^k$, and if we set

$$\Lambda(s) = (2\pi)^{-s}\Gamma(s)L(s), \quad \varepsilon = c \cdot i^k \in \{\pm 1\},$$

then

$$\Lambda(k - s) = \varepsilon N^{s-k/2}\Lambda(s).$$

Note that the condition is rather weak, because we don't require f to even be a modular form! If we in fact have $f \in S_k$, then we can take $N = 1, c = 1$, and then we get the desired analytic continuation and functional equation for $L(f, s)$.

Proof. By definition, we have

$$cf(z) = f \Big|_k \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right) = N^{-k/2} z^{-k} f\left(-\frac{1}{Nz}\right).$$

Applying the matrix once again gives

$$f \Big|_k \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right) \Big|_k \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right) = f \Big|_k \left(\begin{smallmatrix} -N & 0 \\ 0 & -N \end{smallmatrix}\right) = (-1)^k f(z),$$

but this is equal to $c^2 f(z)$. So we know

$$c^2 = (-1)^k.$$

We now apply the Mellin transform. We assume $\mathrm{Re}(s) \gg 0$, and then we have

$$\Lambda(f, s) = M(f(iy), s) = \int_0^\infty f(iy) y^s \, \frac{\mathrm{d}y}{y} = \left(\int_{1/\sqrt{N}}^\infty + \int_0^{1/\sqrt{N}}\right) f(iy) y^s \, \frac{\mathrm{d}y}{y}.$$

By a change of variables, we have

$$\begin{aligned} \int_0^{1/\sqrt{N}} f(iy) y^s \, \frac{\mathrm{d}y}{y} &= \int_{1/\sqrt{N}}^\infty f\left(\frac{i}{Ny}\right) N^{-s} y^{-s} \, \frac{\mathrm{d}y}{y} \\ &= \int_{1/\sqrt{N}}^\infty c i^k N^{k/2-s} f(iy) y^{k-s} \, \frac{\mathrm{d}y}{y}. \end{aligned}$$

So

$$\Lambda(f,s) = \int_{1/\sqrt{N}}^{\infty} f(iy)(y^s + \varepsilon N^{k/2-s} y^{k-s}) \frac{\mathrm{d}y}{y},$$

where

$$\varepsilon = i^k c = \pm 1.$$

Since $f \to 0$ rapidly for $y \to \infty$, this integral is an entire function of s, and satisfies the functional equation

$$\Lambda(f, k-s) = \varepsilon N^{s-\frac{k}{2}} \Lambda(f,s).$$

□

Sometimes, we absorb the power of N into Λ, and define a new function

$$\Lambda^*(f,s) = N^{s/2}\Lambda(f,s) = \varepsilon \Lambda^*(f, k-s).$$

However, we can't get rid of the ε.

What we have established is a way to go from modular forms to L-functions, and we found that these L-functions satisfy certain functional equations. Now is it possible to go the other way round? Given any L-function, does it come from a modular form? This is known as the *converse problem*r One obvious necessary condition is that it should satisfy the functional equation, but is this sufficient?

To further investigate this, we want to invert the Mellin transform.

Theorem (Mellin inversion theorem)**.** Let $f : (0,\infty) \to \mathbb{C}$ be a C^∞ function such that

- for all $N, n \geq 0$, the function $y^N f^{(n)}(y)$ is bounded as $y \to \infty$; and
- there exists $k \in \mathbb{Z}$ such that for all $n \geq 0$, we have $y^{n+k} f^{(n)}(y)$ bounded as $y \to 0$.

Let $\Phi(s) = M(f,s)$, analytic for $\mathrm{Re}(s) > k$. Then for all $\sigma > k$, we have

$$f(y) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} \Phi(s) y^{-s} \, \mathrm{d}s.$$

Note that the conditions can be considerably weakened, but we don't want to do so much analysis.

Proof. The idea is to reduce this to the inversion of the Fourier transform. Fix a $\sigma > k$, and define

$$g(x) = e^{2\pi\sigma x} f(e^{2\pi x}) \in C^\infty(\mathbb{R}).$$

Then we find that for any $N, n \geq 0$, the function $e^{Nx} g^{(n)}(x)$ is bounded as $x \to +\infty$. On the other hand, as $x \to -\infty$, we have

$$\begin{aligned} g^{(n)}(x) &\ll \sum_{j=0}^{n} e^{2\pi(\sigma+j)x} |f^{(j)}(e^{2\pi x})| \\ &\ll \sum_{j=0}^{n} e^{2\pi(\sigma+j)x} e^{-2\pi(j+k)x} \\ &\ll e^{2\pi(\sigma-k)x}. \end{aligned}$$

So we find that $g \in \mathcal{S}(\mathbb{R})$. This allows us to apply the Fourier inversion formula. By definition, we have

$$\begin{aligned}\hat{g}(-t) &= \int_{-\infty}^{\infty} e^{2\pi\sigma x} f(e^{2\pi x}) e^{2\pi i x t}\,\mathrm{d}x \\ &= \frac{1}{2\pi}\int_0^\infty y^{\sigma+it} f(y)\frac{\mathrm{d}y}{y} = \frac{1}{2\pi}\Phi(\sigma+it).\end{aligned}$$

Applying Fourier inversion, we find

$$\begin{aligned}f(y) &= y^{-\sigma} g\left(\frac{\log y}{2\pi}\right) \\ &= y^{-\sigma}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-2\pi i t(\log y/2\pi)}\Phi(\sigma+it)\,\mathrm{d}t \\ &= \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty}\Phi(s)y^{-s}\,\mathrm{d}s.\end{aligned}$$

□

We can now use this to prove a simple converse theorem.

Theorem. Let

$$L(s) = \sum_{n\geq 1} a_n n^{-s}$$

be a Dirichlet series such that $a_n = O(n^R)$ for some R. Suppose there is some even $k \geq 4$ such that

- $L(s)$ can be analytically continued to $\{\mathrm{Re}(s) > \frac{k}{2} - \varepsilon\}$ for some $\varepsilon > 0$;
- $|L(s)|$ is bounded in vertical strips $\{\sigma_0 \leq \mathrm{Re}\, s \leq \sigma_1\}$ for $\frac{k}{2} \leq \sigma_0 < \sigma_1$.
- The function
$$\Lambda(s) = (2\pi)^{-s}\Gamma(s)L(s)$$
satisfies
$$\Lambda(s) = (-1)^{k/2}\Lambda(k-s)$$
for $\frac{k}{2} - \varepsilon < \mathrm{Re}\, s < \frac{k}{2} + \varepsilon$.

Then

$$f = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma(1)).$$

Note that the functional equation allows us to continue the Dirichlet series to the whole complex plane.

Proof. Holomorphicity of f on $\mathcal{H}$ follows from the fact that $a_n = O(n^R)$, and since it is given by a q series, we have $f(z+1) = f(z)$. So it remains to show that

$$f\left(-\frac{1}{z}\right) = z^k f(z).$$

By analytic continuation, it is enough to show this for

$$f\left(\frac{i}{y}\right) = (iy)^k f(iy).$$

Using the inverse Mellin transform (which does apply in this case, even if it might not meet the conditions of the version we proved), we have

$$\begin{aligned} f(iy) &= \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} \Lambda(s) y^{-s}\,\mathrm{d}s \\ &= \frac{1}{2\pi i}\int_{\frac{k}{2}-i\infty}^{\frac{k}{2}+i\infty} \Lambda(s) y^{-s}\,\mathrm{d}s \\ &= \frac{(-1)^{k/2}}{2\pi i}\int_{\frac{k}{2}-i\infty}^{\frac{k}{2}+i\infty} \Lambda(k-s) y^{-s}\,\mathrm{d}s \\ &= \frac{(-1)^{k/2}}{2\pi i}\int_{\frac{k}{2}-i\infty}^{\frac{k}{2}+i\infty} \Lambda(s) y^{s-k}\,\mathrm{d}s \\ &= (-1)^{k/2} y^{-k} f\left(\frac{i}{y}\right). \end{aligned}$$

Note that for the change of contour, we need

$$\int_{\frac{k}{2}\pm iT}^{\sigma\pm iT} \Lambda(s) y^{-s}\,\mathrm{d}s \to 0$$

as $T \to \infty$. To do so, we need the fact that $\Gamma(\sigma + iT) \to 0$ rapidly as $T \to \pm\infty$ uniformly for σ in any compact set, which indeed holds in this case. □

This is a pretty result, but not really of much interest at this level. However, it is a model for other proofs of more interesting things, which we unfortunately would not go into.

Recall we previously defined the Eisenstein series E_2, and found that

$$E_2(z) = 1 - 24\sum_{n\geq 1} \sigma_1(n) q^n.$$

We know this is not a modular form, because there is no modular form of weight 2. However, E_2 does satisfy $E_2(z+1) = E(z)$, as it is given by a q-expansion. So we know that $E_2(-\frac{1}{z}) \neq z^2 E_2(z)$. But what is it?

We let

$$f(y) = \frac{1 - E_2(iy)}{24} = \sum_{n\geq 1} \sigma_1(n) e^{-2\pi n y}.$$

Proposition. We have

$$M(f, s) = (2\pi)^{-s}\Gamma(s)\zeta(s)\zeta(s-1).$$

This is a really useful result, because we understand Γ and ζ well.

Proof. Doing the usual manipulations, it suffices to show that

$$\sum \sigma_1(m)m^{-s} = \zeta(s)\zeta(s-1).$$

We know if $(m,n) = 1$, then

$$\sigma_1(mn) = \sigma_1(m)\sigma_1(n).$$

So we have

$$\sum_{m\geq 1} \sigma_1(m)m^{-s} = \prod_p (1 + (p+1)p^{-s} + (p^2+p+1)p^{-2s} + \cdots).$$

Also, we have

$$(1-p^{-s})(1 + (p+1)p^{-s} + (p^2+p+1)p^{-2s} + \cdots) \\ = 1 + p^{1-s} + p^{2-2s} + \cdots = \frac{1}{1-p^{1-s}}.$$

Therefore we find

$$\sum \sigma_1(m)m^{-s} = \zeta(s)\zeta(s-1).$$

□

The idea is now to use the functional equation for ζ and the inverse Mellin transform to obtain the transformation formula for E_2. This is the *reverse* of what we did for genuine modular forms. This argument is due to Weil.

Recall that we defined

$$\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma\left(\frac{s}{2}\right), \quad Z(s) = \Gamma_{\mathbb{R}}(s)\zeta(s).$$

Then we found the functional equation

$$Z(s) = Z(1-s).$$

Similarly, we defined

$$\Gamma_{\mathbb{C}}(s) = 2(2\pi)^{-s}\Gamma(s) = \Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s+1),$$

where the last equality follows from the duplication formula. Then we know

$$(2\pi)^{-s}\Gamma(s) = (2\pi)^{-s}(s-1)\Gamma(s-1) = \frac{s-1}{4\pi}\Gamma_{\mathbb{R}}(s)\Gamma_{\mathbb{R}}(s-1).$$

This implies we have the functional equation

Proposition.

$$M(f,s) = \frac{s-1}{4\pi}Z(s)Z(s-1) = -M(f,2-s).$$

This also tells us the function is holomorphic except for poles at $s = 0, 1, 2$, which are all simple.

Theorem. We have

$$f(y) + y^{-2} f\left(\frac{1}{y}\right) = \frac{1}{24} - \frac{1}{4\pi} y^{-1} + \frac{1}{24} y^{-2}.$$

Proof. We will apply the Mellin inversion formula. To justify this application, we need to make sure our f behaves sensibly ass $y \to 0, \infty$. We use the absurdly terrible bound

$$\sigma_1(m) \leq \sum_{1 \leq d \leq m} d \leq m^2.$$

Then we get

$$f^{(n)}(y) \ll \sum_{m \geq 1} m^{2+n} e^{-2\pi m y}$$

This is certainly very well-behaved as $y \to \infty$, and is $\ll y^{-N}$ for all N. As $y \to 0$, this is

$$\ll \frac{1}{(1 - e^{2\pi y})^{n+3}} \ll y^{-n-3}.$$

So f satisfies conditions of our Mellin inversion theorem with $k = 3$.

We pick any $\sigma > 3$. Then the inversion formula says

$$f(y) = \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} M(f, s) y^{-s} \, \mathrm{d}s.$$

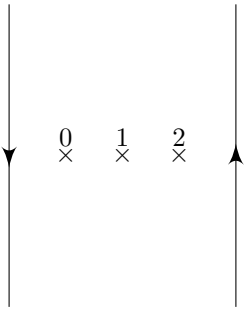
So we have

$$\begin{aligned} f\left(\frac{1}{y}\right) &= \frac{1}{2\pi i} \int_{\sigma - i\infty}^{\sigma + i\infty} -M(f, 2 - s) y^{s} \, \mathrm{d}s \\ &= \frac{-1}{2\pi i} \int_{2 - \sigma - i\infty}^{2 - \sigma + i\infty} M(f, s) y^{2 - s} \, \mathrm{d}s \end{aligned}$$

So we have

$$f(y) + y^{-2} f\left(\frac{1}{y}\right) = \frac{1}{2\pi i} \left(\int_{\sigma - i\infty}^{\sigma + i\infty} - \int_{2 - \sigma - i\infty}^{2 + \sigma + i\infty} \right) M(f, s) y^{-s} \, \mathrm{d}s.$$

This contour is pretty simple. It just looks like this:



Using the fact that $M(f, s)$ vanishes quickly as $|\operatorname{Im}(s)| \to \infty$, this is just the sum of residues

$$f(y) + y^{-2} f\left(\frac{1}{y}\right) = \sum_{s_0 = 0, 1, 2} \operatorname*{res}_{s = s_0} M(f, s) y^{-s_0}.$$

It remains to compute the residues. At $s = 2$, we have

$$\operatorname*{res}_{s=2} M(f,s) = \frac{1}{4\pi} Z(2) \operatorname*{res}_{s=1} Z(s) = \frac{1}{4\pi} \cdot \frac{\pi}{6} \cdot 1 = \frac{1}{24}.$$

By the functional equation, this implies

$$\operatorname*{res}_{s=0} M(f,s) = \frac{1}{24}.$$

Now it remains to see what happens when $s = 1$. We have

$$\operatorname*{res}_{s=1} M(f,s) = \frac{1}{4\pi} \operatorname*{res}_{s=1} Z(s) \operatorname*{res}_{s=0} Z(s) = -\frac{1}{4\pi}.$$

So we are done. □

Corollary.

$$E_2\left(-\frac{1}{z}\right) = z^2 E_2(z) + \frac{12z}{2\pi i}.$$

Proof. We have

$$\begin{aligned} E_2(iy) &= 1 - 24f(y) \\ &= 1 - 24y^{-2} f\left(\frac{1}{y}\right) - 1 + \frac{6}{\pi} y^{-1} + y^{-2} \\ &= y^{-2}\left(1 - 24f\left(\frac{1}{y}\right)\right) + \frac{6}{\pi} y^{-1} \\ &= y^{-2} E\left(\frac{-1}{iy}\right) + \frac{6}{\pi} y^{-1}. \end{aligned}$$

Then the result follows from setting $z = iy$, and then applying analytic continuiation. □

Corollary.

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

Proof. Let $D(z)$ be the right-hand-side. It suffices to show this is a modular form, since $S_{12}(\Gamma(1))$ is one-dimensional. It is clear that this is holomorphic on $\mathcal{H}$, and $D(z+1) = D(z)$. If we can show that

$$D \mid_{12} \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) = D,$$

then we are done. In other words, we need to show that

$$D\left(\frac{-1}{z}\right) = z^{12} D(z).$$

But we have

$$\begin{aligned} \frac{D'(z)}{D(z)} &= 2\pi i - 24 \sum_{m \geq 1} \frac{2\pi i m q}{1 - q^m} \\ &= 2\pi i \left(1 - 24 \sum_{m,d \geq 1} m q^{md}\right) \\ &= 2\pi i E_2(z) \end{aligned}$$

So we know

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}z}\left(\log D\left(-\frac{1}{z}\right)\right) &= \frac{1}{z^2}\frac{D'}{D}\left(-\frac{1}{z}\right)\\ &= \frac{1}{z^2}2\pi i E_2\left(-\frac{1}{z}\right)\\ &= \frac{D'}{D}(z) + 12\frac{\mathrm{d}}{\mathrm{d}z}\log z.\end{aligned}$$

So we know that

$$\log D\left(-\frac{1}{z}\right) = \log D + 12\log z + c,$$

for some locally constant function c. So we have

$$D\left(-\frac{1}{z}\right) = z^{12}D(z)\cdot C$$

for some other constant C. By trying $z = i$, we find that $C = 1$ (since $D(i) \neq 0$ by the infinite product). So we are done. □

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

8.1 Definitions

For the rest of the course, we are going to look at modular forms defined on some selected subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

We fix a subgroup $\Gamma \subseteq \Gamma(1)$ of finite index. For $\Gamma(1)$, we defined a modular form to a holomorphic function $f : \mathcal{H} \to \mathbb{C}$ that is invariant under the action of $\Gamma(1)$, and is holomorphic at infinity. For a general subgroup Γ, the invariance part of the definition works just as well. However, we need something stronger than just holomorphicity at infinity.

Before we explain what the problem is, we first look at some examples. Recall that we write $\bar{\Gamma}$ for the image of Γ in $\mathrm{PSL}_2(\mathbb{Z})$.

Lemma. Let $\Gamma \leq \Gamma(1)$ be a subgroup of finite index, and $\gamma_1, \cdots, \gamma_i$ be right coset representatives of $\bar{\Gamma}$ in $\overline{\Gamma(1)}$, i.e.

$$\overline{\Gamma(1)} = \coprod_{i=1}^{d} \bar{\Gamma}\gamma_i.$$

Then

$$\coprod_{i=1}^{d} \gamma_i \mathcal{D}$$

is a fundamental domain for Γ.

Example. Take

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

Recall there is a canonical map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{F}_p)$ that is surjective. Then $\Gamma^0(p)$ is defined to be the inverse image of

$$H = \left\{ \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \right\} \leq \mathrm{SL}_2(\mathbb{F}_p).$$

So we know

$$(\Gamma(1) : \Gamma^0(p)) = (\mathrm{SL}_2(\mathbb{F}_q) : H) = \frac{|\mathrm{SL}_2(\mathbb{F}_q)|}{|H|} = p + 1,$$

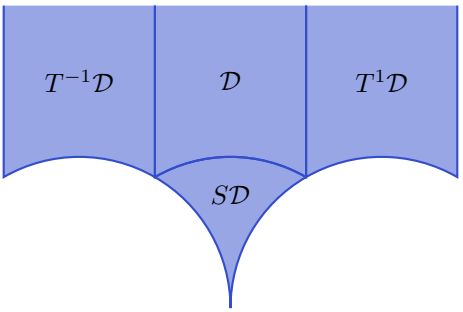
where the last equality follows from counting. In fact, we have an explicit choice of coset representatives

$$\mathrm{SL}_2(\mathbb{F}_p) = \coprod_{b \in \mathbb{F}_p} \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \coprod \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$$

Thus, we also have coset representatives of $\Gamma^0(p)$ by

$$\left\{ T^b = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{F}_p \right\} \cup \left\{ S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \right\}.$$

For example, when $p = 3$, then $b \in \{0, -1, +1\}$. Then the fundamental domain is

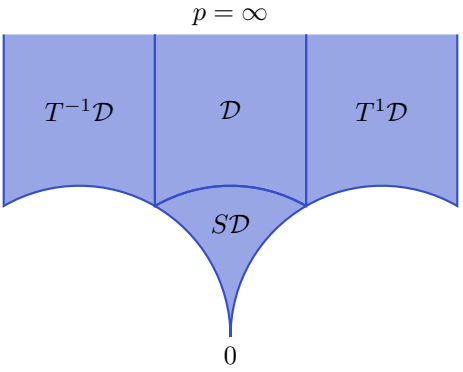


So in defining modular forms, we'll want to control functions as $z \to 0$ (in some way), as well as when $y \to \infty$. In fact, what we really need is that the function has to be holomorphic at all points in $\mathbb{P}^1(\mathbb{Q}) = \mathbb{Q} \cup \{\infty\}$. It happens that in the case of $\Gamma(1)$, the group $\Gamma(1)$ acts transitively on $\mathbb{Q} \cup \{\infty\}$. So by invariance of f under $\Gamma(1)$, being holomorphic at ∞ ensures we are holomorphic everywhere.

In general, we will have to figure out the orbits of $\mathbb{Q} \cup \{\infty\}$ under Γ, and then pick a representative of each orbit. Before we go into that, we first understand what the fundamental domain of Γ looks like.

Definition (Cusps)**.** The *cusps* of Γ (or $\bar{\Gamma}$) are the orbits of Γ on $\mathbb{P}^1(\mathbb{Q})$.

We would want to say a subgroup of index n has n many cusps, but this is obviously false, as we can see from our example above. The problem is that we should could each cusp with "multiplicity". We will call this the *width*. For example, in the fundamental domain above



In this case, we should count $p = \infty$ three times, and $p = 0$ once. One might worry this depends on which fundamental domain we pick for Γ. Thus, we will define it in a different way. From now on, it is more convenient to talk about $\bar{\Gamma}$ than Γ.

Since $\overline{\Gamma(1)}$ acts on $\mathbb{P}^1(\mathbb{Q})$ transitively, it actually just suffices to understand how to define the multiplicity for the cusp of ∞. The stabilizer of ∞ in $\overline{\Gamma(1)}$ is

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

For a general subgroup $\overline{\Gamma} \leq \overline{\Gamma(1)}$, the stabilizer of ∞ is $\overline{\Gamma}_\infty = \overline{\Gamma} \cap \overline{\Gamma(1)}_\infty$. Then

this is a finite index subgroup of $\overline{\Gamma(1)}_\infty$, and hence must be of the form

$$\overline{\Gamma}_\infty = \left\langle \pm \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \right\rangle$$

for some $m \geq 1$. We define the width of the cusp to be this m.

More generally, for an arbitrary cusp, we define the width by conjugation.

Definition (Width of cusp)**.** Let $\alpha \in \mathbb{Q} \cup \{\infty\}$ be a representation of a cusp of Γ. We pick $g \in \Gamma(1)$ with $g(\infty) = \alpha$. Then $\gamma(\alpha) = \alpha$ iff $g^{-1}\gamma g(\infty) = \infty$. So

$$g^{-1}\overline{\Gamma}_\alpha g = (\overline{g^{-1}\Gamma g})_\infty = \left\langle \pm \begin{pmatrix} 1 & m_\alpha \\ 0 & 1 \end{pmatrix} \right\rangle$$

for some $m_\alpha \geq 1$. This m_α is called the *width* of the cusp α (i.e. the cusp $\Gamma\alpha$).

The g above is not necessarily unique. But if g' is another, then

$$g' = g \begin{pmatrix} \pm 1 & n \\ 0 & \pm 1 \end{pmatrix}$$

for some $n \in \mathbb{Z}$. So m_α is independent of the choice of g.

As promised, we have the following proposition:

Proposition. Let Γ have ν cusps of widths $m_1, \cdots, m_\nu$. Then

$$\sum_{i=1}^{\nu} m_i = (\overline{\Gamma(1)} : \bar{\Gamma}).$$

Proof. There is a surjective map

$$\pi : \bar{\Gamma} \setminus \overline{\Gamma(1)} \to \text{cusps}$$

given by sending

$$\bar{\Gamma} \cdot \gamma \mapsto \bar{\Gamma} \cdot \gamma(\infty).$$

It is then an easy group theory exercise that $|\pi^{-1}([\alpha])| = m_\alpha$. □

Example. Consider the following subgroup

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

Then we have

$$\big(\overline{\Gamma(1)} : \overline{\Gamma_0(p)}\big) = \big(\Gamma(1) : \Gamma_0(p)\big) = p + 1.$$

We can compute

$$\Gamma_\infty = \left\langle -1, \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\rangle = \Gamma(1)_\infty.$$

So $m_\infty = 1$. But we also have

$$\Gamma_0 = \left\langle -1, \begin{pmatrix} 1 & 0 \\ p & 1 \end{pmatrix} \right\rangle,$$

and this gives $m_0 = p$. Since $p + 1 = p = 1$, these are the only cusps of $\Gamma_0(p)$. Likewise, for $\Gamma^0(p)$, we have $m_\infty = p$ and $m_0 = 1$.

Equipped with the definition of a cusp, we can now define a modular form!

Definition (Modular form)**.** Let $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ be of finite index, and $k \in \mathbb{Z}$. A *modular form of weight k on Γ* is a holomorphic function $f : \mathcal{H} \to \mathbb{C}$ such that

(i) $f \underset{k}{|} \gamma = f$ for all $\gamma \in \Gamma$.

(ii) f is holomorphic at the cusps of Γ.

If moreover,

(iii) f vanishes at the cusps of Γ,

then we say f is a *cusp form*.

As before, we have to elaborate a bit more on what we mean by (ii) and (iii). We already know what it means when the cusp is ∞ (i.e. it is $\Gamma\infty$). Now in general, we write our cusp as $\Gamma\alpha = \Gamma g(\infty)$ for some $g \in \Gamma(1)$.

Then we know

$$\bar{\Gamma}_\alpha = g \left\langle \pm \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \right\rangle g^{-1}.$$

This implies we must have

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \text{ or } -\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \in g^{-1}\Gamma_\alpha g.$$

Squaring, we know that we must have

$$\begin{pmatrix} 1 & 2m \\ 0 & 1 \end{pmatrix} \in g^{-1}\Gamma_\alpha g.$$

So we have

$$f \underset{k}{|} g \underset{k}{|} \left(\begin{smallmatrix} 1 & 2m \\ 0 & 1 \end{smallmatrix}\right) = f \underset{k}{|} g.$$

So we know

$$(f \underset{k}{|} g)(z + 2m) = (f \underset{k}{|} g)(z).$$

Thus, we can write

$$f \underset{k}{|} g = \tilde{f}_g(q) = \sum_{n \in \mathbb{Z}} (\text{constant}) q^{n/2m} = \sum_{\substack{n \in \mathbb{Q} \\ 2mn \in \mathbb{Z}}} a_{g,n}(f) q^n,$$

where we define

$$q^{a/b} = e^{2\pi i a z / b}.$$

Then f is holomorphic at the cusp $\alpha = g(\infty)$ if

$$a_{g,n}(f) = 0$$

for all $n < 0$, and *vanishes* at α if moreover

$$a_{g,0}(f) = 0.$$

Note that if $-I \in \Gamma$, then in fact

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \in g^{-1}\Gamma_\alpha g.$$

So the "q expansion at α" is actually series in $q^{1/m}$.

There is a better way of phrasing this. Suppose we have $g(\infty) = \alpha = g'(\infty)$, where $g' \in \mathrm{GL}_2(\mathbb{Q})^+$. Then we have

$$g' = gh$$

for some $h \in \mathrm{GL}_2(\mathbb{Q})^+$ such that $h(\infty) = \infty$. So we can write

$$h = \pm \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$$

where $a, b, d \in \mathbb{Q}$ and $a, d > 0$.

Then, we have

$$\begin{aligned} f \underset{k}{|} g' &= (f \underset{k}{|} g) \underset{k}{|} h \\ &= \sum_{\substack{n \in \mathbb{Q} \\ 2mn \in \mathbb{Z}}} a_{g,n}(f) q^n \underset{k}{|} \pm \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \\ &= (\pm i)^k \sum_n a_{g,n}(f) q^{an/d} e^{2\pi bn/d}. \end{aligned}$$

In other words, we have

$$f \underset{k}{|} g = \sum_{n \geq 0} c_n q^{rn}$$

for some positive $r \in \mathbb{Q}$. So condition (ii) is equivalent to (ii'):

(ii') For all $g \in \mathrm{GL}_2(\mathbb{Q})^+$, the function $f \underset{k}{|} g$ is holomorphic at ∞.

Note that (ii') doesn't involve Γ, which is convenient. Likewise, (iii) is consider to

(iii') $f \underset{k}{|} g$ vanishes at ∞ for all $g \in \mathrm{GL}_2(\mathbb{Q})^+$.

We can equivalently replace $\mathrm{GL}_2(\mathbb{Q})^+$ with $\mathrm{SL}_2(\mathbb{Z})$.

Modular form and cusp forms of weight k on Γ form a vector space $M_k(\Gamma) \supseteq S_k(\Gamma)$.

Recall that for $\Gamma = \Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$, we knew $M_k = 0$ if k is odd, because

$$f \underset{k}{|} (-I) = (-1)^k f.$$

More generally, if $-I \in \Gamma$, then $M_k = 0$ for all odd k. But if $-I \notin \Gamma$, then usually there can be non-zero forms of odd weight.

Let's see some examples of such things.

Proposition. Let $\Gamma \subseteq \Gamma(1)$ be of finite index, and $g \in G = \mathrm{GL}_2(\mathbb{Q})^+$. Then $\Gamma' = g^{-1}\Gamma g \cap \Gamma(1)$ also has finite index in $\Gamma(1)$, and if $f \in M_k(\Gamma)$ or $S_k(\Gamma)$, then $f \underset{k}{|} g \in M_k(\Gamma')$ or $S_k(\Gamma')$ respectively.

Proof. We saw that (G, Γ) has property (H). So this implies the first part. Now if $\gamma \in \Gamma'$, then $g\gamma g^{-1} \in \Gamma$. So

$$f \underset{k}{|} g\gamma g^{-1} = f \Rightarrow f \underset{k}{|} g \underset{k}{|} \gamma = f \underset{k}{|} g.$$

The conditions (ii') and (iii') are clear. □

This provides a way of producing a lot of modular forms. For example, we can take $\Gamma = \Gamma(1)$, and take

$$g = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}.$$

Then it is easy to see that $\Gamma' = \Gamma_0(N)$. So if $f(z) \in M_k(\Gamma(1))$, then $f(Nz) \in M_k(\Gamma_0(N))$. But in general, there are lots of modular forms in $\Gamma_0(N)$ that cannot be constructed this way.

As before, we don't have a lot of modular forms in low dimensions, and there is also an easy bound for those in higher dimensions.

Theorem. We have

$$M_k(\Gamma) = \begin{cases} 0 & k < 0 \\ \mathbb{C} & k = 0 \end{cases},$$

and

$$\dim_{\mathbb{C}} M_k(\Gamma) \leq 1 + \frac{k}{12}(\Gamma(1) : \Gamma).$$

for all $k > 0$.

In contrast to the case of modular forms of weight 1, we don't have explicit generators for this.

Proof. Let

$$\Gamma(1) = \coprod_{i=1}^{d} \Gamma\gamma_i.$$

We let

$$f \in M_k(\Gamma),$$

and define

$$\mathcal{N}_f = \prod_{1 \leq i \leq d} f \underset{k}{|} \gamma_i.$$

We claim that $\mathcal{N}_f \in M_{kd}(\Gamma(1))$, and $\mathcal{N}_f = 0$ iff $f = 0$. The latter is obvious by the principle of isolated zeroes.

Indeed, f is certainly holomorphic on $\mathcal{H}$, and if $\gamma \in \Gamma(1)$, then

$$\mathcal{N}_f \underset{k}{|} \gamma = \prod_i f \underset{k}{|} \gamma_i\gamma = \mathcal{N}_f.$$

As $f \in M_k(\Gamma)$, each $f \underset{k}{|} \gamma_i$ is holomorphic at ∞.

– If $k < 0$, then $\mathcal{N}_f \in M_{kd}(\Gamma(1)) = 0$. So $f = 0$.

– If $k \geq 0$, then suppose $\dim M_k(G) > N$. Pick $z_1, \cdots, z_N \in \mathcal{D} \setminus \{i, \rho\}$ distinct. Then there exists $0 \neq f \in M_k(\Gamma)$ with

$$f(z_1) = \cdots = f(z_N) = 0.$$

So

$$\mathcal{N}_f(z_1) = \cdots = \mathcal{N}_f(z_N) = 0.$$

Then by our previous formula for zeros of modular forms, we know $N \leq \frac{kd}{12}$. So $\dim M_k(\Gamma) \leq 1 + \frac{kd}{12}$.

– If $k = 0$, then $M_0(\Gamma)$ has dimension ≤ 1. So $M_0(\Gamma) = \mathbb{C}$.

□

8.2 The Petersson inner product

As promised earlier, we define an inner product on the space of cusp forms.

We let $f, g \in S_k(\Gamma)$. Then the function $y^k f(z)\overline{g(z)}$ is Γ-invariant, and is bounded on $\mathcal{H}$, since f and g vanish at cusps. Also, recall that $\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}$ is an $\mathrm{GL}_2(\mathbb{R})^+$-invariant measure. So we can define

$$\langle f, g\rangle = \frac{1}{v(\Gamma)} \int_{\Gamma\backslash\mathcal{H}} y^k f(z)\overline{g(z)} \frac{\mathrm{d}x\,\mathrm{d}y}{y^z} \in \mathbb{C},$$

where $\int_{\Gamma\backslash\mathcal{H}}$ means we integrate over any fundamental domain, and $v(\Gamma)$ is the volume of a fundamental domain,

$$v(\Gamma) = \int_{\Gamma\mathcal{H}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} = (\overline{\Gamma(1)} : \bar{\Gamma}) \int_{\mathcal{D}} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2}.$$

The advantage of this normalization is that if we replace Γ by a subgroup Γ' of finite index, then a fundamental domain for Γ' is the union of $(\bar{\Gamma} : \bar{\Gamma}')$ many fundamental domains for Γ. So the expression $(*)$ is *independent* of Γ, as long as both $f, g \in S_k(\Gamma)$.

This is called the *Petersson inner product*.

Proposition.

(i) $\langle \cdot, \cdot \rangle$ is a Hermitian inner product on $S_k(\Gamma)$.

(ii) $\langle \cdot, \cdot \rangle$ is invariant under translations by $\mathrm{GL}_2(\mathbb{Q})^+$. In other words, if $\gamma \in \mathrm{GL}_2(\mathbb{Q})^+$, then

$$\langle f\underset{k}{|}\gamma, g\underset{k}{|}\gamma\rangle = \langle f, g\rangle.$$

(iii) If $f, g \in S_k(\Gamma(1))$, then

$$\langle T_n f, g\rangle = \langle f, T_n g\rangle.$$

This completes our previous proof that the T_n can be simultaneously diagonalized.

Proof.

(i) We know $\langle f, g\rangle$ is $\mathbb{C}$-linear in f, and $\overline{\langle f, g\rangle} = \langle g, f\rangle$. Also, if $\langle f, f\rangle = 0$, then

$$\int_{\Gamma\backslash\mathcal{H}} y^{k-2}|f|^2 \,\mathrm{d}x\,\mathrm{d}y = 0,$$

but since f is continuous, and y is never zero, this is true iff f is identically zero.

(ii) Let $f' = f\big|_k \gamma$ and $g' = g\big|_k \gamma \in S_k(\Gamma')$, where $\Gamma' = \Gamma \cap \gamma^{-1}\Gamma\gamma$. Then

$$y^k f'\bar{g}' = y^k \frac{(\det\gamma)^k}{|cz+d|^{2k}} \cdot f(\gamma(z))\overline{g(\gamma(z))} = (\operatorname{Im}\gamma(z))^k f(\gamma(z))\overline{g(\gamma(z))}.$$

Now $\operatorname{Im}\gamma(z)$ is just the y of $\gamma(z)$. So it follows that Then we have

$$\begin{aligned}\langle f', g'\rangle &= \frac{1}{v(\Gamma')}\int_{\mathcal{D}_{\Gamma'}} y^k f\bar{g}\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}\Big|_{\gamma(z)}\\ &= \frac{1}{v(\Gamma')}\int_{\gamma(\mathcal{D}_{\Gamma'})} y^k f\bar{g}\frac{\mathrm{d}x\,\mathrm{d}y}{y^2}.\end{aligned}$$

Now $\gamma(\mathcal{D}_{\Gamma'})$ is a fundamental domain for $\gamma\Gamma'\gamma^{-1} = \gamma\Gamma\gamma^{-1}\Gamma$, and note that $v(\Gamma') = v(\gamma\Gamma'\gamma^{-1})$ by invariance of measure. So $\langle f', g'\rangle = \langle f, g\rangle$.

(iii) Note that T_n is a polynomial with integer coefficients in $\{T_p : p \mid n\}$. So it is enough to do it for $n = p$. We claim that

$$\langle T_p f, g\rangle = p^{\frac{k}{2}-1}(p+1)\langle f\big|_k \delta, g\rangle,$$

where $\delta \in \mathrm{Mat}_2(\mathbb{Z})$ is any matrix with $\det(\delta) = p$.

Assuming this, we let

$$\delta^a = p\delta^{-1} \in \mathrm{Mat}_2(\mathbb{Z}),$$

which also has determinant p. Now as

$$g\big|_k \left(\begin{smallmatrix} p & 0\\ 0 & p\end{smallmatrix}\right) = g,$$

we know

$$\begin{aligned}\langle T_p f, g\rangle &= p^{\frac{k}{2}-1}(p+1)\langle f\big|_k \delta, g\rangle\\ &= p^{\frac{k}{2}-1}(p+1)\langle f, g\big|_k \delta^{-1}\rangle\\ &= p^{\frac{k}{2}-1}(p+1)\langle f, g\big|_k \delta^a\rangle\\ &= \langle f, T_p g\rangle\end{aligned}$$

To prove the claim, we let

$$\Gamma(1)\begin{pmatrix} p & 0\\ 0 & 1\end{pmatrix}\Gamma(1) = \coprod_{0\le j\le p}\Gamma(1)\delta\gamma_i$$

for some $\gamma_i \in \Gamma(1)$. Then we have

$$\begin{aligned}\langle T_p f, g\rangle &= p^{\frac{k}{2}-1}\left\langle \sum_j f\,|_k\,\delta\gamma_j, g\right\rangle \\ &= p^{\frac{k}{2}-1}\sum_j \langle f\,|_k\,\delta\gamma_j, g\,|_k\,\gamma_j\rangle \\ &= p^{\frac{k}{2}-1}(p+1)\langle f\,|_k\,\delta, g\rangle,\end{aligned}$$

using the fact that $g\,|_k\,\gamma_j = g$.

□

8.3 Examples of modular forms

We now look at some examples of (non-trivial) modular forms for different subgroups. And the end we will completely characterize $M_2(\Gamma_0(4))$. This seems like a rather peculiar thing to completely characterize, but it turns out this understanding $M_2(\Gamma_0(4))$ can allow us to prove a rather remarkable result!

Eisenstein series

At the beginning of the course, the first example of a modular form we had was Eisenstein series. It turns out a generalized version of Eisenstein series will give us more modular forms.

Definition ($G_{\mathbf{r},k}$)**.** Let $k \geq 3$. Pick any vector $\mathbf{r} = (r_1, r_2) \in \mathbb{Q}^2$. We define

$$G_{\mathbf{r},k}(z) = \sum_{\mathbf{m}\in\mathbb{Z}^2}{}' \frac{1}{((m_1+r_1)z + m_2 + r_2)^k},$$

where $\sum'$ means we omit any $\mathbf{m}$ such that $\mathbf{m} + \mathbf{r} = \mathbf{0}$.

For future purposes, we will consider $\mathbf{r}$ as a row vector.

As before, we can check that this converges absolutely for $k \geq 3$ and $z \in \mathcal{H}$. This obviously depends only on $\mathbf{r}$ mod $\mathbb{Z}^2$, and

$$G_{\mathbf{0},k} = G_k.$$

Theorem.

(i) If $\gamma \in \Gamma(1)$, then

$$G_{\mathbf{r},k}\,|_k\,\gamma = G_{\mathbf{r}\gamma,k}.$$

(ii) If $N\mathbf{r} \in \mathbb{Z}^2$, then $G_{\mathbf{r},k} \in M_k(\Gamma(N))$.

Proof.

(i) If $g \in \mathrm{GL}_2(\mathbb{R})^+$ and $\mathbf{u} \in \mathbb{R}^2$, then

$$\frac{1}{(u_1 z + u_2)^k}\,|_k\,g = \frac{(\det g)^{k/2}}{((au_1 + cu_2)z + (bu_1 + du_2))^k} = \frac{(\det g)^{k/2}}{(v_1 z + v_2)^k},$$

where $\mathbf{v} = \mathbf{n} \cdot g$. So

$$\begin{aligned} G_{\mathbf{r},k}\big|_k \gamma &= \sum_{\mathbf{m}}{}' \frac{1}{(((\mathbf{m}+\mathbf{r})_1\gamma)z + ((\mathbf{m}+\mathbf{r})\gamma)_2)^k} \\ &= \sum_{\mathbf{m}'} \frac{1}{((m_1'+r_1')z + m_2' + r_2')^k} \\ &= G_{\mathbf{r}\gamma,k}(z), \end{aligned}$$

where $\mathbf{m}' = \mathbf{m}\gamma$ and $\mathbf{r}' = \mathbf{r}\gamma$.

(ii) By absolute convergence, $G_{\mathbf{r},k}$ is holomorphic on the upper half plane. Now if $N\mathbf{r} \in \mathbb{Z}^2$ and $\gamma \in \Gamma(N)$, then $N\mathbf{r}\gamma \equiv N\mathbf{r} \pmod N$. So $\mathbf{r}\gamma \equiv \mathbf{r} \pmod{\mathbb{Z}^2}$. So we have

$$G_{\mathbf{r},k}\big|_k \gamma = G_{\mathbf{r}\gamma,k} = G_{\mathbf{r},k}.$$

So we get invariance under $\Gamma(N)$. So it is enough to prove $G_{\mathbf{r},k}$ is holomorphic at cusps, i.e. $G_{\mathbf{r},k}\big|_k \gamma$ is holomorphic at ∞ for all $\gamma \in \Gamma(1)$. So it is enough to prove that for *all* $\mathbf{r}$, $G_{\mathbf{r},k}$ is holomorphic at ∞.

We can write

$$G_{\mathbf{r},k} = \left(\sum_{m_1+r_1>0} + \sum_{m_1+r_1=0} + \sum_{m_1+r_1<0}\right) \frac{1}{((m_1+r_1)z + m_2 + r_2)^k}.$$

The first sum is

$$\sum_{m_1+r_1>0} = \sum_{m_1>-r_1} \sum_{m_2\in\mathbb{Z}} \frac{1}{([(m_1+r_1)z + r_2] + m_2)^k}.$$

We know that $(m_1+r_1)z + r_2 \in \mathcal{H}$. So we can write this as a Fourier series

$$\sum_{m_1>-r_1} \sum_{d\geq 1} \frac{(-2\pi)^k}{(k-1)!} d^{k-1} e^{2\pi r_2 d} q^{(m_1+r_1)d}.$$

We now see that all powers of q are positive. So this is holomorphic.

The sum over $m_1 + r_1 = 0$ is just a constant. So it is fine.

For the last term, we have

$$\sum_{m_1+r_1<0} = \sum_{m_1<-r_1} \sum_{m_2\in\mathbb{Z}} \frac{(-1)^k}{((-m_1-r_1)z - r_2 - m_2)^k},$$

which is again a series in positive powers of $q^{-m_1-r_1}$.

□

ϑ functions

Our next example of modular forms is going to come from ϑ functions. We previously defined a ϑ function, and now we are going to call it ϑ_3:

$$\vartheta_3(z) = \vartheta(z) = \sum_{n\in\mathbb{Z}} e^{\pi i n^2 z} = 1 + 2\sum_{n\geq 1} q^{n^2/2}.$$

We proved a functional equation for this, which was rather useful.

Unfortunately, this is not a modular form. Applying elements of $\Gamma(1)$ to ϑ_3 will give us some new functions, which we shall call ϑ_2 and ϑ_4.

Definition (ϑ_2 and ϑ_4)**.**

$$\vartheta_2(z) = \sum_{n\in\mathbb{Z}} e^{\pi i (n+1/2)^2 z} = q^{1/8}\sum_{n\in\mathbb{Z}} q^{n(n+1)/2} = 2q^{1/8}\sum_{n\geq 0} q^{n(n+1)/2}$$

$$\vartheta_4(z) = \sum_{n\in\mathbb{Z}} (-1)^n e^{\pi i n^2 z} = 1 + 2\sum_{n\geq 1} (-1)^n q^{n^2/2}.$$

Theorem.

(i) $\vartheta_4(z) = \vartheta_3(z \pm 1)$ and $\theta_2(z+1) = e^{\pi i/4}\vartheta_2(z)$.

(ii)

$$\vartheta_3\left(-\frac{1}{z}\right) = \left(\frac{z}{i}\right)^{1/2}\vartheta_3(z)$$

$$\vartheta_4\left(-\frac{1}{z}\right) = \left(\frac{z}{i}\right)^{1/2}\vartheta_2(z)$$

Proof.

(i) Immediate from definition, e.g. from the fact that $e^{\pi i} = 1$.

(ii) The first part we've seen already. To do the last part, we use the Poisson summation formula. Let

$$h_t(x) = e^{-\pi t(x+1/2)^2} = g_t\left(x + \frac{1}{2}\right),$$

where

$$g_t(x) = e^{-\pi t x^2}.$$

We previously saw

$$\hat{g}_t(y) = t^{-1/2}e^{-\pi y^2/t}.$$

We also have

$$\begin{aligned}\hat{h}_t(y) &= \int e^{-2\pi i x y} g_t\left(x + \frac{1}{2}\right)\,\mathrm{d}x\\ &= \int e^{-2\pi i (x-1/2)y} g_t(x)\,\mathrm{d}x\\ &= e^{\pi i y}\hat{g}_t(y).\end{aligned}$$

So by the Poisson summation formula,

$$\vartheta_2(it) = \sum_{n\in\mathbb{Z}} h_t(n) = \sum_{n\in\mathbb{Z}} \hat{h}_t(n) = \sum_{n\in\mathbb{Z}} (-1)^n t^{-1/2} e^{-\pi n^2/t} = t^{-1/2}\vartheta_4\left(\frac{i}{t}\right).$$

□

There is also a ϑ_1, but we have $\vartheta_1 = 0$. Of course, we did not just invent a funny name for the zero function for no purpose. In general, we can define functions $\vartheta_j(u, z)$, and the ϑ functions we defined are what we get when we set $u = 0$. It happens that ϑ_1 has the property that

$$\vartheta_1(u, z) = -\vartheta_1(-u, z),$$

which implies $\vartheta_1(z) = 0$.

We now see that the action of $\mathrm{SL}_2(\mathbb{Z})$ send us between ϑ_2, ϑ_3 and ϑ_4, up to simple factors.

Corollary.

(i) Let

$$F = \begin{pmatrix} \vartheta_2^4 \\ \vartheta_3^4 \\ \vartheta_4^4 \end{pmatrix}.$$

Then

$$F(z+1) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} F, \quad z^{-2} F\left(-\frac{1}{z}\right) = \begin{pmatrix} 0 & 0 & -1 \\ 0 & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix} F$$

(ii) $\vartheta_j^4 \in M_2(\Gamma)$ for a subgroup $\Gamma \leq \Gamma(1)$ of finite index. In particular, $\vartheta_j^4 \big|_2 \gamma$ is holomorphic at ∞ for any $\gamma \in \mathrm{GL}_2(\mathbb{Q})^+$.

Proof.

(i) Immediate from the theorem.

(ii) We know $\overline{\Gamma(1)} = \langle S, T \rangle$, where $T = \pm\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $S = \pm\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. So by (i), there is a homomorphism $\rho : \Gamma(1) \to \mathrm{GL}_3(\mathbb{Z})$ and $\rho(-I) = I$ with

$$F \big|_2 \gamma = \rho(\gamma) F,$$

where $\rho(\gamma)$ is a signed permutation. In particular, the image of ρ is finite, so the kernel $\Gamma = \ker \rho$ has finite index, and this is the Γ we want.

It remains to check holomorphicity. But each ϑ_j is holomorphic at ∞. Since $F \big|_2 \gamma = \rho(\gamma) F$, we know $\vartheta_j^4 \big|_2$ is a sum of things holomorphic at ∞, and is hence holomorphic at ∞.

□

It would be nice to be more specific about exactly what subgroup ϑ_j^4 is invariant under. Of course, whenever $\gamma \in \Gamma$, then we have $\vartheta_j^4 \big|_2 \gamma = \vartheta_j^4$. But in fact ϑ_j^4 is invariant under a larger group.

To do this, it suffices to do it for $\vartheta^4 = \vartheta_3^4$, and the remaining follow by conjugation.

We introduce a bit of notation

Notation. We write

$$W_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}$$

Note that in general, W_N does *not* have determinant 1.

Theorem. Let $f(z) = \vartheta(2z)^4$. Then $f(z) \in M_2(\Gamma_0(4))$, and moreover, $f\big|_2 W_4 = -f$.

To prove this, we first note the following lemma:

Lemma. $\Gamma_0(4)$ is generated by

$$-I, \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 1 & 0 \\ 4 & 1 \end{pmatrix} = W_4 \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} W_4^{-1}.$$

Four is special. This is not a general phenomenon.

Proof. It suffices to prove that $\overline{\Gamma_0(4)}$ is generated by T and $U = \pm\left(\begin{smallmatrix} 1 & 0 \\ 4 & 1 \end{smallmatrix}\right)$.

Let

$$\gamma = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \overline{\Gamma_0(4)}.$$

We let

$$s(\gamma) = a^2 + b^2.$$

As c is even, we know $a \equiv 1 \pmod 2$. So $s(\gamma) \geq 1$, and moreover $s(\gamma) = 1$ iff $b = 0, a = \pm 1$, iff $\gamma = T^n$ for some n.

We now claim that if $s(\gamma) \neq 1$, then there exists $\delta \in \{T^{\pm 1}, U^{\pm 1}\}$ such that $s(\gamma\delta) < s(\gamma)$. If this is true, then we are immediately done.

To prove the claim, if $s(\gamma) \neq 1$, then note that $|a| \neq |2b|$ as a is odd.

- If $|a| < |2b|$, then $\min\{|b \pm a|\} < |b|$. This means $s(\gamma T^{\pm 1}) = a^2 + (b \pm a)^2 < s(\gamma)$.
- If $|a| > |2b|$, then $\min\{|a \pm 4b|\} < |a|$, so $s(\gamma U^{\pm 1}) = (a \pm 4b)^2 + b^2 < s(\gamma)$.

□

Proof of theorem. It is enough to prove that

$$f\big|_2 T = f\big|_2 U = f.$$

This is now easy to prove, as it is just a computation. Since $\vartheta(z+2) = \vartheta(z)$, we know

$$f\big|_2 T = f(z+1) = f(z).$$

We also know that

$$f\big|_2 W_4 = 4(4z)^{-2} f\left(\frac{-1}{4z}\right) = \frac{1}{4z^2}\vartheta\left(-\frac{1}{2z}\right)^4 = -f(z),$$

as

$$\vartheta\left(-\frac{1}{z}\right) = \left(\frac{z}{i}\right)^{1/2} \vartheta(z).$$

So we have

$$f\big|_2 U = f\big|_2 W_4 \big|_2 T^{-1} \big|_2 W_4 = (-1)(-1) f = f.$$

□

We look at $\Gamma_0(2)$ and $\Gamma_0(4)$ more closely. We have

$$\Gamma_0(2) = \left\{ \begin{pmatrix} a & b \\ 2c & d \end{pmatrix} \right\} = \left\{ \gamma \in \Gamma(1) : \gamma = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod 2 \right\}.$$

We know $|\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})| = 6$, and so $(\Gamma(1) : \Gamma_0(2)) = 3$. We have coset representatives

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

We also have a map

$$\begin{aligned} \Gamma_0(2) &\twoheadrightarrow \mathbb{Z}/2\mathbb{Z} \\ \begin{pmatrix} a & b \\ 2c & d \end{pmatrix} &\mapsto c, \end{aligned}$$

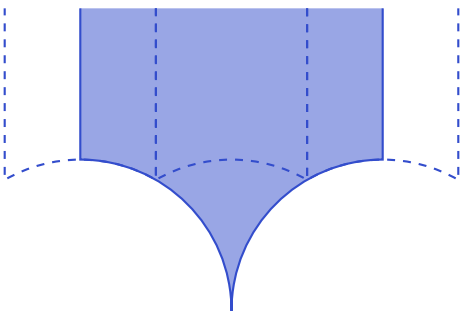
which one can directly check to be a homomorphism. This has kernel $\Gamma_0(4)$. So we know $(\Gamma_0(2) : \Gamma_0(4)) = 2$, and has coset representatives

$$I, \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$$

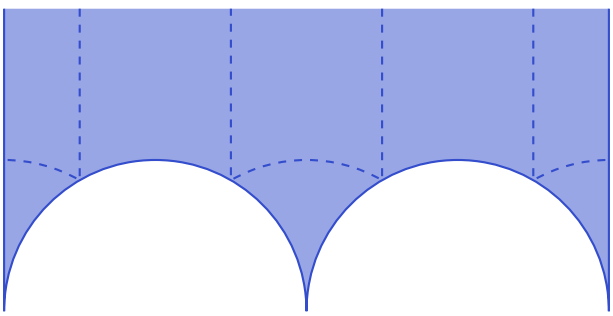
So

$$\overline{\Gamma_0(2)} = \left\langle T, \pm \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} = W_2 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} W_2^{-1} \right\rangle.$$

We can draw fundamental domains for $\Gamma^0(2)$:



and $\Gamma^0(4)$:



We are actually interested in $\Gamma_0(2)$ and $\Gamma_0(4)$ instead, and their fundamental domains look "dual".

Consider

$$g(z) = E_2|_2 \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right) - E_2 = 2E_2(2z) - E_2(z).$$

Recall that we had

$$E_2(z) = 1 - 24\sum_{n\geq 1} \sigma_1(n) q^n = z^{-2} E_2\left(-\frac{1}{z}\right) - \frac{12}{2\pi i z}.$$

Proposition. We have $g \in M_2(\Gamma_0(2))$, and $g|_2 W_2 = -g$.

Proof. We compute

$$\begin{aligned} g|_2 W_2 &= \frac{2}{(2z)^2} g\left(-\frac{1}{2z}\right) \\ &= \frac{1}{z^2} E_2\left(-\frac{1}{z}\right) - \frac{2}{(2z)^2} E_2\left(\frac{-1}{2z}\right) \\ &= E_2(z) + \frac{1}{2\pi i z} - 2\left(E_2(2z) + \frac{12}{2\pi i \cdot 2z}\right) \\ &= -g(z). \end{aligned}$$

We also have

$$g|_2 T = g(z+1) = g(z),$$

and so

$$g|_2 \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right) = g|_2 W_2 T^{-1} W_2^{-1} = g.$$

Moreover, g is holomorphic at ∞, and hence so is $g|_2 W_2 = -g$. So g is also holomorphic at $0 = W_2(\infty)$. As ∞ has width 1 and 0 has width 2, we see that these are all the cusps, and so g is holomorphic at the cusps. So $g \in M_2(\Gamma_0(2))$. □

Now we can do this again. We let

$$h = g(2z) = \frac{1}{2} g|_2 \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right) = 2E_2(4z) - E_2(2z).$$

Since $g \in M_2(\Gamma_0(2))$, this implies $h \in M_2(\Gamma_0(4)) \supseteq M_0(\Gamma_0(2))$.

The functions g and h are obviously linearly independent. Recall that we have

$$\dim M_2(\Gamma_0(4)) \leq 1 + \frac{k(\Gamma(1) : \Gamma_0(4))}{12} = 2.$$

So the inequality is actually an equality. We have therefore shown that

Theorem.

$$M_2(\Gamma_0(4)) = \mathbb{C}g \oplus \mathbb{C}h.$$

Recall we also found an $f(z) = \vartheta(2z)^4 \in M_2(\Gamma_0(4))$. So we know we must have

$$f = ag + bh$$

for some constants $a, b \in \mathbb{C}$.

It is easy to find what a and b are. We just write down the q-expansions. We have

$$\begin{aligned} f &= \vartheta(2z)^4 \\ &= (1 + 2q + 2q^4 + \cdots)^4 \\ &= 1 + 8q + 24q^2 + 32q^3 + \cdots \\ g &= 2E_2(2z) - E_2(z) \\ &= 1 + 24\sum_{n\geq 1} \sigma_1(n)(q^n - 2q^{2n}) \\ &= 1 + 24q + 24q^2 + 96q^3 + \cdots \\ h &= g(2z) \\ &= 1 + 24q^2 + .. \end{aligned}$$

By looking at the first two terms, we find that we must have

$$f = \frac{1}{3}g + \frac{2}{3}h = \frac{1}{3}(4E_2(4z) - E_2(z)) = 1 + 8\sum_{k\geq 1}\left(\sigma_1(n) - 4\sigma_1\left(\frac{n}{4}\right)\right)q^n,$$

where $\sigma_1\left(\frac{n}{4}\right) = 0$ if $\frac{n}{4} \notin \mathbb{Z}$.

But recall that

$$f = \left(\sum_{n\in\mathbb{Z}} q^{n^2}\right)^4 = \sum_{a,b,c,d\in\mathbb{Z}} q^{a^2+b^2+c^2+d^2} = \sum_{n\in\mathbb{N}} r_4(n)q^n,$$

where $r_4(n)$ is the number of ways to write n as a sum of 4 squares (where order matters). Therefore,

Theorem (Lagrange's 4-square theorem)**.** For all $n \geq 1$, we have

$$r_4(n) = 8\left(\sigma_1(n) - 4\sigma_1\left(\frac{n}{4}\right)\right) = 8\sum_{d|n4\nmid d} d.$$

In particular, $r_4(n) > 0$.

We could imagine ourselves looking at other sums of squares. For example, we can look instead at $\vartheta_2(2z)^2$, which turns out to be an element of $M_1(\Gamma_1(4))$, one can get a similar formula for the number of ways of writing n as a sum of 2 squares.

We can also consider higher powers, and then get *approximate formulae* for $r_{2k}(n)$, because the appropriate Eisenstein series no longer generate M_k. There may be a cusp form as well, which gives an error term.

In general, if we have

$$\gamma = \begin{pmatrix} a & b \\ Nc & d \end{pmatrix} \in \Gamma_0(N),$$

then we find

$$W_N \gamma W_N^{-1} \in \begin{pmatrix} d & -c \\ -Nb & a \end{pmatrix} \in \Gamma_0(N).$$

So W_N normalizes the group $\Gamma_0(N)$. Then if $f \in M_k(\Gamma_0(N))$, then $f|_k W_N \in M_k(\Gamma_0(N))$, and this also preserves cusp forms.

Moreover, we have

$$f|_k W_N^2 = f|_k \left(\begin{smallmatrix} -N & 0 \\ 0 & -N \end{smallmatrix}\right) = f,$$

as $-I \in \Gamma_0(N)$. So

$$M_k(\Gamma_0(N)) = M_k(\Gamma_0(N))^+ \oplus M_k(\Gamma_0(N))^-,$$

where we split into the (± 1)-eigenspaces for W_N, and the cusp forms decompose similarly. This W_N is the *Atkin-Lehner involution*. This is the "substitute" for the the operator $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ in $\Gamma(1)$.

9 Hecke theory for $\Gamma_0(N)$

Note that it is possible to do this for other congruence subgroups. The key case is

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

What is special about this? There are two things

- The map $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \mapsto d \bmod N$ is a homomorphism $\Gamma_0(N) \to (\mathbb{Z}/n\mathbb{Z})^\times$, and the kernel is $\Gamma_1(N)$.

So we can describe

$$S_k(\Gamma_1(N)) = \bigoplus_{\chi \in \widehat{(\mathbb{Z}/N\mathbb{Z})^\times}} S_k(\Gamma_1(N), \chi),$$

where $f \in S_k(\Gamma_1(n), \chi)$ if

$$f \mid_k \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \chi(d) f \text{ for all } \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(\mathbb{N}).$$

Of course, $S_k(\Gamma_1(N), \chi_{\text{trivial}}) = S_k(\Gamma_0(N))$. In general, everything we can do for $\Gamma_0(N)$ can be done for $S_k(\Gamma_1(N), \chi)$.

But why not study $\Gamma(N)$ itself? We can check that

$$\begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix} \Gamma(N) \begin{pmatrix} 1 & 0 \\ 0 & N^{-1} \end{pmatrix} \supseteq \Gamma_1(N^2).$$

So we can go from modular forms on $\Gamma(N)$ to ones on $\Gamma_1(N')$.

For various representation-theoretic reasons, things work much better on $\Gamma_1(N)$.

Last time, we used in various places the matrix

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

Then we decomposed

$$S_k(\Gamma_0(N)) = S_k(\Gamma_2(N))^+ \oplus S_k(\Gamma_0(N))^-,$$

according to the $\pm$-eigenspaces of the operator W_N. A while ago, we proved some theorem about the function equation of L-functions.

Theorem. Let $f \in S_k(\Gamma_0(N))^\varepsilon$, where $\varepsilon = \pm 1$. Then define

$$L(f, s) = \sum_{n \geq 1} a_n n^{-s}.$$

Then $L(f, s)$ is am entire function, and satisfies the functional equation

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

Proof. We have $f \mid_k W_N = \varepsilon f$, and then we can apply our earlier result. □

This is a rather remarkable thing. We didn't really use much about the properties of f.

Now we want to talk about Hecke operators on $\Gamma_0(N)$. These are a bit more complicate. It is much better to understand these in terms of representation theory instead, but that is outside the scope of the course. So we will just state the relevant definitions and results.

Recall that a modular form of level 1 is defined by the q-expansion, and if what we have is in fact a Hecke eigenform, then it suffices to know the Hecke eigenvalues, i.e. the values of a_p. We describe this as having "multiplicity one".

Theorem (Strong multiplicity one for $\mathrm{SL}_2(\mathbb{Z})$)**.** Let $f, g \in S_k(\Gamma(1))$ be normalized Hecke eigenforms, i.e.

$$f|T_p = \lambda_p f \qquad \lambda_p = a_p(f)$$
$$g|T_p = \mu_p g \qquad \mu_p = a_p(g).$$

Suppose there exists a finite set of primes S such that such that for all $p \notin S$, then $\lambda_p = \mu_p$. Then $f = g$.

Note that since the space of modular forms is finite dimensional, we know that the modular forms can only depend on finitely many of the coefficients. But this alone does not give the above result. For example, it might be that $a_2(f)$ is absolutely crucial for determining which modular form we are, and we cannot miss it out. The strong multiplicity one theorem says this does not happen.

Idea of proof. We use the functional equations

$$\Lambda(f, k-s) = (-1)^{k/2}\Lambda(f, s)$$
$$\Lambda(g, k-s) = (-1)^{k/2}\Lambda(g, s)$$

So we know

$$\frac{L(f, k-s)}{L(f, s)} = \frac{L(g, k-s)}{L(g, s)}.$$

Since these are eigenforms, we have an Euler product

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

and likewise for g. So we obtain

$$\prod_p \frac{1 - \lambda_p p^{s-k} + p^{2s-k-1}}{1 - \lambda_p p^{-s} + p^{k-1-2s}} = \prod_p \frac{1 - \mu_p p^{s-k} + p^{2s-k-1}}{1 - \mu_p p^{-s} + p^{k-1-2s}}.$$

Now we can replace this $\prod_p$ with $\prod_{p \in S}$. Then we have some very explicit rational functions, and then by looking at the appropriate zeroes and poles, we can actually get $\lambda_p = \mu_p$ for all p. □

This uses L-functions in an essential way.

The reason we mention this here is that a naive generalization of this theorem does not hold for, say, $\Gamma_0(N)$. To even make sense of this statement, we need to say what the Hecke operators are for $\Gamma_0(N)$. We are going to write the definition in a way as short as possible.

Definition (Hecke operators on $\Gamma_0(N)$)**.** If $p \nmid N$, we define

$$T_p f = p^{\frac{k}{2}-1}\left(f\Big|_k \left(\begin{smallmatrix} p & 0 \\ 0 & 1\end{smallmatrix}\right) + \sum_{k=0}^{p-1} f\Big|_k \left(\begin{smallmatrix} 1 & b \\ 0 & p\end{smallmatrix}\right)\right)$$

which is the same as the case with $\Gamma(1)$.

When $p \mid N$, then we define

$$U_p f = p^{\frac{k}{2}-1}\sum_{n=0}^{p-1} f\Big|_k \left(\begin{smallmatrix} 1 & b \\ 0 & p\end{smallmatrix}\right).$$

Some people call this T_p instead, and this is very confusing.

We can compute the effect on q-expansions as follows — when $p \nmid N$, then we have

$$a_n(T_p f) = a_{np}(f) + p^{k-1}a_{n/p}(f),$$

where the second term is set to 0 if $p \nmid n$. If $p \mid N$, then we have

$$a_n(U_p f) = a_{np}(f).$$

Proposition. T_p, U_p send $S_k(\Gamma_0(N))$ to $S_k(\Gamma_0(N))$, and they all commute.

Proof. T_p, U_p do correspond to double coset actions

$$\Gamma_0(N)\left(\begin{smallmatrix} 1 & 0 \\ 0 & p\end{smallmatrix}\right)\Gamma_0(N) = \begin{cases} \Gamma_0(N)\left(\begin{smallmatrix} p & 0 \\ 0 & 1\end{smallmatrix}\right) \amalg \coprod_b \Gamma_0(N)\left(\begin{smallmatrix} 1 & b \\ 0 & p\end{smallmatrix}\right) & p \nmid N \\ \coprod_b \Gamma_0(N)\left(\begin{smallmatrix} 1 & b \\ 0 & p\end{smallmatrix}\right) & p \mid N \end{cases}.$$

Commutativity is checked by carefully checking the effect on the q-expansions. □

However, these do not generate all the Hecke operators. For example, we have W_N!

Example. Consider $S_{12}(\Gamma_0(2))$. This contains $f = \Delta(z)$ and

$$g = f\Big|_{12} \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1\end{smallmatrix}\right) = 2^6\Delta(2z) = \Delta\Big|_{12} W_2,$$

using the fact that

$$\Delta\Big|_k \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0\end{smallmatrix}\right) = \Delta.$$

So the matrix of W_2 on $\operatorname{span}\{f, g\}$ is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0\end{pmatrix}.$$

We can write out

$$\begin{aligned} f &= \sum \tau(n) q^n = q - 24q^2 + 252q^3 - 1472q^4 + \cdots \\ g &= 2^6 \sum \tau(n) q^{2n} = 2^6(q^2 + 24q^4 + \cdots) \end{aligned}$$

So we find that

$$U_2 g = 2^6 f.$$

It takes a bit more work to see what this does on f. We in fact have

$$U_2 f = \sum \tau(2n) q^n = -24q - 1472q^4 + \cdots = -24f - 32g.$$

So in fact we have

$$U_2 = \begin{pmatrix} 24 & 64 \\ -32 & 0 \end{pmatrix}.$$

Now U_2 and W_2 certainly do not commute. So the Hecke algebra is *not* commutative. In fact, Δ generates a two-dimensional representation of the Hecke algebra.

This makes life much worse. When we did Hecke algebras for $\Gamma(1)$, all our representations are 1-dimensional, and we can just work with linear spans. Now everything has higher dimensional, and things go rather wrong. Similarly, we can consider $\Delta(dz) \in S_{12}(\Gamma_0(N))$ for any $d \mid N$, and this gives a whole bunch of things like this.

This turns out to be the only obstruction to the commutativity of the action of the Hecke algebra. We know $S_k(\Gamma_0(N))$ contains

$$\{f(dz) : f \in S_k(\Gamma_0(M)), dM \mid N, M \neq N\}.$$

We let $S_k(\Gamma_0(N))^{\text{old}}$ be the span of these. These are all the forms that come from a smaller level.

Now $S_k(\Gamma_0(N))$ has an inner product! So the natural thing to do is to consider the orthogonal complement of $S_k(\Gamma_0(N))^{\text{old}}$, and call it $S_k(\Gamma_0(N))^{\text{new}}$.

Theorem (Atkin–Lehner)**.** The Hecke algebra $\mathcal{H}(G, \Gamma_0(N))$ fixes $S_k(\Gamma_0(N))^{\text{new}}$ and $S_k(\Gamma_0(N))^{\text{old}}$, and on $S_k(\Gamma_0(N))^{\text{new}}$, it acts as a *commutative* subalgebra of the endomorphism ring, is closed under adjoint, and hence is diagonalizable. Moreover, strong multiplicity one holds, i.e. if S is a finite set of primes, and we have $\{\lambda_p : p \notin S\}$ given, then there exists at most one $N \geq 1$ and at most one $f \in S_k(\Gamma_0(N), 1)^{\text{new}}$ (up to scaling, obviously) for which

$$T_p f = \lambda_p f \text{ for all } p \nmid N, p \notin S.$$

10 Modular forms and rep theory

In this final chapter, we are going to talk about the relation between modular forms and representation theory. The words "representation theory" are a bit vague. We are largely going to talk about automorphic representations, and this is related to Langlands programme.

Recall that f is a modular form on $\mathrm{SL}_2(\mathbb{Z})$ if

(i) f is holomorphic $\mathcal{H} \to \mathbb{C}$

(ii) $f|_k\gamma = (cz+d)^{-k} f(\gamma(z)) = f(z)$ for all

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

(iii) It satisfies suitable growth conditions at the cusp ∞.

Let's look at the different properties in turn. The second is the modularity condition, which is what gave us nice properties like Hecke operators. The growth condition is some "niceness" condition, and for example this gives the finite-dimensionality of the space of modular forms.

But how about the first condition? It seems like an "obvious" condition to impose, because we are working on the complex plane. Practically speaking, it allows us to use the tools of complex analysis. But what if we dropped this condition?

Example. Recall that we had an Eisenstein series of weight 2,

$$E_2(z) = 1 - 24 \sum_{n \geq 1} \sigma_1(n) q^n.$$

This is *not* a modular form. Of course, we have $E_2(z) = E_2(z+1)$, but we saw that

$$E_2\left(-\frac{1}{z}\right) - z^2 E_2(z) = \frac{12z}{2\pi i} \neq 0.$$

However, we can get rid of this problem at the expense of making a non-holomorphic modular form. Let's consider the function

$$f(z) = \frac{1}{y} = \frac{1}{\mathrm{Im}(z)} = f(z+1).$$

We then look at

$$f\left(-\frac{1}{z}\right) - z^2 f(z) = \frac{|z|^2}{y} - \frac{z^2}{y} = \frac{z(\bar{z} - z)}{y} = -2iz.$$

Aha! This is the same equation as that for E_2 apart from a constant factor. So if we let

$$\tilde{E}_2(z) = E_2(z) - \frac{3}{\pi y},$$

then this satisfies

$$\tilde{E}_2(z) = \tilde{E}_2(z+1) = z^{-2} \tilde{E}_2\left(-\frac{1}{z}\right).$$

The term $\frac{3}{\pi y}$ certainly tends to 0 rapidly as $|z| \to \infty$, so if we formulate the growth condition in (iii) without assuming holomorphicity of f, then we will find that $\tilde{E}_2$ satisfies (ii) and (iii), but not (i). This is an example of a non-holomorphic modular form of weight 2.

Perhaps this is a slightly artificial example, but it is one.

Let's explore what happens when our functions satisfy (ii) and (iii), but not (i).

Definition (Non-holomorphic modular forms)**.** We let $W_k(\Gamma(1))$ be the set of all C^∞ functions $\mathcal{H} \to \mathbb{C}$ such that

(ii) $f|_k \gamma = f$ for all $\gamma \in \Gamma(1)$

(iii) $f(x + iy) = O(y^R)$ as $y \to \infty$ for some $R > 0$, and the same holds for all derivatives.

Note that the notation is not standard.

Before we proceed, we need to introduce some notation from complex analysis. As usual, we write $z = x + iy$, and we define the operators

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} + \frac{\partial}{i\partial y}\right)$$
$$\frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} - \frac{\partial}{i\partial y}\right).$$

We can check that these operators satisfy

$$\frac{\partial z}{\partial z} = \frac{\partial \bar{z}}{\partial \bar{z}} = 1, \quad \frac{\partial \bar{z}}{\partial z} = \frac{\partial z}{\partial \bar{z}} = 0.$$

Moreover, the Cauchy–Riemann equations just says $\frac{\partial f}{\partial \bar{z}} = 0$, and if this holds, then the complex derivative is just $\frac{\partial f}{\partial z}$. Thus, if we are working with potentially non-holomorphic functions on the complex plane, it is often useful to consider the operators $\frac{\partial}{\partial z}$ and $\frac{\partial}{\partial \bar{z}}$ separately.

Using this notation, given $f \in W_k$, we have

$$f \in M_k \iff \frac{\partial f}{\partial \bar{z}} = 0.$$

So suppose f is not holomorphic, then $\frac{\partial f}{\partial \bar{z}} \neq 0$. We can define a new operator by

$$L_k^*(f) = -2iy^2 \frac{\partial f}{\partial \bar{z}}.$$

Note that this is slightly strange, because we have a subscript k, but the function doesn't depend on k. Also, we put a star up there for some reason. It turns out there is a related operator called L_k, which does depend on k, and this L_k^* is a slight modification that happens not to depend on k.

This has the following properties:

Proposition.

- We have $L_k^* f = 0$ iff f is holomorphic.

– If $f \in W_K(\Gamma(1))$, then $g \equiv L_k^* f \in W_{k-2}(\Gamma(1))$.

Thus, L_k^* is a "lowering" operator.

Proof. The first part is clear. For the second part, note that we have

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

We now differentiate both sides with respect to $\bar{z}$. Then (after a bit of analysis), we find that

$$(c\bar{z} + d)^{-2} \frac{\partial f}{\partial \bar{z}}(\gamma(z)) = (cz + d)^k \frac{\partial f}{\partial \bar{z}}.$$

On the other hand, we have

$$(\operatorname{Im} \gamma(z))^2 = \frac{y^2}{|cz + d|^4}.$$

So we find

$$g(\gamma(z)) = -2i \frac{y^2}{|2z + d|^4} (c\bar{z} + d)^2 (cz + d)^k \frac{\partial f}{\partial \bar{z}} = (cz + d)^{k-2} g(z).$$

The growth condition is easy to check. □

Example. Consider $\tilde{E}_2$ defined previously. Since E_2 is holomorphic, we have

$$L_k^* \tilde{E}_2 = \frac{6i}{\pi} y^2 \frac{\partial}{\partial \bar{z}} \left(\frac{1}{y} \right) = \text{constant},$$

which is certainly a (holomorphic) modular form of weight 0.

In general, if $L_k^* f$ is actually holomorphic, then it is in M_{k-2}. Otherwise, we can just keep going! There are two possibilities:

– For some $0 \leq \ell < \frac{k}{2}$, we have

$$0 \neq L_{k-2\ell}^* \cdots L_{k-2}^* L_k^* f \in M_{k-2\ell}.$$

– The function $g = L_2^* L_4^* \cdots L_k^* f \in W_0(\Gamma(1))$, and is non-zero. In this case, $g(\gamma(z)) = g(z)$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

What does $W_0(\Gamma(1))$ look like? Since it is invariant under $\Gamma(1)$, it is just a C^∞ function on the fundamental domain $\mathcal{D}$ satisfying suitable C^∞ conditions on the boundary. This space is *huge*. For example, it contains any C^∞ function on $\mathcal{D}$ vanishing in a neighbourhood of the boundary.

This is too big. We want to impose some "regularity" conditions. Previously, we imposed a very strong regularity condition of holomorphicity, but this is too strong, since the only invariant holomorphic functions are constant.

A slightly weaker condition might be to require it is harmonic, i.e. $\tilde{\Delta} f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0$. However, the maximum principle also implies f must vanish.

A weaker condition would be to require that f is an *eigenfunction* of $\tilde{\Delta}$, but there is a problem that $\tilde{\Delta}$ is not invariant under $\Gamma(1)$. It turns out we need a slight modification, and take

$$\Delta = -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

It is a straightforward verification that this is indeed invariant under $\mathrm{SL}_2(\mathbb{R})$, i.e.

$$\Delta(f(\gamma(z))) = (\Delta f)(\gamma(z)).$$

In fact, this Δ is just the Laplacian under the hyperbolic metric.

Definition (Maass form)**.** A *Maass form* on $\mathrm{SL}_2(\mathbb{Z})$ is an $f \in W_0(\Gamma(1))$ such that

$$\Delta f = \lambda f$$

for some $\lambda \in \mathbb{C}$.

There are interesting things we can prove about these. Recall that our first examples of modular forms came from Eisenstein series. There are also non-holomorphic Eisenstein series.

Example. Let $s \in \mathbb{C}$ and $\mathrm{Re}(s) > 0$. We define

$$E(z,s) = \frac{1}{2}\sum_{(c,d)=1c,d\in\mathbb{Z}} \frac{y^s}{|cz+d|^{2s}} = \frac{1}{2}\sum_{\gamma=\pm\left(\begin{smallmatrix}*&*\\c&d\end{smallmatrix}\right)\in\pm\left(\begin{smallmatrix}1&*\\0&1\end{smallmatrix}\right)\backslash \mathrm{PSL}_2(\mathbb{Z})} (\mathrm{Im}\,\gamma(z))^s.$$

It is easy to see that this converges. From the second definition, we see that $E(z,s)$ is invariant under $\Gamma(1)$, and after some analysis, this is C^∞ and satisfies the growth condition.

Finally, we check the eigenfunction condition. We can check

$$\Delta y^s = -y^2\frac{\partial^2}{\partial y^2}(y^s) = s(1-s)y^s.$$

But since Δ is invariant under $\mathrm{SL}_2(\mathbb{R})$, it follows that we also have

$$\Delta E(z,s) = s(1-s)E(z,s).$$

In the case of modular forms, we studied the cusp forms in particular. To study similar phenomena here, we look at the Fourier expansion of f. We have the periodicity condition

$$f(x+iy+1) = f(x+iy).$$

Since this is not holomorphic, we cannot expand it as a function of $e^{2\pi iz}$. However, we can certainly expand it as a function in $e^{2\pi ix}$. Thus, we write

$$f(x+iy) = \sum_{n=-\infty}^{\infty} F_n(y)e^{2\pi inx}.$$

This looks pretty horrible, but now recall that we had the eigenfunction condition. Then we have

$$\lambda f = \Delta f = -y^2\sum_{n=-\infty}^{\infty}(F_n''(y) - 4\pi n^2 F_n(y))e^{2\pi inx}.$$

This tells us $F_n(y)$ satisfies the differential equation

$$-y^2F_n''(y) + (\lambda - 4\pi^2n^2y^2)F_n(y) = 0. \qquad (*)$$

It isn't terribly important what exactly the details are, but let's look what happens in particular when $n = 0$. Then we have

$$y^2 F_0'' + \lambda F_0 = 0.$$

This is pretty easy to solve. The general solution is given by

$$F_0 = Ay^s + By^{s'},$$

where s and $s' = 1 - s$ are the roots of $s(1 - s) = \lambda$.

What about the other terms? We see that if y is large, then, if we were an applied mathematician, then we would say the $\lambda F(y)$ term is negligible, and then the equation looks like

$$F_n''(y) = 4\pi^2 n^2 F(y).$$

This has two independent solutions, and they are $e^{\pm 2\pi n y}$. It is in fact true that the true solutions to the equation grow as $e^{\pm 2\pi n y}$ for large y. To satisfy the growth condition, we must only pick those that grow as $e^{-2\pi n y}$. We call this $\kappa_{|n|,\lambda}(y)$. These are known as the *Bessel functions*.

Thus, we find that we have

$$f(z) = \underbrace{Ay^s + By^{1-s}}_{\text{"constant term"}} + \sum_{n \neq 0} a_n(f) \kappa_{|n|,\lambda}(y) e^{2\pi i n x}.$$

The exact form isn't really that important. The point is that we can separate out these "constant terms". Then it is now not difficult to define cusp forms.

Definition (Cusp form)**.** A Maass form is a *cusp form* if $F_1 = 0$, i.e. $A = B = 0$.

Similar to modular forms, we have a theorem classifying Maass cusp forms.

Theorem (Maass)**.** Let $S_{\text{Maass}}(\Gamma(1), \lambda)$ be the space of Maass cusp forms with eigenvalue λ. This space is finite-dimensional, and is non-zero if and only if $\lambda \in \{\lambda_n : n \geq 0\}$, where $\{\lambda_n\}$ is a sequence satisfying

$$0 < \lambda_0 < \lambda_1 < \lambda_2 < \cdots \to \infty.$$

Given this, we can define Hecke operators just as for holomorphic forms (this is easier as $k = 0$), and most of the theory we developed for modular forms carry over.

Even though we proved all these very nice properties of these cusps forms, it took people a lot of time to actually come up with such a cusp form! Nowadays, we are able to compute these with the aid of computers, and there exists tables of λ's and Hecke eigenforms.

Now recall that we had this mysterious operator

$$L_k^* = -2iy^2 \frac{\partial}{\partial \bar{z}},$$

which had the property that if $f \underset{k}{|} \gamma = f$, then $(L_k^* f) \underset{k-2}{|} \gamma = (L_k^* f)$.

With a bit of experimentation, we can come up with something that raises the weight.

Definition (R_k^*)**.** Define

$$R_k^* = 2i\frac{\partial}{\partial z} + \frac{1}{y}k.$$

Now this has a property that

Proposition. If $f\underset{k}{|}\gamma = f$, then $(R_k^* f) \underset{k+2}{|} \gamma = R_k^* f$.

Note that this time, since we are differentiating against z, the $cz + d$ term will be affected, and this is where the $\frac{1}{y}k$ term comes in.

Suppose we have $f = f_0 \in M_k(\Gamma(1))$. Then we can apply R to it to obtain $f_1 = R_k^* f_0$. We can now try to apply L_{k+2}^* to it. Then we have

$$L_{k+2}^* R_k^* f = -2iy^2 \frac{\partial}{\partial \bar{z}}\left(2if' + \frac{k}{y}f\right) = -2iy^2 kf\frac{\partial y^{-1}}{\partial \bar{z}} = -kf.$$

So we don't get anything new.

But of course, we can continue in the other direction. We can recursively obtain

$$f_2 = R_{k+2}^* f_1, \quad f_3 = R_{k+4}^* f_2, \cdots.$$

Then we can compute L_{k+2n}^* and R_{k+2n}^* of these, and we find that

$$(R^*L^* - L^*R^*)f_n = (k + 2n)f_n.$$

This looks suspiciously like the representation of the Lie algebra of $\mathfrak{sl}_2$, where we have operators that raise and lower weights. The only slightly non-trivial part is that this is an infinite-dimensional representation, as we can keep on raising and (at least in general) it doesn't get to 0.

It turns out it is much easier to make sense of this by replacing functions on $\mathcal{H}$ with functions on $G = \mathrm{SL}_2(\mathbb{R})$. By the orbit-stabilizer theorem, we can write $\mathcal{H} = G/K$, where

$$K = \mathrm{SO}(2) = \{g \in \mathrm{SL}_2(\mathbb{R}) : g(i) = 1\} = \left\{r_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}\right\}.$$

Recall that we defined the function $j(\gamma, z) = cz + d$, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. This satisfied the property

$$j(\gamma\delta, z) = j(\gamma, \delta(z))j(\delta, z).$$

The main theorem is the following:

Proposition. For $\Gamma \subseteq \Gamma(1)$, there is a bijection between functions $f : \mathcal{H} \to \mathbb{C}$ such that $f\underset{k}{|}\gamma = f$ for all $\gamma \in \Gamma$, and functions $\Phi : G \to \mathbb{C}$ such that $\Phi(\gamma g) = \Phi(g)$ for all $\gamma \in \Gamma$ and $\Phi(gr_\theta) = e^{ik\theta}\Phi(g)$.

The real reason of this is that such an f is a section of a certain line bundle $\mathcal{L}_k$ on $\Gamma\backslash\mathcal{H} = \Gamma\backslash G/K$. The point is that this line bundle can be made trivial either by pulling to $\mathcal{H} = G/K$, or to $\Gamma\backslash G$. Of course, to actually prove it, we don't need such fancy language. We just need to write down the map.

Proof. Given an f, we define

$$\Phi(g) = (ci + d)^{-k} f(g(i)) = j(g, i)^{-k} f(g(i)).$$

We can then check that

$$\begin{aligned}\Phi(\gamma g) &= j(\gamma g, i)^{-k} f(\gamma(g(i))) \\ &= j(\gamma g, i)^{-k} j(\gamma, g(i))^k f(g(i)) \\ &= \Phi(g).\end{aligned}$$

On the other hand, using the fact that r_θ is in the stabilizer of i, we obtain

$$\begin{aligned}\Phi(g r_\theta) &= j(g r_\theta, i)^{-k} f(g r_\theta(i)) \\ &= j(g r_\theta, i)^{-k} f(g(i)) \\ &= j(g, r_\theta(i)) j(r_\theta, 1) f(g(i)) \\ &= \Phi(g) j(r_\theta, i)^{-k}.\end{aligned}$$

But $j(r_\theta, i) = -\sin\theta + \cos\theta$. So we are done. □

What we can do with this is that we can cast everything in the language of these functions on G. In particular, what do these lowering and raising operators do? We have our C^∞ function $\Phi : \Gamma\backslash G \to \mathbb{C}$. Now if $X \in \mathfrak{g} = \mathfrak{sl}_2(\mathbb{R})$, then this acts on Φ by differentiation, since that's how Lie algebras and Lie groups are related. Explicitly, we have

$$X\Phi = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \Phi(g e^{Xt}).$$

When we compute these things explicitly, we find that, up to conjugacy, L^* and R^* just correspond to the standard elements

$$X_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad X_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{sl}_2,$$

and we have

$$[X_+, X_-] = 2H, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Then the weight k is just corresponds to H.

What does the Laplacian correspond to? It corresponds to a certain product of these operators, the *Casimir operator*, given by

$$\Omega = X_+ X_- + X_- X_+ + \frac{1}{2} H^2.$$

This leads to the notion of automorphic forms.

Definition (Automorphic form)**.** An *automorphic form* on Γ is a C^∞ function $\Phi : \Gamma\backslash G \to \mathbb{C}$ such that $\Phi(g r_\theta) = e^{ik\theta}\Phi(g)$ for some $k \in \mathbb{Z}$ such that

$$\Omega\Phi = \lambda\Phi$$

for some $\lambda \in \mathbb{C}$, satisfying a growth condition given by

$$\Phi\begin{pmatrix} a & b \\ c & d \end{pmatrix} \leq \text{polynomial in } a, b, c, d.$$

The condition of this being a cusp function is then

$$\int_0^1 \Phi\left(\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} g\right)\, \mathrm{d}x = 0.$$

These things turn out to be exactly what we've looked at before.

Proposition. The set of cuspoidal automorphic forms bijects with representations of $\mathfrak{sl}_2$ generated by holomorphic cusp forms f and their conjugates $\bar{f}$, and Maass cusp forms.

The holomorphic cusp forms f generate a representation of $\mathfrak{sl}_2$ with lowest weight; The conjugates of holomorphic cusp forms generate those with highest weight, while the Maass forms generate the rest.

This is now completely susceptible to generalization. We can replace G, Γ with any semi-simple Lie group (e.g. $\mathrm{SL}_n(\mathbb{R})$, $\mathrm{Sp}_{2n}(\mathbb{R})$), and Γ by some arithmetic subgroup. This leads to the general theory of automorphic forms, and is one half of the Langlands' program.

Index