Modular forms*

Avi Zeff

The goal of these notes is to summarize the theory of modular forms. We define the action of the modular group on the upper half-plane and modular forms as functions with good behavior under this action (or the action of certain subgroups), and study the space of such objects. We'll look at particular examples given by Eisenstein series, and then study certain operators (Hecke operators) on spaces of modular forms. Finally we'll generalize to "twisted" actions and give a way to check when a function built in a certain way is a modular form, which will be useful in practice. Finally we'll try and say something about how modular forms are an incarnation of a more abstract notion in a simple case: automorphic representations.

1. The modular action

We write $\mathcal{H}$ for the upper half-plane, i.e. the set of complex numbers with positive imaginary part. There is an interesting action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathcal{H}$ by

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

The restriction to $\mathrm{SL}_2(\mathbb{R})$ rather than $\mathrm{GL}_2(\mathbb{R})$ is necessary: for example,

$$\begin{pmatrix} 1 & \\ & -1 \end{pmatrix} \cdot i = \frac{i}{-1} = -i,$$

so the corresponding action of $\mathrm{GL}_2(\mathbb{R})$ would not preserve the upper half-plane.

This action factors through $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$, since $-1 = \begin{pmatrix} -1 & \\ & -1 \end{pmatrix}$ acts trivially on $\mathcal{H}$. The latter action is faithful, which will sometimes be more convenient to work with.

It is perhaps more natural to think about these linear fractional transformations as the action of $\mathrm{SL}_2(\mathbb{C})$ on $\mathbb{P}^1(\mathbb{C})$, which we can think of as the Riemann sphere, $\mathbb{C}$ plus a point at infinity. Once we fix $\mathcal{H}$ within $\mathbb{P}^1(\mathbb{C})$, the subgroup of $\mathrm{SL}_2(\mathbb{C})$ fixing $\mathcal{H}$ is precisely $\mathrm{SL}_2(\mathbb{R})$.

The action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathcal{H}$ is transitive; indeed the action of upper triangular matrices B is already transitive, since

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

for any real x and $y > 0$. If we allow all of $\mathrm{SL}_2(\mathbb{R})$ to act on i, we get a stabilizer consisting of matrices such that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot i = \frac{ai+b}{ci+d} = i,$$

*These notes are based on sections 1.2-1.5 of [2] and sections 2-3 of [1].

i.e. $ai + b = di - c$, so $c = -b$ and $a = d$, i.e. matrices of the form $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$. Adding the requirement that they have determinant 1, i.e. $a^2 + b^2 = 1$, this is exactly $\mathrm{SO}(2)$. This already gives a proof that $\mathrm{SL}_2(\mathbb{R}) = B \cdot \mathrm{SO}(2)$, known as the Iwasawa decomposition for $\mathrm{SL}_2(\mathbb{R})$.

For many purposes, $\mathrm{SL}_2(\mathbb{R})$ is too big, and we would like to be able to pick out discrete subgroups. An obvious such subgroup is $\mathrm{SL}_2(\mathbb{Z})$, which we will sometimes call $\Gamma(1)$. For every positive integer N, there is a canonical map $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ induced by $\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$; its kernel is a subgroup of $\mathrm{SL}_2(\mathbb{Z}) \subset \mathrm{SL}_2(\mathbb{R})$ consisting of matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ such that $a \equiv d \equiv 1 \pmod N$ and $b \equiv c \equiv 0 \pmod N$. We call this group $\Gamma(N)$; more generally, we say that a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup if it contains $\Gamma(N)$ for some N.

One can check that if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then the imaginary part of $g \cdot z$ is $|cz + d|^{-2} \cdot \mathrm{Im}(z)$. One can use this to prove the following.

Proposition 1.1. *The action of $\Gamma(1)$ on $\mathcal{H}$ is discontinuous, i.e. for every compact subset $K \subset \mathcal{H}$ there are only finitely many g such that $gK \cap K$ is nonempty.*

Proof. Fix a compact subset K of $\mathcal{H}$. Since K is closed, there is some $\epsilon > 0$ such that $\mathrm{Im}(z) > \epsilon$ for every $z \in K$. Fixing some z, the pairing $(c, d) \mapsto |cz + d|^2$ is a positive-definite quadratic form. Fix some $g \in \Gamma(1)$, and assume that $z \in gK \cap K$, so in particular $\mathrm{Im}(g \cdot z) = |cz + d|^{-2} \mathrm{Im}(z) > \epsilon$. On the other hand, for fixed z, as (c, d) grow $|cz + d|^{-2}$ tends to 0, so this inequality can only hold for finitely many pairs (c, d).

Suppose that g_1 and g_2 in $\Gamma(1)$ have the same bottom row (c, d). Then (keeping in mind that $\det g_1 = \det g_2 = 1$)

$$g' = g_1 g_2^{-1} = \begin{pmatrix} a_1 & b_1 \\ c & d \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ c & d \end{pmatrix}^{-1} = \begin{pmatrix} 1 & b_1 a_2 - a_1 b_2 \\ & 1 \end{pmatrix}$$

is of the form $g' = \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix}$ for some integer n. If z, $g_1 \cdot z$, and $g_2 \cdot z$ are all in K, then

$$g' \cdot g_2 \cdot z = g_1 z$$

is in K, so if we fix g_2 and view g_1 as $g' g_2$ then looking for g_1 is equivalent to looking for g' which keep $g_2 \cdot z$ in K. But since g' is of this form it acts by translation by n, so since K is bounded there are finitely many pairs g_1, g_2 keeping z in K with the same bottom row, and we know that there are finitely many bottom rows keeping a fixed z in K from above, from which the claim follows. □

We can even find a fundamental domain for the action of $\Gamma(1)$, i.e. an open subset $F \subset \mathcal{H}$ such that for every $z \in \mathcal{H}$, we can find $g \in \Gamma(1)$ such that $g \cdot z \in \overline{F}$ and if $z_1, z_2 \in F$ and $g \cdot z_1 = z_2$ for some $g \in \Gamma(1)$, then $z_1 = z_2$ and $g = \pm 1$. In this case we take F to be the set of $z \in \mathcal{H}$ such that $-\frac{1}{2} < \mathrm{Re}(z) < \frac{1}{2}$ and $|z| > 1$.

Proposition 1.2. *The set F defined above is a fundamental domain for the action of $\Gamma(1)$ on $\mathcal{H}$.*

Proof. For a fixed z, since $|cz+d|^2$ is a positive-definite quadratic form it takes a minimum value, and so $\operatorname{Im}(g\cdot z) = |cz+d|^{-2}\operatorname{Im}(z)$ takes a maximum value for $g \in \Gamma(1)$; let g be a minimizing element. By acting by $\begin{pmatrix} 1 & n \\ & 1 \end{pmatrix}$, we can make $\gamma \cdot z + n = \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix} \cdot \gamma \cdot z$ have real part between $-\frac{1}{2}$ and $\frac{1}{2}$, and the value of the imaginary part does not change, so we can assume that $\gamma \cdot z$ has the real part as described. If it has absolute value less than 1, then acting by $\begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$ would give it larger imaginary part, and the imaginary part is already maximal; this shows that F satisfies one half of the definition of a fundamental domain.

For the other half, suppose that $z \in F$ and $g \cdot z \in F$. Write $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $c = 0$, then since the determinant is 1 and we work up to sign we can assume $g = \begin{pmatrix} 1 & b \\ & 1 \end{pmatrix}$ for some integer b, which will take z out of the fundamental domain unless $b = 0$, in which case it is the identity, so in this case we're done; therefore assume $c \neq 0$. The minimal imaginary part of an element of $\overline{F}$ is at the corners $\pm\frac{1}{2} + \frac{\sqrt{3}}{2}i$, which are not included in F itself, so if z and $g \cdot z$ are both in F then

$$\frac{\sqrt{3}}{2} < \operatorname{Im}(g\cdot z) = \frac{\operatorname{Im}(z)}{|cz+d|^2} \leq \frac{1}{c^2 \operatorname{Im}(y)} < \frac{2}{c^2\sqrt{3}}$$

and so $c^2 \leq \frac{4}{3}$. Since c is an integer, the only possibilities are $c = \pm 1$, which by scaling by -1 we can assume is $c = 1$. Since $ad - bc = ad - b = 1$, we have the decomposition

$$g = \begin{pmatrix} a & b \\ 1 & d \end{pmatrix} = \begin{pmatrix} 1 & a \\ & 1 \end{pmatrix} \begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \begin{pmatrix} 1 & d \\ & 1 \end{pmatrix}.$$

Therefore

$$g \cdot z = -\frac{1}{z+d} + a.$$

Since $|\operatorname{Re}(z)| < \frac{1}{2}$, for any integer d we have $|\operatorname{Re}(z) + d| \geq |\operatorname{Re}(z)|$ and so $|z+d| \geq |z| > 1$; the same argument applies to $g \cdot z$ since it is also in F, so in particular $|(g\cdot z) - a| > 1$. But then $\frac{1}{|z+d|} = |g \cdot z - a| > 1$ and so $|z + d| < 1$, a contradiction, so $c \neq 0$ is impossible and we are left in the situation above. □

We can describe $\mathrm{SL}_2(\mathbb{Z})$ explicitly be generators: it is generated by $T = \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$ and $S = \begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$. To see this, we use the fact that F is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$: pick $z \in \mathcal{H}$. If it has real part greater than $\frac{1}{2}$ or less than or equal to $-\frac{1}{2}$, we can adjust it by powers of T until it has real part in the correct range. If it now has absolute value less than 1, applying S sends z to $-\frac{1}{z}$, with absolute value $\frac{1}{|z|} > 1$. This may change the real part. However, it now has imaginary part $\operatorname{Im}(-1/z) = \frac{\operatorname{Im}(z)}{|z|^2} > \operatorname{Im}(z)$, so once we adjust it back to the right real range by powers of T it will have greater imaginary part, and if the absolute value is still less than 1 we can repeat the process. One might worry that this could take infinitely many steps, but since each $g \in \Gamma(1)$ carries F to a different fundamental domain, which collectively cover $\mathcal{H}$, given any z we can find a path of adjacent fundamental domains going to F; the adjacent domains to F are given by the images under S, T, and T^{-1}, so by homogeneity the same is true for any domain and so moving through this chain corresponds to multiplication by these elements. One can use similar methods to find generators for other discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$.

As mentioned before, it is natural to view $\mathcal{H}$ with the fractional linear action as living inside $\mathbb{P}^1(\mathbb{C})$ with a similar action, where its boundary is $\mathbb{P}^1(\mathbb{R})$, including a point at infinity. For a discrete subgroup Γ of $\mathrm{SL}_2(\mathbb{R})$, the cusps are the points at which the closure of the fundamental domain intersects this boundary; thus for $\Gamma(1)$ there is only one cusp, at the point at infinity. For other congruence subgroups there may also be intersections along the real line, which are treated the same way.

Translating from the fundamental domain to the language of orbits, $\mathrm{SL}_2(\mathbb{Z})$ acts transitively on the rational points $\mathbb{P}^1(\mathbb{Q})$, so any finite index subgroup Γ has finitely many orbits; these orbits are the cusps of Γ, and correspond to the cusps in the above sense. This definition can be generalized to arbitrary discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$ acting discontinuously on $\mathcal{H}$.

For a congruence subgroup Γ, since it acts discontinuously we can form the quotient $\Gamma\backslash\mathcal{H}$. Upon adding one point for each cusp of Γ, we will obtain a compact Riemann surface; we do this by adding a point at infinity and at every rational number, topologized by open sets given by a ball tangent to those points (i.e. $|x| > C$ for some C for the point at infinity). We call this completed upper half-plane $\mathcal{H}^*$. Then we can equip $\Gamma\backslash\mathcal{H}^*$ with the quotient topology; this is a manifold.

By specifying charts at each point, we can even give it a complex structure. For most points, we can just take a neighborhood in $\mathcal{H}$. The difficult points are the elliptic points, i.e. points of $\mathcal{H}$ with a nontrivial stabilizer in Γ (after factoring through its quotient by the trivial scalar action).

For $\Gamma = \Gamma(1)$, the elliptic points are (the orbits of) i and $\rho = e^{2\pi i/3}$, with stabilizers of order 2 and 3 respectively, generated by $\begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$ and $\begin{pmatrix} & -1 \\ 1 & 1 \end{pmatrix}$, since $-\frac{1}{i} = i$ and $-\frac{1}{\rho+1} = \rho$.

If a is an elliptic point, we use the Cayley transform

$$z \mapsto \frac{z-a}{z-\bar{a}}.$$

This carries $\mathcal{H}$ to the unit disk and sends a to 0, which gives a chart about a.

Finally, at the cusps, by acting by an element g of $\Gamma(1)$ we can send any cusp to the point at infinity, which is stabilized by the group generated by T^n for some integer n. Thus $z \mapsto e^{2\pi i(g\cdot z)/n}$ sends a neighborhood of a homeomorphically to a neighborhood of the identity, giving a chart near a. All of these charts live naturally in the complex plane and so give $\Gamma\backslash\mathcal{H}^*$ a complex structure, making it a compact Riemann surface.

2. Modular forms

It is natural to ask about smooth functions on the manifold $\Gamma\backslash\mathcal{H}^*$. Let's start with the simplest case, $\Gamma = \Gamma(1)$, so we're looking for functions on $\mathcal{H}^*$ which are invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$. This is equivalent to functions on $\mathcal{H}$, which are easier to think about, satisfying the additional requirement of being holomorphic at the cusp ∞. What this means is that for any $\mathrm{SL}_2(\mathbb{Z})$-invariant function f on $\mathcal{H}$, since it is invariant under the action of T in particular it is periodic, $f(z+1) = f(z)$, and so has a Fourier expansion of the form

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

where $q = e^{2\pi i z}$. If a_n vanishes for $n < -N$ for N large enough, we say that f is meromorphic at ∞; if $a_n = 0$ for $n < 0$, we say that it is holomorphic at infinity. We could require a stronger condition, that $a_0 = 0$ as well, or $a_n = 0$ for $n \leq 0$; in this case we say that f is cuspidal at ∞, or a cusp form.

Unfortunately this turns out to be a fairly trivial notion: any such function is constant. This is because it descends to a compact manifold $\Gamma\backslash\mathcal{H}^*$ and so takes a maximum value, but this is impossible for any nonconstant function by the maximum modulus principle.

It is better to look at a looser notion than Γ-invariance: we should allow functions which under the action of Γ are not invariant but transformed in some nice way. For $\Gamma = \Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$ and an integer k, we say that a function f on $\mathcal{H}$ transforms with weight k if for ever $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1)$ we have

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

Since for $g = T$ we have $cz + d = 1$, such functions are still periodic with period 1 and so still have a Fourier expansion, so we can form the same condition as above; if f transforms with weight k under $\mathrm{SL}_2(\mathbb{Z})$ and is holomorphic at ∞, we call it a modular form of weight k for $\mathrm{SL}_2(\mathbb{Z})$. We write $M_k(\Gamma(1))$ for the space of such functions. If f is further cuspidal at ∞, we call it a cusp form of weight k, and write the space of such things as $S_k(\Gamma(1)) \subset M_k(\Gamma(1))$. Thus our previous notion corresponds to $k = 0$, and so we see that the space of modular forms of weight 0 is one-dimensional, given by constant functions, and the space of cusp forms of weight k is trivial.

To do anything with these spaces, we would first like to know that they are in fact finite-dimensional for all k.

Proposition 2.1. *For every integer k, the spaces $M_k(\Gamma(1))$ and $S_k(\Gamma(1))$ are finite-dimensional.*

Proof. First, note that the claim for $S_k(\Gamma(1))$ follows from that for $M_k(\Gamma(1))$ since it is a subspace, so we focus on general modular forms. Write X for the compactification of the quotient $\Gamma(1)\backslash\mathcal{H}$ described above. If $M_k(\Gamma(1))$ is trivial, we are done, so we can assume it has some nonzero element f_0. For any other function $f \in M_k(\Gamma(1))$, the function f/f_0 is a $\Gamma(1)$-invariant meromorphic function, with poles at the zeros of f_0 of orders at most those of the zeros. We now apply a lemma (a weaker version of Riemann-Roch).

Lemma 2.2. *The space V of meromorphic functions on a compact Riemann surface holomorphic away from a finite set of points $p_1, \ldots, p_m$ and bounded in order by a set of positive integers $r_1, \ldots, r_m$ has dimension at most $r_1 + \cdots + r_m + 1$.*

Proof. Choose such a meromorphic function $\phi \in V$; we can associate to it around each p_j its Laurent expansion $a_{j,-r_j} z^{-r_j} + a_{j,-r_j+1} z^{-r_j+1} + \cdots$ in a local coordinate z, and keep track of the negative coordinates $a_{j,-r_j+s}$ for $0 \le s < r_j$; taking all the j together gives $r = r_1 + \cdots + r_m$ coefficients, which we bundle together as a vector $A(\phi)$. If we take $N > r$ linearly independent such functions $\phi_1, \ldots, \phi_N$ in V, then the $A(\phi_j)$ cannot be linearly independent, so we can find constants c_j not all 0 such that $\sum_j c_j A(\phi_j) = 0$; since A is linear in ϕ, this is the vector of negative-indexed Laurent coefficients of $\sum_j c_j \phi_j$, so in particular $\sum_j c_j \phi_j$ has no poles. But then by the maximum modulus principle it is constant, so any subspace of V of dimension $N > r$ contains a constant function, which is nonzero since the ϕ_j are linearly independent in V. Since the space of constant functions is one-dimensional, it is only possible that any subspace of V of dimension at least $r+1$ contains it if there is only one such subspace, V itself, so $\dim V \le r+1$ as desired. □

Applying this lemma to X with poles bounded by the orders of the zeros of f_0 (counted with multiplicity of the stabilizer), functions f/f_0 are in the vector space V for $f \in M_k(\Gamma(1))$ and so $\dim M_k(\Gamma(1)) \le \dim V$ which is finite by the lemma. □

This gives us a finiteness result: the space of modular forms is not infinite-dimensional (at least when graded by weight). We'd like to have a result in the other direction: are there actually any nontrivial modular forms at all?

The answer is yes. For even integers $k \ge 4$, we can define

$$E_k(z) = \frac{1}{2} \sum_{(m,n) \in \mathbb{Z}^2 \setminus \{(0,0)\}} (mz+n)^{-k}.$$

This is absolutely convergent, and some series rearrangements show that its Fourier expansion is given by

$$E_k(z) = \zeta(k) + \frac{(2\pi)^k (-1)^{k/2}}{(k-1)!} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n,$$

where $q = e^{2\pi i z}$ and $\sigma_r(n) = \sum_{d|n} d^r$. Sometimes we normalize this by dividing by $\zeta(k)$; we call the result $G_k(z)$, and note that all of the Fourier coefficients are rational in that case.

We will often restrict to even k. In that case we've shown that $M_k(\Gamma(1))$ has dimension at least 1 for $k \ge 4$. Indeed, we've displayed an element $G_k(z)$ with constant term equal to 1, so $\dim M_k(\Gamma(1)) - \dim S_k(\Gamma(1)) \ge 1$; in fact this difference is equal to 1, since given another $f \in M_k(\Gamma(1))$ with constant term 1 we have $f - G_k$ a cusp form, so the space orthogonal to cusp forms has dimension at most 1 and thus equal to 1 when $k \ge 4$.

This leaves the question of whether there are any cusp forms. The answer is again yes, at least in suitable weights: modular forms form a graded ring $M_*(\Gamma(1))$, and one can compute for example

$$G_4(z) = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n) q^n, \qquad G_6(z) = 1 - 504 \sum_{n=1}^{\infty} \sigma_5(n) q^n,$$

so we can combine these to find a nontrivial cusp form of weight 12:

$$\frac{G_4(z)^3 - G_6(z)^2}{1728} = q - 24q^2 + 252q^3 - 1472q^4 + 4830q^5 + \cdots .$$

By manipulating Jacobi's triple product formula, we can obtain another cusp form of weight 12, given by the formula

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

This is a convergent product of nonzero functions on $\mathcal{H}$ and so is nonzero.

Proposition 2.3. *The space $S_{12}(\Gamma(1))$ is one-dimensional, spanned by Δ.*

In particular it follows that Δ is equal to $\frac{G_4^3-G_6^2}{1728}$ up to a scalar, and inspecting Fourier coefficients reveals that this scalar is 1.

Proof. Let f be a cusp form of weight 12. Since Δ is nonvanishing on $\mathcal{H}$, f/Δ is holomorphic on $\mathcal{H}$; at the cusp, Δ has a zero of order 1, and f also has a zero of some positive order by virtue of being a cusp form so it is holomorphic on all of $\mathcal{H}^*$, and it is weight 0 and therefore constant by the arguments above. □

For general discrete subgroups, one can compute the dimensions of the spaces of modular forms via the Riemann-Roch formula and the Selberg trace formula. For the case of $\Gamma(1)$, we can compute the dimensions in general using simpler tools.

Proposition 2.4. *Let k be an even nonnegative integer, and write $k = 12j+r$ for $0 \le r \le 10$ (since r is also even). Then $\dim M_{12j+r}(\Gamma(1))$ is $j+1$ if $r \in \{0,4,6,8,10\}$ and j if $r=2$.*

We can compute the dimensions of the cuspidal spaces from here as one less than the dimension of the corresponding space of modular forms (when it is nonzero).

Proof. First, suppose that $k < 12$, i.e. $j = 0$. We've already seen that $\dim M_0(\Gamma(1)) = 1$; in the cases $k \in \{4,6,8,10\}$, we know that the dimension is at least 1, so it remains to show it is at most 1, i.e. there are no nonzero cusp forms. Let $f \in M_k(\Gamma(1))$ be a cusp form. Then $E_{12-k}f/\Delta$ is a modular form of weight 0 for $k \le 8$, and so constant; in particular if f is nontrivial then E_{12-k} has no zeros. The same goes for $E_{(12-k)m}f^m/\Delta^m$, so $E_{(12-k)m}$ has no zeros and up to a scalar $f^m = \Delta^m/E_{(12-k)m}$ is a holomorphic function (this now works for $k = 10$ as well). To get something of weight 0, we multiply by some Δ^r such that $12(r+m) = (12-k)m$, so $r = -mk/12$; then $\Delta^{m+r}/E_{(12-k)m}$ is holomorphic of weight 0 and so constant; but it is a multiple of f^m and so has a zero at infinity, so it is everywhere 0, a contradiction since Δ is nonzero.

Next, we want to show that $M_2(\Gamma(1))$ is trivial. Suppose that f is a nonzero modular form of weight 2. Then fE_4 is a nonzero modular form of weight 6, so it is a nonzero scalar multiple of E_6; one can check that $E_4(\rho) = 0$, where $\rho = e^{2\pi i/3}$ as above, so $E_6(\rho) = 0$ and therefore $\Delta(\rho) = 0$, but Δ is nonvanishing on $\mathcal{H}$, a contradiction.

Finally, suppose $k \ge 12$. Then multiplication by Δ gives an injection $M_{k-12}(\Gamma(1)) \to S_k(\Gamma(1))$, and if $f \in S_k(\Gamma(1))$ then f/Δ is holomorphic and so lies in $M_{k-12}(\Gamma(1))$, i.e. this map is also surjective and so $\dim S_k(\Gamma(1)) = \dim M_{k-12}(\Gamma(1))$. The formula follows from the fact that the cusp forms are a codimension 1 subspace. □

Corollary 2.5. *The ring of modular forms is generated by G_4 and G_6.*

Proof. Let R be the subring generated by G_4 and G_6. Since M_4, M_6, M_8, and M_10 are one-dimensional, their elements are all scalar multiples of G_4, G_6, G_4^2, and G_4G_6 respectively, so $R_k = M_k$ for $k \leq 10$. We know from Proposition 2.3 that Δ is generated by G_4 and G_6, so if f is a cusp form of weight k then f/Δ is a modular form of weight $k-12$, which we can assume by induction is in R, so f is also in R; and up to a scalar there is only one non-cuspidal form, which is given by $E_4^a E_6^b$ for a, b such that $4a + 6b = k$, possible for every even $k \geq 4$ since $\gcd(4, 6) = 2$. □

Suppose that f and g are cusp forms of weight k. Then $f(z)\overline{g(z)}\,\mathrm{Im}(z)^k$ is invariant under the action of $\Gamma(1)$: replacing z by γz, where $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, gives

$$f(\gamma z)g(\gamma z)\,\mathrm{Im}(\gamma z)^k = (cz+d)^k\overline{(cz+d)^k}|cz+d|^{-2k} f(z)\overline{g(z)}\,\mathrm{Im}(z)^k = f(z)\overline{g(z)}\Im(z)^k.$$

Therefore it descends to the quotient and so it makes sense to define

$$\langle f, g\rangle = \int_{\Gamma(1)\backslash\mathcal{H}} f(z)\overline{g(z)}\,\mathrm{Im}(z)^k\,Dz$$

where Dz is the invariant measure under the action of $\mathrm{SL}_2(\mathbb{R})$, sometimes written $\frac{dx\,dy}{y^2}$ where $z = x+iy$. This is the Petersson inner product on $S_k(\Gamma(1))$; the integral is rapidly convergent by the properties of cusp forms, and gives a positive-definite Hermitian inner product.

To any modular form

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

we can associate an L-function

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

We would like it to converge for $\mathrm{Re}(s)$ sufficiently large. To prove this, we use the following estimate.

Proposition 2.6. *If f is a cusp form of weight k, then its Fourier coefficients a_n satisfy $|a_n| \leq Cn^{k/2}$ for some constant C independent of n.*

This is the "trivial estimate," due to Hardy and Hecke; a better bound is $Cn^{(k-1)/2+\epsilon}$ for any $\epsilon > 0$, conjectured by Ramanujan (for $f = \Delta$) and Petersson (in general) and finally proven by Deligne, using the Weil conjectures.

Proof. After taking absolute values, $|f(z)\,\mathrm{Im}(z)^{k/2}|$ is $\Gamma(1)$-invariant as above, and since f is cuspidal it is bounded over all $z \in \mathcal{H}$. We can obtain the Fourier coefficients by integrating: for any fixed $y > 0$,

$$\int_0^1 f(x+iy)e^{-2\pi inx}\,dx = \sum_{m=1}^{\infty} a_m \int_0^1 e^{2\pi im(x+iy)}e^{-2\pi inx}\,dx = a_n e^{-2\pi ny},$$

so

$$|a_n| \leq \int_0^1 |f(x+iy)|\,dx \leq Ce^{2\pi ny}y^{-k/2}$$

since $f(z)\,\mathrm{Im}(z)^{k/2}$ is bounded. Taking $y = \frac{1}{n}$ gives the claimed bound. □

Note that this bound is only valid when f is a cusp form. However, another bound of a similar type is possible for Eisenstein series: the nth Fourier coefficient of E_k is $O(\sigma_{k-1}(n)) = O(n^{k-1}\log n)$. In either case the L-function is convergent for $\mathrm{Re}(s)$ sufficiently large.

Proposition 2.7. *Let f be a modular form of weight k. Then $L(s,f)$ has meromorphic continuation to the complex plane; if*

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

then

$$\Lambda(s,f) = (-1)^{k/2}\Lambda(k-s,f),$$

and if f is a cusp form $\Lambda(s,f)$ is holomorphic; otherwise it has simple poles at $s=0$ and $s=k$.

Proof. First, assume that f is a cusp form. Let $z = iy$ for y real. The action of $S = \begin{pmatrix} & -1 \\ 1 & \end{pmatrix}$ is by

$$f(-1/(iy)) = f(i/y) = z^k f(iy) = (-1)^{k/2} y^k f(iy),$$

so as $y \to \infty$ since f is cuspidal the right-hand side goes to 0 rapidly, so so does the left-hand side, i.e. $f(iy)$ goes rapidly to 0 as $y \to 0$ as well. Therefore the integral

$$\int_0^\infty f(iy)y^s\,\frac{dy}{y}$$

converges and gives an analytic function of s. Substituting the Fourier expansion of f, for $\mathrm{Re}(s)$ sufficiently large we can exchange the summation and integration to get

$$\sum_{n=1}^\infty a_n \int_0^\infty e^{-2\pi n y} y^s \frac{dy}{y},$$

and substituting $u = 2\pi n y$ this is

$$\sum_{n=1}^\infty a_n (2\pi n)^{-s} \int_0^\infty e^{-u} u^s\,\frac{du}{u} = (2\pi)^{-s}\Gamma(s)L(s,f) = \Lambda(s,f).$$

On the other hand, using the modular properties of f above and substituting $1/y$ for y gives for the same integral

$$\int_0^\infty f(i/y)y^{-s}\,\frac{dy}{y} = \int_0^\infty (-1)^{k/2}y^{k-s}f(iy)\,\frac{dy}{y} = (-1)^{k/2}\Lambda(s,f)$$

by the above, so the functional equation follows. By standard L-function methods Proposition 2.6 shows that $L(s,f)$, and therefore $\Lambda(s,f)$, converges for $\mathrm{Re}(s) > 1 + \frac{k}{2}$, and so $\Lambda(k-s,f)$ is holomorphic for $s < \frac{k}{2} - 1$ and so the functional equation gives an extension of $\Lambda(s,f)$ to a holomorphic function on the whole plane.

Since $L(s,-)$ and $\Lambda(s,-)$ are linear in f, to conclude the result for all modular forms it suffices to prove it for Eisenstein series. In this case $f(iy)$ does not go to 0 as $y \to \infty$ or

$y \to 0$, but is only bounded, and the integral does not converge. Instead, we compute the L-function directly: we know that for $n \geq 1$ the nth Fourier coefficient is $\frac{(2\pi)^k(-1)^{k/2}}{(k-1)!}\sigma_{k-1}(n)$, so

$$L(s, E_k) = \frac{(2\pi)^k(-1)^{k/2}}{(k-1)!}\sum_{n=1}^{\infty}\sigma_{k-1}(n)n^{-s}.$$

By Dirichlet convolution it is not hard to see that

$$\sum_{n=1}^{\infty}\sigma_{k-1}(n)n^{-s} = \zeta(s)\zeta(s-k+1).$$

Thus

$$\Lambda(s, E_k) = \frac{(2\pi)^{k-s}(-1)^{k/2}}{(k-1)!}\Gamma(s)\zeta(s)\zeta(s-k+1).$$

Recall that the completed Riemann zeta function is

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

so

$$\Lambda(s, E_k) = \frac{2^{k-s}\pi^{(k+1)/2}\Gamma(s)}{\Gamma(\frac{s}{2})\Gamma(\frac{s-k+1}{2})}\Lambda(s)\Lambda(s-k+1).$$

Therefore

$$\Lambda(k-s, E_k) = \frac{2^{s}\pi^{(k+1)/2}\Gamma(k-s)}{\Gamma(\frac{k-s}{2})\Gamma(\frac{1-s}{2})}\Lambda(k-s)\Lambda(1-s).$$

Since $\Lambda(s)$ satisfies the functional equation $\Lambda(1-s) = \Lambda(s)$, this is

$$\frac{2^{s}\pi^{(k+1)/2}\Gamma(k-s)}{\Gamma(\frac{k-s}{2})\Gamma(\frac{1-s}{2})}\Lambda(s-k+1)\Lambda(s),$$

so it suffices to check that

$$\frac{2^{s}\Gamma(k-s)}{\Gamma(\frac{k-s}{2})\Gamma(\frac{1-s}{2})} = \frac{2^{k-s}\Gamma(s)}{\Gamma(\frac{s}{2})\Gamma(\frac{s-k+1}{2})},$$

which we can verify using the reflection formula. □

Thus these L-functions have many of the properties of arithmetic L-functions, such as Dirichlet L-functions; this is not a coincidence. Another property we would like to see is an Euler product expansion; to get one, we need the theory of Hecke operators.

3. Hecke operators

To define Hecke operators, it is convenient to enlarge the groups we are considering. We can extend the action on $\mathcal{H}$ from $\mathrm{SL}_2(\mathbb{R})$ to the subgroup $\mathrm{GL}_2(\mathbb{R})^+ \subset \mathrm{GL}_2(\mathbb{R})$ given by the connected component with positive determinant; this still preserves $\mathcal{H}$, though some formulas become messier. Of course, this still factors through $\mathrm{PGL}_2(\mathbb{R})^+ \simeq \mathrm{PSL}_2(\mathbb{R})$, so it is essentially the same action on $\mathcal{H}$. However, the induced action on functions is more general: for f a holomorphic function on $\mathcal{H}$, $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})^+$, and k a fixed weight, which we now allow to be either even or odd, define a new function $f|\gamma$ by

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

One can check that this gives a genuine right action of $\mathrm{GL}_2(\mathbb{R})^+$ on the space of holomorphic functions on $\mathcal{H}$. If k is even, then scalar matrices $\begin{pmatrix} \lambda & \\ & \lambda \end{pmatrix}$ act trivially as expected; if k is odd, though, they act by $\mathrm{sign}(\lambda)$. Observe that f transforms with weight k under a group $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$ if and only if Γ fixes f under this action.

We can now generalize the definition of modular forms. Let Γ be a subgroup of $\mathrm{SL}_2(\mathbb{R})$ acting discontinuously on $\mathcal{H}$ such that $\Gamma\backslash\mathcal{H}$ has finite volume, such as a congruence subgroup of $\Gamma(1)$. Then a holomorphic function f on $\mathcal{H}$ is a modular form of weight k for Γ if Γ fixes f under the above action and f is holomorphic at the cusps of $\Gamma\backslash\mathcal{H}$. To make this precise, we can use the same notion as for $\mathrm{SL}_2(\mathbb{Z})$ for the cusp at ∞ (except that now the period may be some integer t greater than 1); for the other cusps, let $g \in \Gamma$ be an element taking the cusp to ∞. Then $f|g^{-1}$ has a Fourier expansion $\sum_n a_{g,n}e^{2\pi inz/t}$ and we can make the usual definitions.

Note that if $-1 \in \Gamma$, then there are no nonzero modular forms of odd weight for Γ:

$$f(z) = (f|-1)(z) = (-1)^{-k}f(z) = -f(z)$$

for all z, so $f = 0$. Since we want to focus primarily on $\mathrm{SL}_2(\mathbb{Z})$, which contains -1, we'll assume that k is even.

Proposition 3.1. *For every $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$, the double coset $\Gamma(1)\alpha\Gamma(1)$ is a finite union of right cosets*

$$\Gamma(1)\alpha\Gamma(1) = \bigcup_{i=1}^{N} \Gamma(1)\alpha_i$$

for $\alpha_i \in \mathrm{GL}_2(\mathbb{Q})^+$, with $N = [\Gamma(1) : \alpha^{-1}\Gamma(1)\alpha \cap \Gamma(1)]$.

Proof. It is easy to check that conjugates of congruence subgroups are congruence subgroups, so $\alpha^{-1}\Gamma(1)\alpha \cap \Gamma(1)$ is a congruence subgroup of $\Gamma(1)$ and so has finite index in it. Right multiplication by α^{-1} gives a bijection of $\mathrm{GL}_2(\mathbb{Q})^+$ with itself, and applying this to the set of cosets $\Gamma(1)\backslash\Gamma(1)\alpha\Gamma(1)$ gives a bijection to $\Gamma(1)\backslash\Gamma(1)(\alpha\Gamma(1)\alpha^{-1})$. By the second isomorphism theorem, this is isomorphic to $(\Gamma(1) \cap \alpha\Gamma(1)\alpha^{-1})\backslash\alpha\Gamma(1)\alpha^{-1}$. Multiplying on the left by α^{-1} and on the right by α is a further bijection carrying this to $(\alpha^{-1}\Gamma(1)\alpha \cap \Gamma(1))\backslash\Gamma(1)$, and so taking cardinalities all around we conclude that the number of right cosets in the decomposition of the proposition is equal to the index claimed. $\square$

Given any $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$, by Proposition 3.1 we can find $\alpha_1, \ldots, \alpha_N$ with N as described giving a decomposition of $\Gamma(1)\alpha\Gamma(1)$ into right cosets. We define an operator T_α on $M_k(\Gamma(1))$ by

$$f|T_\alpha = \sum_{i=1}^{N} f|\alpha_i.$$

This appears to depend on the choice of the representatives α_i, but each is well-defined up to an element of $\Gamma(1)$, and since f is modular it is fixed by the action of $\Gamma(1)$. Moreover the action of $\gamma \in \Gamma(1)$ on the right does not change $\Gamma(1)\alpha\Gamma(1)$ and so only permutes the α_i, so $f|T_\alpha$ is still modular. The subspace of cusp forms is invariant under this action.

To see how these operators compose, choose another $\beta \in \mathrm{GL}_2(\mathbb{R})^+$, with corresponding β_i. Then

$$f|T_\alpha T_\beta = \sum_{i,j} f|\alpha_i\beta_j = \sum_{\sigma\in\Gamma(1)\backslash \mathrm{GL}_2(\mathbb{Q})^+} m(\alpha,\beta;\sigma) f|\sigma$$

where $m(\alpha,\beta;\sigma)$ is the number of pairs of indices (i,j) such that $\sigma \in \Gamma(1)\alpha_i\beta_j$. Multiplying σ by an element of $\Gamma(1)$ on the right does not change this number, so we can write this as

$$f|T_\alpha T_\beta = \sum_{\sigma\in\Gamma(1)\backslash \mathrm{GL}_2(\mathbb{Q})^+/\Gamma(1)} m(\alpha,\beta;\sigma)|T_\sigma.$$

This motivates the following definition: let R be the free abelian group generated by T_α for α representatives for $\Gamma(1)\backslash \mathrm{GL}_2(\mathbb{Q})^+/\Gamma(1)$. We equip R with a multiplication making it into a ring by

$$T_\alpha \cdot T_\beta = \sum_{\sigma\in\Gamma(1)\backslash \mathrm{GL}_2(\mathbb{Q})^+/\Gamma(1)} m(\alpha,\beta;\sigma)T_\sigma,$$

i.e. such that $f|T_\alpha \cdot T_\beta = (f|T_\alpha)|T_\beta$. Thus R acts on $M_k(\Gamma(1))$, and is called the Hecke algebra.

Proposition 3.2. *A complete set of representatives for* $\Gamma(1)\backslash \mathrm{GL}_2(\mathbb{Q})^+/\Gamma(1)$ *is given by the diagonal matrices* $\begin{pmatrix} d_1 & \\ & d_2 \end{pmatrix}$ *with* d_1 *and* d_2 *rational numbers such that* d_1/d_2 *is a positive integer.*

Proof. For each $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$, choose an integer N large enough that $N\alpha$ has integer coefficients. We need to apply an algebraic fact, the elementary divisor theorem, which states that if R is a principle ideal domain, Λ_1 is a free R-module of rank n, and $\Lambda_2 \subset \Lambda_1$ a free R-submodule also of rank n, then there is a basis $\xi_1,\dots,\xi_n$ of Λ_1 and nonzero elements $D_1,\dots,D_n$ of R such that $D_{i+1}|D_i$ for $1 \le i < n$ and $D_1\xi_1,\dots,D_n\xi_n$ is a basis of Λ_2. In our case, if $R = \mathbb{Z}$, $\Lambda_1 = \mathbb{Z}^2$, and Λ_2 is the sublattice spanned by the rows of $N\alpha$, we see that there is a basis ξ_1, ξ_2 of $\mathbb{Z}^2$ and integers D_1, D_2 such that $D_2|D_1$ and $D_1\xi_1, D_2\xi_2$ is a basis of Λ_2; by flipping the signs as necessary, we can assume that D_1 and D_2 are positive and the matrix ξ whose rows are ξ_1 and ξ_2 has determinant 1. This is therefore an element of $\mathrm{SL}_2(\mathbb{Z})$, and $\begin{pmatrix} D_1 & \\ & D_2 \end{pmatrix}\xi$ gives a matrix whose rows span the same lattice as $N\alpha$ and so is a change of basis away, i.e. there is some $\gamma \in \mathrm{GL}_2(\mathbb{Z})$ such that $\gamma N\alpha = \begin{pmatrix} D_1 & \\ & D_2 \end{pmatrix}\xi$. Taking determinants, we find that $\det\gamma > 0$ and so $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, and so

$$\alpha = \gamma^{-1}\begin{pmatrix} D_1/N & \\ & D_2/N \end{pmatrix}\xi$$

gives the desired decomposition, with $(D_1/N)/(D_2/N) = D_1/D_2$ an integer since D_2 divides D_1.

This process shows that not only can we find a diagonal matrix in the coset of α, but it is actually unique. In particular, we can recover D_1 as the greatest common divisor of entries of $N\alpha$, and D_2 as $\frac{N^2}{D_1}\det\alpha$. □

Proposition 3.3. *For any* $\alpha \in \mathrm{GL}_2(\mathbb{Q})^+$*, we have* $\Gamma(1)\alpha\Gamma(1) = \Gamma(1)\alpha^{\mathsf{T}}\Gamma(1)$.

Proof. By Proposition 3.2, we can find a representative that is diagonal and therefore symmetric, so its transpose gives the same coset, which is stable under transposition. □

Proposition 3.4. *We can choose the representatives* α_i *for the decomposition from Proposition 3.1 such that*

$$\Gamma(1)\alpha\Gamma(1) = \bigcup_{i=1}^{N} \alpha_i\Gamma(1)$$

also holds.

Proof. By Proposition 3.3, the double coset is stable under transposition, so

$$\Gamma(1)\alpha\Gamma(1) = \bigcup_{i=1}^{n} \alpha_i^{\mathsf{T}}\Gamma(1).$$

Since α_i and α_i^{T} generate the same double coset, there is some $g_i \in \Gamma(1)$ such that $\alpha_i^{\mathsf{T}} g_i \alpha^{-1} \in \Gamma(1)$. Then $\alpha_i^{\mathsf{T}} g_i$ is equal to $h_i\alpha_i$ for some $h \in \Gamma(1)$, and so they give a set of representatives with the desired property:

$$\bigcup_{i=1}^{N} \alpha_i^{\mathsf{T}} g_i \Gamma(1) = \bigcup_{i=1}^{N} \Gamma(1)\alpha_i = \Gamma(1)\alpha\Gamma(1),$$

and work for the original decomposition as well:

$$\bigcup_{i=1}^{N} \Gamma(1)\alpha_i^{\mathsf{T}} g_i = \bigcup_{i=1}^{N} \Gamma(1)h_i\alpha_i = \bigcup_{i=1}^{N} \Gamma(1)\alpha_i = \Gamma(1)\alpha\Gamma(1).$$

□

Theorem 3.5. *The Hecke algebra* R *is commutative.*

The idea is that transposition induces an antiautomorphism of R, which by Proposition 3.3 is really the identity, so in order for this to be an antiautomorphism R must be commutative.

Proof. We proceed by showing that the structure constants $m(\alpha, \beta; \sigma)$ are actually symmetric in α and β. Write $\Gamma(1)\sigma\Gamma(1) = \bigcup_l \Gamma(1)\sigma_l$. Then $\sigma \in \Gamma(1)\alpha_i\beta_j$ if and only if some σ_l is in $\Gamma(1)\alpha_i\beta_j$, and so

$$\sum_l m(\alpha, \beta; \sigma_l) = \sum_l m(\alpha, \beta; \sigma) = m(\alpha, \beta; \sigma) \cdot |\Gamma(1)\backslash\Gamma(1)\sigma\Gamma(1)|.$$

Choose representatives α_i and β_i as in Proposition 3.4, so that

$$\Gamma(1)\alpha\Gamma(1) = \bigcup_i \Gamma(1)\alpha_i = \bigcup_i \alpha_i\Gamma(1) = \bigcup_i \Gamma(1)\alpha_i^{\mathsf{T}}$$

and

$$\Gamma(1)\beta\Gamma(1) = \bigcup_i \Gamma(1)\beta_i = \bigcup_i \beta_i\Gamma(1) = \bigcup_i \Gamma(1)\beta_i^{\mathsf{T}}.$$

Then

$$\begin{aligned} m(\alpha,\beta;\sigma)\cdot|\Gamma(1)\backslash\Gamma(1)\sigma\Gamma(1)| &= |\{(i,j)|\sigma\in\Gamma(1)\alpha_i\beta_j\Gamma(1)\}| \\ &= |\{(i,j)|\sigma\in\Gamma(1)\alpha_i^{\mathsf{T}}\beta_j^{\mathsf{T}}\Gamma(1)\}| \\ |\{(i,j)|\sigma^{\mathsf{T}}\in\Gamma(1)\beta_j\alpha_i\Gamma(1)\}| \\ &= m(\beta,\alpha;\sigma^{\mathsf{T}})\cdot|\Gamma(1)\backslash\Gamma(1)\sigma^{\mathsf{T}}\Gamma(1)|. \end{aligned}$$

We can choose the representative σ such that $\sigma^{\mathsf{T}} = \sigma$, so $m(\alpha,\beta;\sigma) = m(\beta,\alpha;\sigma)$, and so the multiplication in R is commutative. □

This means we can dismiss with the awkward $f|T_\alpha$ notation and instead write $T_\alpha f$.

We extend the Petersson inner product to cusp forms for any congruence subgroup by picking N such that f and g are both modular forms for $\Gamma(N)$ and setting

$$\langle f,g\rangle = \frac{1}{[\Gamma(1):\Gamma(N)]}\int_{\Gamma(N)\backslash\mathcal{H}} f(z)\overline{g(z)}\,\mathrm{Im}(z)^k\,Dz$$

with Dz the invariant measure as before.

Theorem 3.6. *The Hecke operators on $S_k(\Gamma(1))$ are self-adjoint with respect to the Petersson inner product, i.e. $\langle T_\alpha f, g\rangle = \langle f, T_\alpha g\rangle$.*

Proof. It can be checked that by replacing z with $\alpha^{-1}z$ in the integral we get

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

The expression on the left is invariant under replacing α by $\gamma\alpha$ for $\gamma\in\Gamma(1)$, and similarly the expression on the right is invariant under replacing α by $\alpha\gamma$; thus both sides depend only on the coset of α in $\Gamma(1)\backslash\mathrm{GL}_2(\mathbb{Q})^+/\Gamma(1)$, which contains all of the α_i of Proposition 3.4. Therefore

$$\langle T_\alpha f, g\rangle = \sum\langle f|\alpha_i, g\rangle = \langle f|\alpha, g\rangle\cdot|\Gamma(1)\backslash\Gamma(1)\alpha\Gamma(1)| = \left\langle f, g|\alpha^{-1}\right\rangle\cdot|\Gamma(1)\backslash\Gamma(1)\alpha\Gamma(1)|.$$

By Proposition 3.3, this is the same thing as

$$\left\langle f, g|\alpha^{-\mathsf{T}}\right\rangle\cdot|\Gamma(1)\backslash\Gamma(1)\alpha\Gamma(1)|.$$

We can check that $\alpha^{-\mathsf{T}}\det\alpha = S\alpha S^{-1}$, so because scalars act trivially and $S\in\Gamma(1)$ this is

$$\langle f, g|\alpha\rangle\,|\Gamma(1)\backslash\Gamma(1)\alpha\Gamma(1)|,$$

which by the same argument as above is $\langle f, T_\alpha g\rangle$. □

For any commutative algebra of self-adjoint operators on a finite-dimensional vector space we can find a basis of eigenvectors, in this case eigenfunctions in $S_k(\Gamma(1))$, i.e. nonzero cusp forms f which are eigenfunctions for every Hecke operator. In this case we say that f is a Hecke eigenform; since these form a basis we can often restrict attention to them, and for certain purposes they are much better-behaved. In particular, the L-function of a normalized Hecke eigenform has an Euler product expansion.

First, we need to associate Hecke operators to integers. We do this by taking all diagonal matrices $\operatorname{diag}(d_1, d_2)$ with d_1, d_2 integers such that $d_2|d_1$ and $d_1d_2 = n$; by Proposition 3.2 these are the cosets corresponding of the set Δ_n of integer matrices with determinant n. If we sum over the Hecke operators corresponding to all such matrices (of which there are finitely many), we call the result $T(n)$.

One can make these explicit: Δ_n decomposes as the right cosets $\Gamma(1)\begin{pmatrix} a & b \\ & d \end{pmatrix}$ for $a, d > 0$, $ad = n$, and b ranging over any set of representatives of congruence classes modulo d. Thus $T(n)$ can be viewed as the sum of the Hecke operators for these matrices, which act by

$$\left(f\middle|\begin{pmatrix} a & b \\ & d \end{pmatrix}\right)(z) = (ad)^{k/2}d^{-k}f\left(\frac{az+b}{d}\right) = \left(\frac{a}{d}\right)^{k/2} f\left(\frac{az+b}{d}\right).$$

Suppose that f has Fourier expansion $\sum_n a_n q^n$. We'd like to understand the Fourier expansion $\sum_n b_n q^n$ of $T(n)f$. For notational convenience, we write a_n and b_n even for n rational rather than integral, by setting them to be 0 whenever n is not an integer, and write $e(z)$ for $e^{2\pi i z}$. Then

$$\begin{aligned}
(T(n)f)(z) &= \sum_{\substack{a,d>0 \\ ad=n}} \sum_{b=0}^{d-1} \left(\frac{a}{d}\right)^{k/2} f\left(\frac{az+b}{d}\right) \\
&= \sum_{\substack{a,d>0 \\ ad=n}} \sum_{b=0}^{d-1} \left(\frac{a}{d}\right)^{k/2} \sum_{m=1}^{\infty} a_m e\left(\frac{amz}{d}\right) e\left(\frac{mb}{d}\right) \\
&= \sum_{m=1}^{\infty} a_m \sum_{\substack{a,d>0 \\ ad=n}} \left(\frac{a}{d}\right)^{k/2} e\left(\frac{amz}{d}\right) \sum_{b=0}^{d-1} e\left(\frac{mb}{d}\right).
\end{aligned}$$

The innermost sum vanishes unless $d|m$, in which case it is d, so this is

$$\sum_{m=1}^{\infty} a_m \sum_{\substack{a,d>0 \\ ad=n \\ d|m}} \left(\frac{a}{d}\right)^{k/2} de\left(\frac{amz}{d}\right).$$

On the other hand by definition

$$(T(n)f)(z) = \sum_{l=1}^{\infty} b_l e(lz),$$

so comparing the corresponding terms, so $l = \frac{a}{m}d$, we get

$$b_l = \sum_{\substack{a,d>0 \\ ad=n \\ a|l}} \left(\frac{a}{d}\right)^{k/2} d a_{dl/a}.$$

Now suppose that f is a Hecke eigenform; we normalize the eigenvalues by $n^{k/2-1}T(n)f = \lambda(n)f$.

Proposition 3.7. *Let f be a Hecke eigenform with $\lambda(n)$ as above and Fourier coefficients a_n. Then $a_1 \neq 0$, and if we normalize f such that $a_1 = 1$, then $a_n = \lambda(n)$ for every n and $a_{mn} = a_m a_n$ if $\gcd(m,n) = 1$.*

Proof. By the discussion above together with the fact that f is a Hecke eigenform, we have

$$n^{1-k/2}\lambda(n)a_m = \sum_{\substack{a,d>0 \\ ad=n \\ a|m}} \left(\frac{a}{d}\right)^{k/2} d a_{dm/a}.$$

Suppose that $\gcd(m,n) = 1$. Then if $a|m$ and $a|n$ then $a = 1$, so the only term in the sum is $a = 1$ and $d = n$, so it is just $n^{1-k/2}a_{nm}$, i.e.

$$\lambda(n)a_m = a_{nm}.$$

For $m = 1$, we always have $\gcd(m,n) = 1$, and so $\lambda(n)a_1 = a_n$, so a_1 cannot be zero since otherwise f is everywhere zero, and it follows that if $a_1 = 1$ then $\lambda(n) = a_n$ for all n. Therefore we have

$$a_n a_m = a_{nm}$$

for $\gcd(n,m) = 1$. □

Thus it is most natural to normalize Hecke eigenforms such that $a_1 = 1$; we call such an eigenform normalized. In particular $S_k(\Gamma(1))$ has a basis of normalized eigenforms.

Theorem 3.8. *Let f be a normalized Hecke eigenform. Then its L-function has an Euler product*

$$L(s,f) = \sum_{n=1}^{\infty} a_n n^{-s} = \prod_p \frac{1}{1 - a_p p^{-s} + p^{k-1-2s}}.$$

Proof. By the multiplicativity of the coefficients and general L-function theory (essentially the fundamental theorem of arithmetic), we have an expansion

$$L(s,f) = \sum_{n=1}^{\infty} a_n n^{-s} = \prod_p \left(\sum_{r=0}^{\infty} a_{p^r} p^{-rs}\right).$$

We have from the formula for the Fourier expansion of eigenforms

$$p^{1-k/2}\lambda(p)a_{p^r} = \sum_{\substack{a,d>0 \\ ad=p \\ a|p^r}} \left(\frac{a}{d}\right)^{k/2} d a_{dp^r/a}.$$

There are only two possibilities for $ad = p$, namely $a = 1$ and $d = p$ or $a = p$ and $d = 1$, so the right-hand side is

$$p^{1-k/2}a_{p^{r+1}} + p^{k/2}a_{p^{r-1}},$$

and since $\lambda(p) = a_p$ we have

$$a_{p^{r+1}} - a_p a_{p^r} + p^{k-1}a_{p^{r-1}} = 0.$$

One can obtain from this recursion

$$\sum_{r=0}^{\infty} a_{p^r}x^r = \frac{1}{1 - a_p x + p^{k-1}x^2},$$

and setting $x = p^{-s}$ gives the desired identity. □

It is also possible to find a theory of Hecke operators for congruence subgroups, in more or less similar ways.

4. Twisting and a converse theorem

We have seen that the L-function of a normalized eigenform of weight k has analytic continuation to all of $\mathbb{C}$, a functional equation relating its values at s and $k - s$, and an Euler product. There is a sense in which the converse is true.

Proposition 4.1. *Let $a = a_n$ be a sequence of complex numbers with $|a_n| = O(n^K)$ for some sufficiently large real number K, and let*

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

Suppose that $\Lambda(s, a) = (2\pi)^{-s}\Gamma(s)L(s, a)$ has analytic continuation to all of $\mathbb{C}$, is bounded in vertical strips $\sigma_1 \leq \mathrm{Re}(s) \leq \sigma_2$, and satisfies $\Lambda(s, a) = \Lambda(k - s, a)$. Then

$$f(z) = \sum_{n=1}^{\infty} a_n q^n$$

is a cusp form of weight k. If

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

then f is a Hecke eigenform.

The condition on being bounded in vertical strips is a technical one we skipped earlier, but holds for $L(s, f)$.

Since f is defined by a Fourier expansion, it is T-invariant, so the main thing is to study its behavior under the action of S; this is done by applying Mellin inversion (with the help of the Phragmén-Lindelöf principle) to study f from $\Lambda(s, f)$, which is essentially its Mellin transform. The assertion about the Euler product follows from inverting the reasoning in the proof of Theorem 3.8.

However, in general Proposition 4.1 is too specialized to be terribly useful. To get something more powerful, we need to broaden our definition of modular forms again. A useful congruence subgroup is $\Gamma_0(N)$, the preimage of the upper triangular matrices along $\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$, i.e. matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ with $c \equiv 0 \pmod N$. We study a generalization of modular forms for $\Gamma_0(N)$: instead of requiring that $f|\gamma = f$ for $\gamma \in \Gamma_0(N)$, we instead require that $f|\gamma = \chi(d)f$ for $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$, where χ is a Dirichlet character modulo N. We write the space of forms of this type as $M_k(\Gamma_0(N), \chi)$, with cuspidal subspace $S_k(\Gamma_0(N), \chi)$. To get a nonempty space, we require k even if $\psi(-1) = 1$ and k odd if $\psi(-1) = -1$.

Let $w_N = \begin{pmatrix} & -1 \\ N & \end{pmatrix}$. This normalizes $\Gamma_0(N)$, and one can check that if $f \in S_k(\Gamma_0(N), \chi)$ then $f|w_N|\gamma = \overline{\chi(d)} f|w_n$, i.e. w_N maps $S_k(\Gamma_0(N), \chi)$ to $S_k(\Gamma_0(N), \overline{\chi})$. We have

$$f(iy) = i^k N^{-k/2} y^{-k} (f|w_N)(i/(Ny)),$$

and one can use similar arguments to those appearing in the proof of Proposition 2.7 to show that we have a functional equation

$$\Lambda(s, f) = i^k N^{-s+k/2} \Lambda(k - s, f|w_N),$$

giving extensions of both $\Lambda(s, f)$ and $\Lambda(s, f|w_N)$ to $\mathbb{C}$.

More generally, suppose that ψ is another Dirichlet character with some modulus D; assume that ψ is primitive. Write

$$L(s, f, \psi) = \sum_{n=1}^{\infty} \psi(n) a_n n^{-s},$$

with completion $\Lambda(s, f, \psi)$ as usual. One can prove, by more careful work, the functional equation

$$\Lambda(s, f, \psi) = i^k \psi(N) \chi(D) \frac{\tau(\psi)^2}{D} (D^2 N)^{-s+k/2} \Lambda(k - s, f|w_N, \overline{\psi}),$$

where

$$\tau(\psi) = \sum_{a=0}^{D-1} \psi(a) e^{2\pi i a/D}.$$

The main result here is that a similar converse theorem to Proposition 4.1 is possible.

The proof of Proposition 4.1 relies on the fact that $\Gamma(1)$ has fairly simple generators, which we use a functional equation to govern. For more complicated congruence subgroups, the generators are more complicated, and we need multiple functional equations.

Theorem 4.2 (Weil). *Let N be a positive integer and χ be a Dirichlet character modulo N. Let $a = a_n$ and $b = b_n$ be sequences of complex numbers, both bounded by $O(n^K)$ for some* *sufficiently large real number K. For any primitive Dirichlet character ψ modulo D coprime to N, let*

$$L_1(s,\psi) = \sum_{n=1}^{\infty} \psi(n) a_n n^{-s}, \qquad L_2(s,\psi) = \sum_{n=1}^{\infty} \psi(n) b_n n^{-s},$$

with completions Λ_1, Λ_2 as usual. Suppose that for any D either 1 or a prime other than a finite set and every primitive character ψ modulo D we have an analytic continuation of $\Lambda_1(s,\psi)$ and $\Lambda_2(s,\overline{\psi})$ to $\mathbb{C}$ bounded in vertical strips and satisfying the functional equation

$$\Lambda_1(s,\psi) = i^k \psi(N)\chi(D)\frac{\tau(\psi)^2}{D}(D^2N)^{-s+k/2}\Lambda_2(k-s,\overline{\psi}).$$

Then $f(z) = \sum_n a_n q^n$ is a modular form in $M_k(\Gamma_0(N),\psi)$, and $g(z) = \sum_n b_n q^n$ is a modular form in $M_k(\Gamma_0(N),\overline{\psi})$ given by $g = f|w_N$.

The proof uses broadly similar ideas to that of Proposition 4.1, but with much greater care.

5. The associated automorphic representation

We finally return to the adelic setting. We are interested in representations of $[\mathrm{GL}_2] = \mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\mathbb{A})$, or generalizations to arbitrary number fields; we will not find all such representations here, but some turn out to come from the objects we have been studying.

Recall that $\mathrm{SL}_2(\mathbb{R})$ acts transitively on $\mathcal{H}$, sending i to every point of $\mathcal{H}$ with stabilizer $\mathrm{SO}(2)$. Therefore we have a canonical homeomorphism $\mathrm{SL}_2(\mathbb{R})/\mathrm{SO}(2) \to \mathcal{H}$ sending $g\,\mathrm{SO}(2) \mapsto g\cdot i$. Thus we can think of modular forms, Γ-invariant functions on $\mathcal{H}$ for some suitable subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{R})$, as functions on the quotient $\Gamma\backslash \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$.

In the adelic setting, by strong approximation $\mathrm{SL}_2(\mathbb{A}) = \mathrm{GL}_1(\mathbb{A})\,\mathrm{SL}_2(\mathbb{Q})\,\mathrm{SL}_2(\mathbb{R})K$ for any compact open subgroup $K \subset \mathrm{SL}_2(\mathbb{A}_f)$, where GL_1 is embedded by $t \mapsto \begin{pmatrix} t & \\ & t^{-1}\end{pmatrix}$. In particular for $K = K_0(N)$, consisting of matrices over $\hat{\mathbb{Z}}$ with $c \equiv 0 \pmod N$, we have $\Gamma_0(N)\backslash \mathrm{SL}_2(\mathbb{R}) \simeq \mathrm{GL}_1(\mathbb{A})\,\mathrm{SL}_2(\mathbb{Q})\backslash \mathrm{SL}_2(\mathbb{A})/K_0(N)$. (We could write this in a similar way for GL_2 or GL_2^+, essentially equivalently.) Therefore modular forms for $\Gamma_0(N)$ are functions on $\Gamma_0(N)\backslash \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}(2) \simeq \mathrm{GL}_1(\mathbb{A})\,\mathrm{SL}_2(\mathbb{Q})\backslash \mathrm{SL}_2(\mathbb{A})/K_0(N)\,\mathrm{SO}(2)$. If $f \in S_k(\Gamma_0(N),\chi)$, we can view χ as a character on $\mathbb{Q}^\times\backslash\mathbb{A}^\times$, and can further evaluate it on $K_0(N)$ by $\begin{pmatrix} a & b\\ c & d\end{pmatrix} \mapsto \chi(d)$. We call this character $\lambda : K_0(N) \to \mathbb{C}^\times$.

We want to lift f to a function on $\mathrm{GL}_2(\mathbb{A}_f)$. We can naturally think of it as a function on $\mathrm{GL}_2(\mathbb{R})^+$ by $\gamma \mapsto (f|\gamma)(i)$. This tells us what happens at the archimedean place; at the nonarchimedean places we act by λ. In other words, we define a function on $\mathrm{GL}_2(\mathbb{A})$ by first using strong approximation to write $\gamma = \gamma_0\gamma_\infty k$ for $\gamma_0 \in \mathrm{GL}_2(\mathbb{Q})$, $\gamma_\infty \in \mathrm{GL}_2(\mathbb{R})$, and $k \in K_0(N)$. Then we define

$$\gamma = \gamma_0 g_\infty k \mapsto (f|\gamma_\infty)(i)\cdot\lambda(k).$$

One has to check that this is well-defined, but it turns out to be, and defines an adelic automorphic form for $\mathrm{GL}_2(\mathbb{Q})\backslash \mathrm{GL}_2(\mathbb{A})$. It generates a representation of this quotient, as a subrepresentation of the space of automorphic forms, which is is irreducible, admissible, and automorphic; it factors over the places of $\mathbb{Q}$, and it is possible to describe each place explicitly as well as the action of the Hecke operators, though beyond what we have done here.

References

[1] Jeremy Booher. Viewing modular forms as automorphic representations, 2015.

[2] Daniel Bump. *Automorphic forms and representations.* Number 55. Cambridge University Press, 1998.