Modular Forms

Jacob Denson

October 8, 2016

Table Of Contents

Chapter 1

Modular Forms

1.1   An Example From Number Theory

Here is an interesting problem of number theory. The number 4 can written in 5 different ways as the sum of positive integers

$$4$$
$$1 + 3$$
$$2 + 2$$
$$1 + 1 + 2$$
$$1 + 1 + 1 + 1$$

The number 5 can be written in 7 different ways

$$5$$
$$4 + 1$$
$$3 + 2$$
$$3 + 1 + 1$$
$$2 + 2 + 1$$
$$2 + 1 + 1 + 1$$
$$1 + 1 + 1 + 1 + 1$$

We define the partition function $P(n)$ to be the number of ways a number n can be written as the sum of positive integers. The challenge is to determine the properties of $P(n)$, since we are very unlikely to find an

exact formula for the function. Thus $P(4) = 5$. The order of terms is irrelevant, so we normally assume the numbers are written in ascending or descending order. By bundling numbers which are repeated together, we can also define $P(n)$ as the number of ways we can write $n = \sum_{k=1}^{\infty} k a_k$, where $a_k \geqslant 0$. For 4, we have

$$1 \cdot 4$$
$$1 \cdot 1 + 1 \cdot 3$$
$$2 \cdot 2$$
$$2 \cdot 1 + 1 \cdot 2$$
$$4 \cdot 1$$

If we consider the infinite product

$$\prod_{n=1}^{\infty} \frac{1}{1 - q^n} = \prod_{n=1}^{\infty} (1 + q^n + q^{2n} + \dots) = \sum_{n=0}^{\infty} p_n q^n$$

which converges for $|q| < 1$, we see that $p_n = P(n)$. Thus η 'encodes' the partition function in someway. It turns out the Dedekind-Eta function

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

undoubtably linked to the partition function, is what is called a Modular form, an object with a beautiful general theory, which enables us to understand the properties of η, and thus of the partition function. These 'modular forms' also occur when counting curves in algebraic geometry, and when understanding the dimensions of finite simple groups, so they are an incredibly useful object to study. Because they occur so often in mathematics, the number theorist Joseph Eichler has been told to have said that there are five fundamental arithmetical operations – addition, subtraction, multiplication, division, and modular forms.

1.2 The Modular Group

The nicest holomorphic functions on $\mathbf{C}^{\infty}$ are the Möbius transformations, obtained via an action from $GL_2(\mathbf{C})$ by the map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} (z) = \frac{az + b}{cz + d}$$

3

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}(\infty) = \lim_{z \to \infty} \frac{az+b}{cz+d} = \frac{a}{c}$$

The kernel of the representation is $K = \left\{ \begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} : a \in \mathbf{C} - \{0\} \right\}$. Indeed, if $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in K$ then $b = 0$, for $M(0) = b/d = 0$. Similarily, $M(\infty) = a/c = \infty$, so $c = 0$. But then the transformation is of the form $z \mapsto (a/d)z$, so $a = d$. From K, we define the projective linear group

$$PGL_2(\mathbf{C}) = GL_2(\mathbf{C})/K$$

which now acts faithfully on $\mathbf{C}^\infty$.

Modular forms arise from a study of the actions of certain Möbius transformations on the hyperbolic plane. We shall take the upper half of the complex plane, denoted $\mathbf{H}$, as our model, whose 'straight lines' consist of circular arcs passing through the origin. The biholomorphisms of $\mathbf{H}$ are exactly the Möbius transformations obtained from matrices with determinant one, and can therefore be described as the image of $SL_2(\mathbf{R})$ in $PGL_2(\mathbf{C})$, denoted $PSL_2(\mathbf{R})$ and called the projective special linear group. The special linear group maps the upper half plane to itself because if $ad - bc = 1$,

$$\begin{aligned} \text{Im}\left(\frac{az+b}{cz+d}\right) &= \frac{1}{2i} \frac{(az+b)(c\bar{z}+d) - (a\bar{z}+b)(cz+d)}{|cz+d|^2} \\ &= \frac{1}{2i} \frac{z - \bar{z}}{|cz+d|^2} = \frac{\text{Im}(z)}{|cz+d|^2} \end{aligned}$$

We shall, perhaps by a little notational abuse, denote an element of $PSL_2(\mathbf{R})$ as an ordinary matrix, even though elements of the group are really cosets. Note that the kernel of $SL_2(\mathbf{R})$ under the Möbius transformation representation is only $\{\pm 1\}$, so two matrices have equal actions on the Riemann sphere if and only if they can be obtained from one another by negation.

We define the modular group $\Gamma = PSL_2(\mathbf{Z})$ to be the discrete counterpart to the projective special linear group. To understand Γ, we study it's action on $\mathbf{H}$. One such way is to determine the equivalence classes of the orbits of Γ. A **fundamental region** of a group action on $\mathbf{C}$ is a subset $D \subset \mathbf{H}$ which contains a point in every orbit, which is unique to that orbit, except perhaps if it is on the boundary of D. Then D 'represents' the group action in some fashion, by paramaterizing it's orbit space. Perhaps the most 'fundamental' fundamental region is the set

$$D = \{z \in \mathbf{H} : |z| \geqslant 1, |\mathfrak{Re}(z)| \leqslant 1/2\}$$

4
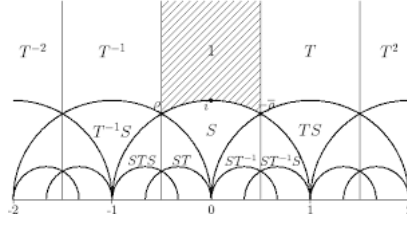
This image shows D's image under two important elements of Γ,

$$S = \begin{pmatrix} 0 & -1 \\ +1 & 0 \end{pmatrix} \qquad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

which induce the maps $z \mapsto -1/z$ and $z \mapsto z+1$ on $\mathbf{H}$.

Theorem 1.1. *Let D be the region described above.*

1. *For any $z \in \mathbf{H}$, there is $\gamma \in \Gamma$ for which $\gamma(z) \in D$.*

2. *If $\gamma(z) = w$, and $z, w \in D$, then z and w both occur on the boundary of D and are obtained from each other by reflection in the y axis.*

3. *For each z, the stabilizers Γ_z are trivial except that*

$$\Gamma_i = \{1, S\}$$
$$\Gamma_{e^{\pi i/3}} = \{1, ST, (ST)^2\}$$
$$\Gamma_{e^{2\pi i/3}} = \{1, TS, (TS)^2\}$$

Proof. Let $G = \langle S, T \rangle$. If D really was a fundamental region, then we could identify the point in D corresponding to each orbit Γx by taking the point in the strip $|\mathcal{R}e(z)| < 1$ such that $|\mathrm{Im}(\Gamma x)|$ is maximized. Since

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

Maximizing $\mathrm{Im}(\gamma z)$ happens by minimizing $|cz + d|$. For each M, there are only finitely many matrices with $|cz + d| < M$. We have

$$|cz + d| \geqslant \mathrm{Im}(z)|c| \qquad |cz + d| \geqslant |c\mathcal{R}e(z) + d| \geqslant |d| - |c\mathcal{R}e(z)|$$

There are only finitely many integers c for which $\mathrm{Im}(z)|c| < M$, and therefore only finitely many values of d for which $|d| - |c\mathcal{R}e(z)| < M$. Let γ

5

maximize the value of $\text{Im}(z)$. Without loss of generality, we may assume $|\mathfrak{Re}(\gamma z)| < 1/2$, for otherwise we may apply T to transport γz to this strip, without affecting the imaginary part. If $|\gamma z| < 1$, then

$$\text{Im}(S\gamma z) = \frac{\text{Im}(\gamma z)}{|\gamma z|} > \text{Im}(z)$$

contradicting the maximality of γ, so $\gamma z \in D$, and we have proved (i).

Now suppose that for $z, w \in D$,

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

We may assume $\text{Im}(w) \geqslant \text{Im}(z)$. Since

$$\text{Im}(w) = \frac{\text{Im}(z)}{|cz + d|^2}$$

$|cz + d| \leqslant 1$. Since $z \in D$, $\text{Im}(z) \geqslant \sqrt{1/2}$, so $|c| \leqslant \sqrt{2} < 2$, and we can split the proof into two cases ($c = \pm 1$ or $c = 0$):

- First assume $c = 0$. Then

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

  since a and a^{-1} are integers, we must have $a = \pm 1$, and we might as well let $a = 1$, since $\gamma = -\gamma$ in $PSL_2(\mathbf{R})$. Then

$$w = z + b$$

  Now $|b| = |\mathfrak{Re}(z - w)| \leqslant 1$, so either $b = 0$, and $w = z$, $b = 1$, and $w = z + 1$, or $b = -1$, and $w = z - 1$.

- Assume $c = 1$ (if $c = -1$, replace γ with $-\gamma$). We must have

$$|z + d|^2 = \text{Im}(z)^2 + (\mathfrak{Re}(z) + d)^2 \leqslant 1$$

  One possibility is that $d = 0$, so that $z \in \partial D \cap S^1$. Because

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

6

$w = a - 1/z$. Now $-1/z \in D$ is also on the unit circle, so either $a = 0$, and $w = -1/z$, which is a just a reflect in the y axis, $a = 1$, $z = e^{i\pi/3}$, $w = z$, and $\gamma = ST$, or $a = -1$, $z = e^{2i\pi/3}$, $w = z$, and $\gamma = S^{-1}T = (TS)^2$.

If $d = 1$, then

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

So

$$w = \frac{az + (a-1)}{z+1} = a - \frac{1}{z+1}$$

Now $|z+1| \geqslant 1$ if $z \in D$, so $|\frac{1}{z+1}| \leqslant 1$, and $a - \frac{1}{z+1}$ is in D if and only if $a = 0$, $z = e^{2\pi i/3}$, $w = z$, and $\gamma = TS$. If $a = 1$, $-\frac{1}{z+1} = e^{2\pi i/3}$, so $z = e^{2\pi i/3}$, and $w = e^{\pi i/3}$. If $a = -1$, $-\frac{1}{z+1} = e^{\pi i/3}$, which implies $z = e^{2\pi i/3} - 1 \notin D$.

Finally, let $d = -1$. Then

$$\gamma = \begin{pmatrix} a & -(1+a) \\ 1 & -1 \end{pmatrix}$$

so

$$w = \frac{az - (1+a)}{z-1} = a - \frac{1}{z-1}$$

As before, $|z-1| \geqslant 1$ if $z \in D$, so $|\frac{1}{z-1}| \leqslant 1$. Either $a = 0$, in which case $|z - 1| = 1$, so $z = w = e^{\pi i/3}$, $\gamma = TS^{-1} = (ST)^2$, or $a = 1$, in which case $\frac{-1}{z-1} = e^{2\pi i/3}$, implying $z = e^{2\pi i/3}$, $w = e^{\pi i/3}$, or $a = -1$, in which case $\frac{-1}{z-1} = e^{\pi i/3}$, which would imply $z = e^{2\pi i/3} - 1 \notin D$.

We have addressed all cases, which shows that (ii) and (iii) hold. $\qquad\square$

Corollary 1.2. $PSL_2(\mathbf{R})$ *is generated by S and T.*

Proof. Given $\gamma \in PSL_2(\mathbf{R})$ pick z in the interior of D, and let $w = \gamma z$. We have verified in the above proof that there is $\phi \in G$ for which $\phi w \in D$, and since the orbit is unique on the interior of D, $\phi w = z$. But then $\phi\gamma \in \Gamma_z = \{I\}$, so $\phi = \gamma^{-1}$, and so $\gamma \in G$. $\qquad\square$

1.3 Congruence Subgroups

An important family of subgroups of Γ are the principal congruence subgroups

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

It can be seen as the kernel of reduction of the coefficients of Γ modulo N, giving us the exact sequence

$$0 \to \Gamma(N) \to \Gamma \to SL_2(\mathbf{Z}_n) \to 0$$

A congruence subgroup of Γ is one which contains a principal congruence subgroup. Two important examples are

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : c \equiv 0 \text{ modulo } N \right\}$$

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a \equiv d \equiv 1 \text{ modulo } N \right\}$$

The matrices of $\Gamma_1(N)$ reduced modulo N are written in the form

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

for $a \in \mathbf{Z}_n$, and the matrices in the image $\Gamma_0(N)$ are therefore exactly the upper triangle matrices modulo N.

As we work with fewer and fewer elements of Γ, the action of Γ on $\mathbf{H}$ gets weaker and weaker, and the fundamental domain gets larger and larger. As an example, let's compute the fundamental domain of $\Gamma(2)$. The index of Γ in $\Gamma(2)$ is easily computed to be

$$[\Gamma : \Gamma(2)] = |SL_2(\mathbf{Z}_2)| = 6$$

Thus we may write $\Gamma = \bigcup_{i=1}^{6} \gamma_i \Gamma(2)$ as the disjoint union of cosets, for some $\gamma_i \in \Gamma$. We can choose these γ_i to be

$$\gamma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \gamma_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \gamma_3 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

$$\gamma_4 = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \quad \gamma_5 = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix} \quad \gamma_6 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

The elements must be in different cosets, since these elements differ modulo 2. It then follows that we have a fundamental domain

$$D_2 = \bigcup_{i=1}^{6} \gamma_i^{-1} D$$
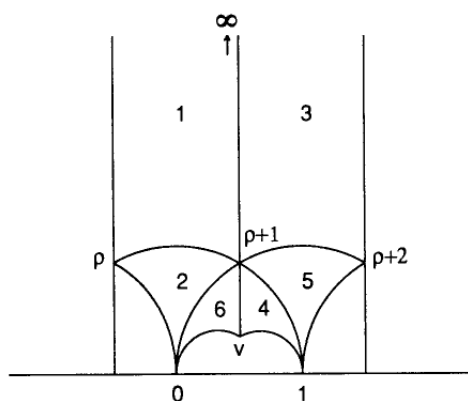
Certainly it contains a point in every orbit. It remains to check that repeated points occur only on the boundary, but I think it's pretty tedious to check this. The points 0, 1, and ∞ are known as cusps, for they are places where the fundamental domain can 'go off to infinity'. D only has one cusp, the one at ∞.

1.4 Modular Forms

A meromorphic function $f : \mathbf{H} \to \mathbf{C}$ is **weakly modular** of weight k (where k is an even integer) such that for any $\phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,

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

We assume k is even, for otherwise $(cz + d)^k \neq (-(cz + d))^k$, and so the theory becomes trivial. Since

$$\phi'(z) = \left(\frac{az + b}{cz + d}\right)\left(\frac{a}{az + b} - \frac{c}{cz + d}\right)$$

$$= \left(\frac{az + b}{cz + d}\right)\frac{1}{(az + b)(cz + d)}$$

$$= \frac{1}{(cz + d)^2}$$

we can write

$$\frac{f(\phi(z))}{f(z)} = (cz + d)^k = \phi'(z)^{-k/2}$$

which we see as a relation between differential forms.

$$\phi_*(f dz^{k/2}) = (f \circ \phi)(\phi')^{k/2} dz^{k/2} = f dz^{k/2}$$

The structure theorem for $SL_2(\mathbf{Z})$ simplifies the verification of the transformation properties of weakly modular functions.

Theorem 1.3. *f is weakly modular of weight k if and only if*

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

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

Proof. Certainly the relations hold if f is weakly modular. Conversely, suppose f satisfies the relations, and define

$$G = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma : f\left(\frac{az + b}{cz + d}\right) = (cz + d)^k f(z) \right\}$$

Certainly S, T, and T^{-1} are in G. If $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$ and $N = \begin{pmatrix} m & n \\ o & p \end{pmatrix}$, then

$$MN = \begin{pmatrix} am + bo & an + bp \\ cm + do & cn + dp \end{pmatrix}$$

and we know

$$f(MN(z)) = (cNz + d)^k f(Nz) = (cNz + d)^k (oz + p)^k f(z)$$

10

since

$$(cNz + d)(oz + p) = d(oz + p) + c(mz + n) = (cm + do)z + (cn + dp)$$

so $MN \in G$. We conclude that all monomials in the variables T, T^{-1}, S are in G. But all elements of Γ are of this form, so $G = \Gamma$. □

A weakly modular function is periodic, so we may consider the map

$$g(z) = e^{2\pi iz}$$

which maps **H** holomorphically onto the punctured unit disk. We may then expand $f \circ g^{-1}$ at the origin in a Laurent series, leading to an expansion of f of the form

$$f(z) = \sum_{k=-\infty}^{\infty} a_k e^{2\pi ikz}$$

We normally write $q = e^{2\pi iz}$, so that we 'expand f in q' to obtain

$$f(z) = \sum_{k=-n}^{\infty} a_k q^k$$

If f is meromorphic at ∞ (so that the coefficients a_k vanish for k small enough), we call it a **modular function** of weight k. If f is actually holomorphic at ∞, so that $a_k = 0$ for $k < 0$, and is holomorphic on **H**, then we call it a **modular form**. If, even further, we have $a_0 = 0$, then we say that f is a **cusp form**, for then f vanishes at ∞, and therefore at all cusps. We denote the set of modular forms, of weight k, by M_k, and the set of cusp forms by M_k°.

To obtain a geometric picture, it is useful to add a point ∞ at infinity on **H**. If we wish for the map $z \mapsto q$ to be a homeomorphism in $\mathbf{H} \cup \{\infty\}$, we put a topology on **H** by saying $z_i \to \infty$ if $\mathrm{Im}(z_i) \to \infty$. Equivalently, the neighbourhoods of ∞ are sets of the form

$$\{z \in \mathbf{H} : \mathrm{Im}(z) > M\}$$

as M ranges from 0 to ∞. This is certainly different to the relative topology induced by the Riemann sphere (if we need to explicitly identify the topology, we often say that $z \to i\infty$ if z converges to ∞ in the topology of

11

the upper half plane. If we wish to analyze the effects of the Möbius group on this set, then we are forced to also consider all rational points, for these points comprise the orbit of ∞ under the modular group. If we now want to make the maps $\gamma \in \Gamma$ homeomorphisms, then we must let the neighbourhoods of each rational point be the images of the neighbourhoods of ∞. These are exactly the circles whose boundary cross through the rational number. These points are known as **cusps**, and are the minimal number of points needed to compactify the orbit space of Γ, and still maintain a nice theory of modular forms. We can then compactify the fundamental region to $\overline{D} = D \cup \{\infty\}$, which is now the fundamental region of Γ over $\overline{\mathbf{H}}$.

The sets of modular functions, forms, and cusp-forms form complex vector spaces. In addition, if f is a modular (function or form) of weight k, and g modular of weight k', then fg is modular of weight $k + k'$, and if $g \neq 0$, f/g modular of weight $k - k'$ (g needs to have no zeroes in $\mathbf{H}$ for f/g to be a modular form).

1.5   Lattices

There is an interesting connection between the theory of modular forms and the theory of lattices. Recall that a **lattice** in a finite dimensional real vector-space V is an additive subgroup L of V such that one of the equivalent conditions holds.

1. L is discrete, and V/L is compact.

2. There is an $\mathbf{R}$ basis $e_1, \dots, e_n$ of V such that $L = \mathbf{Z}e_1 \oplus \dots \oplus \mathbf{Z}e_n$.

3. L is discrete, and spans V.

In our situation, we take $V = \mathbf{C}$, and we let $\mathcal{R}$ denote the set of all lattices on $\mathbf{C}$. Such lattices can be described by pairs of $\mathbf{R}$-independant complex numbers $(z, w) \in \mathbf{C}^2$. Given any particular pair, we let $L(z, w)$ denote the lattice generated by the two. Note that z and w are dependant if and only if $z/w \in \mathbf{R}$, so we may assume that $z/w \in \mathbf{H}$. We let

$$M = \{(z, w) \in \mathbf{C}^2 : z/w \in \mathbf{H}\}$$

And we then view L as a map from M to $\mathcal{R}$.

A lattice can be defined by many different pairs of complex numbers. For instance, if $w' = -z$, and $z' = w$, then $L(z, w) = L(z', w')$, and we find

$$z'/w' = -w/z = S(z/w)$$

Furthermore, if we let $z' = z + w$ and $w' = z$, then $L(z', w') = L(z, w)$ and

$$z'/w' = z/w + 1 = T(z/w)$$

So we begin to see Γ giving information about the various bases of any particular lattice. Indeed, if $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$, and $(z', w') = M(z, w)$ (where we define the action of Γ on M in the obvious way), then $L(z', w') = L(z, w)$, because $(z, w) = M^{-1}(z', w')$. What's more,

$$z'/w' = \frac{az + bw}{cz + dw} = \frac{a(z/w) + b}{c(z/w) + d} = M(z/w)$$

Theorem 1.4. *If $z/w, z'/w' \in \mathbf{H}$, then $L(z, w) = L(z', w')$ if and only if (z, w) is congruent to (z', w') relative to Γ.*

Proof. If $L(z, w) = L(z', w')$, we can write

$$z' = az + bw \qquad w' = cz + dw$$

It then follows that

$$z = \frac{az' - bw'}{ad - bc} \qquad w = \frac{dw' - cz'}{ad - bc}$$

Since all of the coefficients of these relations must be integer valued, we have $ad - bc \mid a, b, c, d$. We may assume the a, b, c, d are relatively prime, for if we write $a = \lambda a_1$, $b = \lambda b_1$, $c = \lambda c_1$, and $d = \lambda d_1$, with a_1, b_1, c_1, and d_1 relatively prime, we find

$$z = \frac{1}{\lambda} \frac{a_1 z' - b_1 w'}{a_1 d_1 - b_1 c_1} \qquad w = \frac{1}{\lambda} \frac{d_1 w' - c_1 z'}{a_1 d_1 - b_1 c_1}$$

from which we conclude that $\lambda \mid a_1, b_1, c_1, d_1$ again, hence $\lambda = 1$. It then follows that $ad - bc = \pm 1$. But if $ad - bc = -1$, we find z'/w' and z/w lie on opposite sides of the complex plane, contradicting the fact that they both lie on $\mathbf{H}$. Thus we find $(z', w') = M(z, w)$ for $M \in \Gamma$, and then $z'/w' = M(z/w)$. $\qquad \square$

13

We can therefore identify $\mathcal{R}$ with M/Γ. Now $\mathbf{C}^*$ acts on $\mathcal{R}$ by scaling, and this is the same map induced on the quotient M/Γ from M by scaling. Now we may identify $M/\mathbf{C}^*$ with $\mathbf{H}$, by the quotient map $(z, w) \mapsto z/w$, and the action of Γ on M induces the action of Γ on $\mathbf{H}$, so

$$\mathcal{R}/\mathbf{C}^* \cong (M/\Gamma)/\mathbf{C}^* \cong (M/\mathbf{C}^*)/\Gamma \cong \mathbf{H}/\Gamma$$

This observation can be used to construct modular functions from lattices. Let F be a complex valued function on $\mathcal{R}$. F is a modular function of weight k if $F(\lambda L) = \lambda^{-k}F(L)$. Then, pulling back, F defines a function on M by $F(z, w) = F(L(z, w))$, and F satisfies $F(\lambda z, \lambda w) = \lambda^{-k}F(z, w)$, $F(\gamma(z, w)) = F(z, w)$ for $\gamma \in \Gamma$. Choosing the right value of λ tells us that

$$F(z, w) = w^{-k}F(z/w, 1)$$

Defining $f(z) = F(z, 1)$, and understanding the Γ invariance of F shows

$$
\begin{aligned}
f\left(\frac{az + b}{cz + d}\right) &= F\left(\frac{az + b}{cz + d}, 1\right) \\
&= (cz + d)^k F(az + b, cz + d) \\
&= (cz + d)^k F(z, 1) \\
&= f(z)
\end{aligned}
$$

So f is weakly modular, provided it is also meromorphic. Conversely, a weakly modular function gives rise to a modular function on lattices.

1.6   Eisenstein Series

To build up a larger family of modular forms, let's build some concrete examples.

Lemma 1.5. *Let L be a lattice in* $\mathbf{C}$. *Then*

$$\sum_{\substack{z \in L \\ z \neq 0}} \frac{1}{|z|^s}$$

converges for $s > 2$.

14

Proof. Let $L = L(z, w)$. Assume that z/w is in the upper half of the plane and has non-negative real part (for the sum for z/w reflected in the y-axis is the same as the sum for z/w). Let $w = s/2$. Then, for $s > 2$,

$$\sum_{(n,m)\neq 0} \frac{1}{|m(z/w) + n|^s} \leqslant \sum_{(n,m)\neq 0} \frac{1}{(m^2|z/w|^2 + n^2)^w}$$

$$\leqslant \int_{x^2+y^2\geqslant 1} \frac{dx\,dy}{(|z/w|^2 x^2 + y^2)^w}$$

$$= |w/z| \int_{u^2|w/z|^2+y^2\geqslant 1} \frac{du\,dx}{(u^2 + y^2)^w}$$

$$\leqslant |w/z| \int_{u^2+y^2\geqslant \varepsilon} \frac{du\,dx}{(u^2 + y^2)^w}$$

$$= 2\pi|w/z| \int_{r=\varepsilon}^{\infty} r^{1-2w} dr < \infty$$

$\square$

This series gives us a modular function on the set of lattices. Namely, for even $k > 2$, we define

$$G_k(L) = \sum_{\substack{z\in L \\ z\neq 0}} \frac{1}{|z|^k}$$

These give rise to modular functions on **H** of the form

$$G_k(z) = \sum_{(n,m)\neq 0} \frac{1}{(mz + n)^k}$$

These functions are known as Eisenstein series.

Theorem 1.6. *G_k is a modular form of weight k, and $G_k(i\infty) = 2\zeta(k)$, where ζ is the Riemann-Zeta function.*

Proof. We know that G_k satisfies the symmetry condition defining modular functions, but we now need to show that it is holomorphic everywhere. Let z be a point in the fundamental domain. Notice that

$$|mz + n|^2 = (mz + n)(m\bar{z} + n)$$

$$= m^2|z|^2 + 2mn\,\mathfrak{Re}(z) + n^2$$

$$\geqslant m^2 + n^2 - mn = |m\rho + n|^2$$

15

$$|m(z+k)+n|^2 \geqslant |m\rho + (mk+n)|^2$$

$$|-m/z+n|^2 = \frac{1}{|z|^2}|n-mz|^2 \geqslant \frac{1}{|z|^2}|n-m\rho|^2$$

This implies that the series

$$\sum_{(n,m)\neq 0} \frac{1}{(mz+n)^k}$$

converges uniformly on D (and by the inequalities, it converges uniformly on a sufficiently small neighbourhood containing all boundary points). It is easy to see this implies the function is holomorphic everywhere. Finally, by uniform convergence on D,

$$G_k(i\infty) = \lim_{z\to i\infty} \sum_{(m,n)\neq 0} \frac{1}{(mz+n)^k} = \sum_{(m,n)\neq 0} \lim_{z\to i\infty} \frac{1}{(mz+n)^k}$$

$$= \sum_{n\neq 0} \frac{1}{n^k} = 2\zeta(k)$$

which is exactly what we wanted to prove. □

Example. *It is known that*

$$\zeta(4) = \frac{4\pi^4}{180} \quad \zeta(6) = \frac{8\pi^6}{3780}$$

If we define $g_2 = 60G_4$, $g_3 = 140G_6$, and $\Delta = (60G_4)^3 - 27(140G_6)^2$ (the coefficients emerge from the theory of elliptic curves), then we find that Δ, a modular form of order 12, satisfies $\Delta(i\infty) = 0$, so Δ is a cusp form.

It turns out that the coefficients of the G_k represent a certain arithmetic function. First, a proposition about the Riemann zeta function.

Lemma 1.7. *If k is a positive integer, then*

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

where B_k is the k'th Bernoulli number, defined by the expansion

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}$$

Proof. Since

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

We have

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

$$= \frac{1}{z} + \sum_{n=1}^{\infty} \frac{1}{n+z} - \frac{1}{n-z}$$

But also

$$\pi \cot(\pi z) = \pi i \frac{e^{i\pi z} + e^{-i\pi z}}{e^{\pi i z} - e^{-\pi i z}} = \pi i \frac{e^{2i\pi z} + 1}{e^{2\pi i z} - 1} = \pi i + \frac{2\pi i}{e^{2\pi i z} - 1}$$

Hence

$$\pi z \cot(\pi z) = \pi i z + 2\pi i \sum_{k=0}^{\infty} B_k \frac{(2\pi i z)^k}{k!}$$

□

Define the k'th divisor sum to be $\sigma_k(n) = \sum_{d \mid n} d^k$.

Theorem 1.8. *The Eisenstein series has the q-expansion*

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

Where B_k is the k'th Bernoulli number, defined by

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}$$

Proof. The logarithmic derivative of the product formula for sine...   □

17

1.7 Classification of Modular Forms

We shall now show that all functions can be represented in terms of the Eisenstein series. Recall that the order $\text{ord}(f, w)$ of a non-zero meromorphic function f at a point w is the unique integer k such that $(z-w)^k f(z)$ is a function non-zero and holomorphic at w. Note that if f is a weakly modular function, then $\text{ord}(f, w) = \text{ord}(f, \gamma w)$. Certainly this is true for $\gamma = T$, since f is periodic, and

$$
w^m f(-1/z + w) = w^m f\left(-\frac{1 - wz}{z}\right)
$$
$$
= w^m \left(\frac{z}{1 - wz}\right)^k f\left(\frac{z}{1 - wz}\right)
$$
$$
= \left(\frac{w(1 - wz)}{z}\right)^m \left(\frac{z}{1 - wz}\right)^k \left(\frac{z}{1 - wz}\right)^m f\left(\frac{z}{1 - wz}\right)
$$

This function is holomorphic at $w = 0$, provided m is the order of $f(z)$, and it is nonzero at $-1/z$, so this shows that the order is preserved by $\gamma = S$. The general case is then true because Γ is generated by S and T. Thus we may talk about the order of a point on $\mathbf{H}/\Gamma$. We also define, for a modular function, the order at ∞ to be the smallest integer whose corresponding q-coefficent is non-zero.
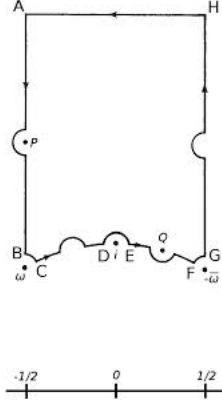
Theorem 1.9. *If $f \neq 0$ is a modular function of weight k, then*

$$
\text{ord}(f, \infty) + \frac{\text{ord}(f, i)}{2} + \frac{\text{ord}(f, e^{\frac{i\pi}{6}})}{3} + \sum_{\substack{p \in \mathbf{H}/\Gamma \\ p \neq i, e^{i\pi/6}}} \text{ord}(f, p) = \frac{k}{12}
$$

or perhaps more lucidly,

$$
\text{ord}(f, \infty) + \sum_{p \in \mathbf{H}/\Gamma} \frac{\text{ord}(f, p)}{\Gamma_p} = \frac{k}{12}
$$

Proof. The idea of the proof results from counting zeroes and poles via integrating the logarithmic derivative. Consider the oriented curve α shown in the image below.

18

which is tall enough that the region contains one of each zero and pole in each orbit (orbits cannot go to infinity because f is meromorphic there). We swerve around points $p \in \partial D$ which are zeroes and poles, such that only one of each point in the orbit class in contained in the region in question. The residue theorem performs its magic, telling us that

$$\frac{1}{2\pi i} \int_\alpha \frac{f'(z)}{f(z)} dz = \sum_{p \in \mathbf{H}/\Gamma, p \neq i, \omega} \operatorname{ord}(f, p)$$

Now let's calculate the integral explicitly. Let $q = e^{2\pi i z}$ and $\tilde{f}(q) = f(z)$. Then

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

Thus

$$\frac{1}{2\pi i} \int_{HA} \frac{f'(z)}{f(z)} dz = \frac{1}{2\pi i} \int_\omega \frac{\tilde{f}'(q)}{\tilde{f}(q)} dq$$

Where ω is now a clockwise rotation around the origin. The only possible poles and zeroes are at $q = 0$, so this evaluates to $-\operatorname{ord}(f, \infty)$. T takes AB to HG, so

$$\frac{1}{2\pi i} \int_{AB} \frac{f'(z)}{f(z)} dz = -\frac{1}{2\pi i} \int_{GH} \frac{f'(z)}{f(z)} dz$$

so the integrals cancel out when they are added together. Similarily, S takes CD to FE, and because $f(Sz) = z^k f(z)$, we find

$$\frac{1}{2\pi i} \left(\int_{CD} \frac{f'(z)}{f(z)} - \int_{FE} \frac{f'(z)}{f(z)} dz \right) = \frac{1}{2\pi i} \int_{CD} \left(\frac{f'(z)}{f(z)} - \frac{f'(Sz)}{f(Sz)} \right) dz$$

$$= -\frac{k}{2\pi i} \int_{CD} \frac{1}{z} dz$$

19

We will be taking the limit as we shrink the radius of the circles defining BC, DE, and FG to zero, in which case we trace out an angle of size $\pi/6$ in the limit, and so the value of the integral is $k/12$. Finally, we integrate BC, FG, and DE. In the limit, as we decrease the size of the circle, the circles have angles $\pi/6$, $\pi/2$, and $\pi/6$, so the path integrals integrate to $-\text{ord}(f,\omega)/6$, $-\text{ord}(f,i)/2$, and $-\text{ord}(f,-\overline{\omega})/6$. Summing up these integrals, and then combining them with our original deduction in terms of the residue theorem, we obtain the required formula. □

Note that this proof could have been simplified hugely by applying the theory of complex analytic structures.

The pole formula for modular forms of weight k places a vast restriction on the degrees of freedom of the functions. Note then that we only have zeroes, and so $\text{ord}(f,p) \geqslant 0$ for all p. If $f \neq 0$ is a modular form of weight $k \leqslant 2$, then

$$0 \leqslant \text{ord}(f,\infty) + \sum \frac{\text{ord}(f,p)}{\Gamma_p} = \frac{k}{12} < 0$$

which is clearly impossible, hence 0 is the only modular function of weight k.

If f has weight zero, then it certainly must have no zeroes on $\mathbf{H}$ (for the sum of zeroes must equal zero), but then if $f(w) = w'$, $f - w'$ is a form of weight zero, with a zero in $\mathbf{H}$, so $f - w = 0$, and so $f = w$ is a constant function.

Suppose that f is a non-zero modular form of weight k, where $k = 4$, $k = 6$, $k = 8$, and $k = 10$, and $k = 14$. Then the zero order formula tells us that f has no zeroes at ∞. But then $2\zeta(k)f - f(i\infty)E_k$ has a zero at $i\infty$, so $2\zeta(k)f = f(i\infty)E_k$, and so $M_k = \mathbf{C}E_k$.

Recall the discriminant Δ, a cusp form of order 12. Δ has a zero at ∞, which must be of order one by the zero summation formula, and can therefore have no zeroes anywhere else. Thus the map $f \mapsto f\Delta$ is an isomorphism from M_k to M_{k+12}°, and since $M_k = M_k^\circ \oplus \mathbf{C}E_k$, we find that all modular forms can be written as the sum and product of Eisenstein series. In particular, $\mathbf{C}[E_4, E_6] = \oplus_{k=0}^\infty M_{2k}$, which can shown by induction since $\Delta \in \mathbf{C}[E_4, E_6]$, and for any even $n = 4m + 6l$, we find $E_4^m E_6^l$ is a modular form of order n, non-zero at ∞, and $M_n = M_n^\circ \oplus \mathbf{C}E_4^m E_6^l$. Thus all elements

in M_n can be written

$$\sum_{4m+6l=n} a_{m,l} E_4^m E_6^l$$

1.8   The j-invariant

Let's get back to modular functions now. Define the function

$$j(z) = \frac{1728 g_2(z)^3}{\Delta(z)} = 1728 \frac{E_4(z)^3}{E_4(z)^3 - E_6(z)^2}$$

a modular function of weight zero.

Theorem 1.10. *j is a bijection from $\overline{H}/\Gamma$ to the Riemann sphere $\mathbf{PC} = \mathbf{C} \cup \{\infty\}$.*

Proof. j has a simple pole at infinity since g_2 does not vanish at infinity and Δ is a cusp form. For any $w \in \mathbf{C}$ the modular form $1728 g_2^3 - w\Delta$ must vanish at exactly one point, for it is a modular form of order 1. But this implies $j(z) - w = 0$ for exactly one value in $\overline{H}/\Gamma$. $\square$

Theorem 1.11. *The Modular functions of weight zero for Γ are precisely the rational functions of j.*

Proof. Certainly every element of $\mathbf{C}(j)$ is a modular function of weight zero, since j is a meromorphic modular function of weight zero. If f has poles $z_1, \ldots, z_l$ of order $n_1, \ldots, n_l$, then

$$g(z) = \prod_{n=1}^{\infty} [j(z) - j(z_1)]^{n_1} f(z)$$

is a modular function of weight zero with no poles in $\mathbf{H}$, so we may assume that f has no finite poles. If f has a pole at $i\infty$, define $g = \Delta^k f$, for a suitable value of k such that the pole at ∞ is removed, and so g is a modular form of weight $12k$. We may therefore write

$$f(z) = \sum_{4n+6m=12k} a_{n,m} \frac{E_4^n E_6^m}{\Delta^k}$$

We find $n = 3n'$, $m = 2m'$ by division laws, and E_4^3/Δ and E_6^2/Δ are each in $\mathbf{C}(j)$ (they are actually linear in j), and because each monomial in the sum is the product of such factors, we conclude that $\mathbf{C}(j)$ is the space of all weight zero modular forms. $\square$

The compactification of $\mathbf{H}/\Gamma$ can be given the structure of a complex analytic manifold, and these propositions imply a Biholomorphism with $\mathbf{CP}^1$ by j, and hence the only holomorphic functions on the compactification are rational functions of j, as the only holomorphic functions on $\mathbf{CP}^1$ are rational functions.

Chapter 2

JOIJ

2.1 The Dedekind Eta Function

We define the Dedekind Eta function to be

$$\eta(z) = e^{i\pi z/12} \prod_{n=1}^{\infty} (1 - e^{2\pi niz})$$

It is defined for $z \in \mathbf{H}$, because $\sum_{n=1}^{\infty} \ln(1 - e^{2\pi niz})$ exists when $\sum |e^{2\pi inz}|$ exists, because $\lim_{x \to 0} \ln(1 - x)/x$