

MODULAR SYMBOLS

PAUL E. GUNNELLS

ABSTRACT. Expanded notes from three lectures given by Paul E. Gunnells at the 2014 UNCG Summer School in Computational Number Theory: Modular Forms and Geometry.
<http://www.uncg.edu/mat/numbertheory/summerschool/2014.html>

LECTURE 1. MODULAR FORMS AND APPLICATIONS

The goal of these lectures is to explain how to compute effectively with classical holomorphic modular forms. The main approach is the *modular symbol method*, due to work of Birch, Manin, Mazur, Merel, and Cremona.

Definitions and notation. Let

$$\mathfrak{H} = \text{upper halfplane} = \{z \in \mathbb{C} : \text{Im}(z) > 0\}$$
$$\text{SL}_2(\mathbb{Z}) = \left\{ \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a, b, c, d \in \mathbb{Z}, \det(\gamma) = 1 \right\}.$$

Then $\text{SL}_2(\mathbb{Z})$ acts on \mathfrak{H} by

$$z \mapsto \frac{az + b}{cz + d}.$$

For each *weight* $k \geq 2$, we get an action on functions $f: \mathfrak{H} \rightarrow \mathbb{C}$ called the *slash operator*:

$$(f|_k\gamma)(z) = f\left(\frac{az + b}{cz + d}\right) (cz + d)^{-k}, \quad \gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \text{SL}_2(\mathbb{Z}).$$

Definition 1.1. A function $f: \mathfrak{H} \rightarrow \mathbb{C}$ is a *modular form of weight k* if

- (a) f is holomorphic
- (b) $(f|_k\gamma) = f$ for all $\gamma \in \text{SL}_2(\mathbb{Z})$
- (c) f is holomorphic “at infinity”, which means as $\text{Im}(z) \rightarrow \infty$, $|f(z)|$ is majorized by a polynomial in $\max\{1, \text{Im}(z)^{-1}\}$

Let M_k denote the \mathbb{C} -vector space of weight k modular forms.

We get the notion of a *cuspidal form* by imposing stronger growth conditions, namely f decays very rapidly as $\text{Im}(z) \rightarrow \infty$. More precisely, replace (c) by (c)': $|f|$ is majorized by $\text{Im}(z)^{k/2}$ as $\text{Im}(z) \rightarrow \infty$. Let $S_k \subset M_k$ be the subspace of cuspidal forms.

Fact 1.2. *The space of weight k modular forms M_k is finite-dimensional.*

Date: April 25, 2015.

2010 Mathematics Subject Classification. Primary 11F75; Secondary 11F11, 11F67.

Key words and phrases. Modular symbols, modular forms, Hecke operators.

The author thanks the organizers for the invitation to speak. He also thanks Dan Yasaki for providing a preliminary LaTeXed version of his notes.

Base revision a124061, Tue Apr 21 11:46:05 2015 -0400, Dan Yasaki.

Fourier expansion of f . Let $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$. Observe $f|_k \gamma = f$ means f is invariant under $z \mapsto z + 1$. Thus f has a Fourier expansion

$$f(z) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n z}, \quad a_n \in \mathbb{C}.$$

Usually we put $q = e^{2\pi i z}$ and write this as a q -expansion of f :

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

One can show that the growth conditions (c) and (c)' are equivalent to

$$\begin{aligned} a_n = 0 \text{ for all } n < 0 &\iff f \in M_k, \\ a_n = 0 \text{ for all } n \leq 0 &\iff f \in S_k. \end{aligned}$$

Under the change of coordinates $z \mapsto q = e^{2\pi i z}$, the upper halfplane maps to the unit disk $\{q \in \mathbb{C} : |q| < 1\}$. The point at $i\infty$ gets taken to the origin in the disk. In these new coordinates, saying $f \in M_k$ means f is bounded as $q \rightarrow 0$ in the disk, and can thus be extended to a function defined on the disk. Similarly, saying $f \in S_k$ means that f extends to a function vanishing at 0 on the disk.

The space S_k is more than just a complex vector space. It actually has a Hermitian product on it, the *Petersson product*:

$$\langle f, g \rangle = \int_D y^k f \bar{g} \, dA$$

where D is a fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$ in \mathfrak{H} and dA is hyperbolic measure. We can't compute the product of two Eisenstein series (the integral doesn't converge), but we can compute the inner product of an Eisenstein series and a cusp form. Using the inner product it's possible to prove

$$M_k \simeq \mathbb{C}E_k \oplus S_k$$

is an orthogonal decomposition.

Why do we study modular forms? As we shall see, sometimes we have a sequence

$$\{\alpha_n : n \in \mathbb{Z}_{\geq 0}\} \subset \mathbb{C}$$

arising naturally. For instance, we might have $\alpha_n \in \mathbb{Z}$, and they may count something.

Combinatoricists use a generating function $\sum \alpha_n x^n$ to organize these numbers. Number theorists, on the other hand, replace x by q and make a q -series. Replacing x by q is trivial, but nevertheless suggestive. One can ask: Is the resulting series the q -expansion of a weight k modular form?

If this is true, then $f \in M_k$, and the latter is a vector space of rather small dimension (roughly $k/12$). We can then take a basis of M_k and can express the function f in terms of this basis; this typically already leads to nontrivial information about the coefficients of f . Another typical phenomenon is that we may have other sequences g_1, g_2, \dots giving rise to modular forms in M_k coming from quite different settings. Since M_k has small dimension, this leads to nontrivial relations among f and the g_i , relations that are not at all obvious from the sources of these series.

This is best understood through examples, as we now illustrate. This also gives us the chance to introduce some key players in the theory.

Example 1.3 (Eisenstein series). The simplest way to try to make a modular form is by averaging: we can average over $\mathrm{SL}_2(\mathbb{Z})$ to force invariance under the slash action. Put $k \geq 4$, and define

$$E_k(z) := \frac{(k-1)!}{2(2\pi i)^k} \sum'_{m,n \in \mathbb{Z}} (mz+n)^{-k}$$

(the normalizing factor is used for convenience). This sum is absolutely convergent if $k \geq 4$, and we get a modular form $E_k \in M_k$, called the *holomorphic weight k Eisenstein series*. Note E_k vanishes identically for odd k . When $k = 2$ the series doesn't converge absolutely, but there is a standard way to sum the series conditionally (*Hecke's trick*). In this case the result is not a modular form, but it's close: It's called a *quasimodular form* and satisfies the

The Eisenstein series E_k has Fourier expansion

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

where σ_r is the r^{th} power divisor sum

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

Note

$$\frac{1}{2} \zeta(1-k) = -\frac{B_k}{2k},$$

where B_k is the k^{th} Bernoulli number. The first few q -expansions are

- (1) $E_4 = \frac{1}{240} + q + 9q^2 + 28q^3 + \dots,$
- (2) $E_6 = -\frac{1}{504} + q + 33q^2 + 244q^3 + \dots,$
- (3) $E_8 = \frac{1}{480} + q + 129q^2 + 2188q^3 + \dots$

Now the direct sum of all the spaces of modular forms

$$M_* = \bigoplus_k M_k$$

forms a graded ring, where the weight gives the grading: if f has weight k and g has weight l , then fg is a modular form of weight $k+l$. One can prove

$$(4) \quad M_* \simeq \mathbb{C}[E_4, E_6].$$

Thus any weight k modular form can be written as a (weighted) homogeneous polynomial in the Eisenstein series E_4, E_6 , which allows one to easily compute the dimension of M_k . Immediately we can get a nontrivial identity: one can check $\dim(M_4) = \dim(M_8) = 1$, which means E_4^2 must be a multiple of E_8 . Checking constant terms of Fourier expansions, we see

$$120E_4^2 = E_8.$$

Now look at the Fourier coefficients. We get

$$\sigma_7(n) = \sigma_3(n) + 120 \sum_{m=1}^{n-1} \sigma_3(m) \sigma_3(n-m),$$

which is not obvious (try to prove it directly!).

Example 1.4 (Delta function). The first weight with $S_k \neq 0$ is $k = 12$: M_{12} is spanned by E_4^3, E_6^2 , and these are not equal. The difference

$$(5) \quad \Delta(q) := 8000E_4^3 - 147E_6^2 = q - 24q^2 + 252q^3 + \dots$$

has no constant term and is thus a cusp form. The coefficients of the q -expansion give the values of Ramanujan's τ -function:

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

Thus our expression in terms of Eisenstein series gives a way to compute $\tau(n)$ using sums of powers of divisors of n . But Δ has even more structure. One can prove that Δ satisfies an infinite product formula

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

most modular forms, of course, have no such product structure. This shows that $\Delta = \eta(q)^{24}$, where η is Dedekind's eta-function.

Example 1.5 (Theta series of even, unimodular lattices). Now we have an arithmetic application. Let L be an even, unimodular lattice in \mathbb{R}^n . This means

- (1) $L \subset \mathbb{R}^n$ is a discrete, cocompact subgroup,
- (2) the inner product in \mathbb{R}^n is \mathbb{Z} -valued when restricted to L ,
- (3) L has a \mathbb{Z} -basis $\{v_1, \dots, v_n\}$ such that the Gram matrix $(v_i \cdot v_j)$ has determinant ± 1 (*unimodular*), and
- (4) $v \cdot v \in 2\mathbb{Z}$ for all $v \in L$ (*even*).

It is known that even, unimodular lattices exist in \mathbb{R}^n if and only if $n \equiv 0 \pmod{8}$. There are finitely many up to rotation. In general, the number of such lattices is unknown except for small values of n (cf. Table 1).

TABLE 1. Even unimodular lattices in \mathbb{R}^n .

| n | $\#L$ | Name |
|-----|-----------------|---|
| 8 | 1 | The root lattice E_8 |
| 16 | 2 | $E_8 \oplus E_8$ and the root lattice D_{16} |
| 24 | 24 | The 24 Niemeier lattices (includes the Leech lattice) |
| 32 | over 1000000000 | |

Define

$$r_L(m) = \#\left\{x \in L: \frac{x \cdot x}{2} = m\right\},$$

and form the q -expansion

$$f_L(q) = \sum_{m \geq 0} r_L(m)q^m.$$

Then one can prove the following:

This is correct although it looks quite ugly. Another typical normalization of the Eisenstein series puts the constant terms to be 1, i.e. $\tilde{E}_4 = 240E_4$, $\tilde{E}_6 = -504E_6$, \dots . With this convention, the expression (5) becomes $\Delta = (\tilde{E}_4^3 - \tilde{E}_6^2)/1728$, which is much more attractive.

Fact 1.6. *Let $L \subset \mathbb{R}^n$ be an even, unimodular lattice. Then $f_L(q)$ is a modular form of weight $n/2$.*

Here are two applications of this fact. First, consider the root lattice of type E_8 . Then $f_{E_8}(q) \in M_4$, which we know is spanned by the Eisenstein series E_4 . Comparing constant terms, we find $f_{E_8} = 240E_4$. This implies

$$r_{E_8}(m) = 240\sigma_3(m);$$

check it for $m = 2!$

Next consider $n = 16$. There are two even unimodular lattices in this dimension, $L_1 = E_8 \oplus E_8$ and a new one L_2 , which is the root lattice D_{16} . Now $f_{L_1}(q)$ and $f_{L_2}(q)$ are both weight 8 modular forms with constant coefficient 1. Since the space of weight 8 modular forms is one-dimensional and is spanned by the Eisenstein series $E_8(q)$ (don't mix this up with the root lattice $E_8!$), both these modular forms must be equal. (In fact by (3) they equal $480E_8(q)$).

Thus these two lattices have the property the number of vectors of a given length is the same for both. This is relevant to a famous problem in differential geometry, which asks *Can you hear the shape of a manifold?* Precisely, the question means *Does the spectrum of the Laplacian on a Riemannian manifold uniquely determine it, up to isometry?* The answer, as observed by Milnor, is no. The lattices determine two 16-dimensional flat tori $T_1 = \mathbb{R}^{16}/L_1$ and $T_2 = \mathbb{R}^{16}/L_2$. If $\Lambda \subset \mathbb{R}^n$ is a lattice with associated flat torus $T = \mathbb{R}^n/\Lambda$, then the eigenfunctions for the Laplacian have the form

$$f_{\lambda^*}(x) := e^{2\pi\sqrt{-1}(\lambda^* \cdot x)},$$

where λ^* is any point in the dual of Λ (by definition the dual of Λ is all λ^* such that $\lambda^* \cdot \lambda \in \mathbb{Z}$ for all $\lambda \in \Lambda$). Furthermore, the eigenvalue of $f_{\lambda^*}(x)$ is $4\pi^2|\lambda^*|^2$. The lattices L_i are self-dual, so the sequence of Laplacian eigenvalues is essentially what's encoded by the q -expansions $f_{L_i}(q)$. Thus $f_{L_1}(q) = f_{L_2}(q)$ implies that T_1 and T_2 are isospectral. On the other hand, T_1 and T_2 are non-isometric (there is no isometry of \mathbb{R}^{16} taking L_1 into L_2).

Level structure. For arithmetic applications, one needs the notion of modular forms with level. To define these, we need congruence subgroups.

Definition 1.7. Fix $N \in \mathbb{Z}_{>0}$. The *principal congruence subgroup* $\Gamma(N)$ is defined by

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

A subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ is called a *congruence subgroup* if Γ contains $\Gamma(N)$ for some N . The minimal such N is called the *level*.

The principal congruence subgroup $\Gamma(N)$ has finite index in $\mathrm{SL}_2(\mathbb{Z})$; indeed, one can show that $\Gamma(N)$ fits into an exact sequence

$$1 \longrightarrow \Gamma(N) \longrightarrow \mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \longrightarrow 1$$

(the tricky part is the surjectivity onto $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$). Thus every congruence subgroup also has finite index. The converse, however, is not true: not every finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is a congruence subgroup.

The most important congruence subgroups besides $\Gamma(N)$ are the *Hecke congruence subgroups*:

$$\Gamma_0(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{bmatrix} * & * \\ 0 & * \end{bmatrix} \pmod{N} \right\},$$

$$\Gamma_1(N) = \left\{ \gamma \in \mathrm{SL}_2(\mathbb{Z}) : \gamma \equiv \begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix} \pmod{N} \right\}.$$

Definition 1.8. Suppose Γ is a congruence subgroup. We say $f: \mathfrak{H} \rightarrow \mathbb{C}$ is a *weight k modular form on Γ* if

- (a) f is holomorphic,
- (b) $f|_k \gamma = f$ for all $\gamma \in \Gamma$, and
- (c) the previous growth condition now holds for $f|_k \gamma$ for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

Let $M_k(\Gamma)$ denote the \mathbb{C} -vector space of weight k modular forms on Γ . If $\Gamma = \Gamma_0(N)$, we usually just write $M_k(N)$ etc.

The last condition is a generalization of holomorphic at ∞ . It is more complicated because there is more than one way to go to infinity, and by requiring the growth condition to hold for $f|_k \gamma$ for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we are holomorphic at infinity for all possible cases. We will say more about this about this later.

Let $M_k(\Gamma)$ be the space of modular forms on Γ . As before this is a finite-dimensional complex vector space, and there is a distinguished subspace $S_k(\Gamma)$ of cusp forms. Just like the case of full level, f is a cusp form if $f|_k \gamma$ decays rapidly to zero as $\Im z$ goes to infinity, where γ varies over all of $\mathrm{SL}_2(\mathbb{Z})$. The Petersson product makes sense (just use the same definition but integrate over a fundamental domain for Γ), and the complement of the cusp forms in $M_k(\Gamma)$ is the subspace of Eisenstein series $\mathrm{Eis}_k(\Gamma)$. We have

$$M_k(\Gamma) = S_k(\Gamma) \oplus \mathrm{Eis}_k(\Gamma),$$

an orthogonal decomposition with respect to the Petersson inner product.

So far everything looks the same, but there is a difference. Unlike the case of full level, it is not true in general that $M_*(\Gamma)$ is a polynomial ring over a fixed set of Eisenstein series. In fact, the Eisenstein series usually aren't sufficient to generate $M_*(\Gamma)$ as a graded ring; some cusp forms must be taken too. And once one has a set of generators, there are usually nontrivial relations among them. However, just like the case of full level, it is still true that the ring of modular forms is always finitely presented.

There is a close connection between the groups $\Gamma_0(N)$ and $\Gamma_1(N)$, and in fact one can investigate modular forms on $\Gamma_1(N)$ by enlarging the scope of objects considered on $\Gamma_0(N)$. Let $\chi: \mathbb{Z} \rightarrow \mathbb{C}$ be a Dirichlet character of level N . This means $\chi(n+N) = \chi(n)$; $\chi(n) = 0$ if and only if $(n, N) > 1$; and $\chi(mn) = \chi(m)\chi(n)$. Thus χ induces a map

$$\chi: \mathbb{Z}/N\mathbb{Z} \rightarrow \mathbb{C}$$

that is nonzero exactly on $(\mathbb{Z}/N\mathbb{Z})^\times$, and when nonzero takes values in the roots of unity. We have

$$(6) \quad \Gamma_0(N)/\Gamma_1(N) \simeq (\mathbb{Z}/N\mathbb{Z})^\times$$

by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \longmapsto d \pmod{N}.$$

Hence we can understand modularity with respect to $\Gamma_1(N)$ by incorporating a character χ into the action of $\Gamma_0(N)$. More precisely, for $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma_0(N)$, put

$$(f|_{k,\chi}\gamma)(z) = \overline{\chi(d)}(cz + d)^{-k} f(\gamma z).$$

We can define the space $M_k(N, \chi)$ by replacing the condition $f|_k\gamma = f$ with $f|_{k,\chi} = f$. This leads to the vector space $M_k(N, \chi)$, which is called the space of weight k modular forms of level N and *nebentype* χ . By (6) we have

$$M_k(\Gamma_1(N)) \simeq \bigoplus_{\chi} M_k(N, \chi).$$

Hecke operators. The space of modular forms M_k admits a huge collection of commuting linear operators, the *Hecke operators*. Moreover, they are Hermitian with respect to the natural inner product on M_k . Thus we can look for simultaneous eigenclasses. It is these eigenclasses and their eigenvalues that reveal the hidden arithmetic information in the modular forms. They are crucial for arithmetic applications, and motivate the main goal of our lectures: how to effectively compute spaces of modular forms and the Hecke action on them.

For now, we just define the Hecke operators; later we will see how to compute them. Let n be a fixed positive integer. Define a subset $\mathcal{X}_n \subset M_2(\mathbb{Z})$ by

$$\mathcal{X}_n = \left\{ \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} : a \geq 1, ad = n, 0 \leq b < d \right\}.$$

Extend the slash action on functions $f: \mathfrak{H} \rightarrow \mathbb{C}$ from matrices in $\mathrm{SL}_2(\mathbb{Z})$ to $\mathrm{GL}_2(\mathbb{Q})$ via

$$(f|_k\gamma)(z) = (\det(\gamma))^{k-1}(cz + d)^{-k} f(\gamma z).$$

Now we can apply the elements of \mathcal{X}_n to modular forms. Suppose f is a weight k modular form of full level. Then the action of the Hecke operator T_n on f is defined by

$$(T_n f)(z) := \sum_{\gamma \in \mathcal{X}_n} (f|_k\gamma)(z).$$

Note that to be pedantic, we really should write fT_n (i.e. the Hecke operator should act on modular forms on the right, since the matrices in \mathcal{X}_n are acting by the slash operator, which is a right action). But as we said, one knows that the Hecke operators commute with each other. Thus it doesn't matter whether we write the operators acting on the right or left.

Why is this an action, and why are these interesting operators? Certainly, if you've never seen it before, it's not clear why this is an action. The main thing to check is that if f is modular, so is $T_n f$. The point is that the set \mathcal{X}_n is in bijection with a certain subset of lattices. Namely, we have

$$\mathcal{X}_n \iff \{L \subset \mathbb{Z}^2 : [\mathbb{Z}^2 : L] = n\},$$

i.e., \mathcal{X}_n is in bijection the set of sublattices of \mathbb{Z}^2 of index n . The bijection itself is easy to describe: any such lattice has a basis of the form $ae_1 + be_2, de_2$, where $\mathbb{Z}^2 = \mathbb{Z}e_1 \oplus \mathbb{Z}e_2$. Now an alternative perspective on modular forms describes them as certain functions on lattices in \mathbb{C} : to any weight k modular form f we can attach a function $F = F_f$, where

$$F: \{\text{lattices in } \mathbb{C}\} \longrightarrow \mathbb{C}$$

satisfies the homogeneity condition

$$F(\lambda L) = \lambda^{-k} F(L), \quad \text{for all } \lambda \in \mathbb{C}^\times.$$

For more discussion, see [Ser73, VII.2.2]. So from this perspective, the effect of the Hecke operator T_n is to define a new function $T_n F$ that averages F over the index n sublattices of its input [Ser73, VII.5.1]. This is certainly a very natural operation on functions defined on lattices, although why this reveals the arithmetic information hidden in M_k is less obvious.

The operators satisfy

$$(7) \quad T_n T_m = T_{nm} \quad \text{if } (n, m) = 1, \text{ and}$$

$$(8) \quad T_{p^n} = T_{p^{n-1}} T_p - p^{k-1} T_{p^{n-2}} \quad \text{for } p \text{ prime.}$$

These identities follow from the description of \mathcal{X}_n in terms of sublattices. We can compute the operators directly on q -expansions. If $f(q) = \sum a_n q^n$, then

$$(9) \quad (T_n f)(q) = \sum_{m \in \mathbb{Z}} \left(\sum_{\substack{d \geq 1 \\ d|(m,n)}} d^{k-1} a_{mn/d^2} \right) q^m.$$

In particular, for p prime (9) becomes

$$(10) \quad (T_p f)(q) = \sum_{m \geq 0} (a_{mp} + p^{k-1} a_{m/p}) q^m.$$

These formulas give an algorithm to compute Hecke operators, although not a very good one: simply compute q -expansions of a basis of M_k as far as one needs, using (4) and the q -expansions of the Eisenstein series, then apply (9) and find the action of T_p in terms of the basis. (What makes this algorithm not great is that computing the coefficient of q^m in $T_p f$ needs the coefficient a_{mp} .) In any case, we see that if f is an eigenform, and if we normalize so that $a_1 = 1$, then the Fourier coefficient a_n is the eigenvalue of T_n , and from (7)–(8) the Fourier coefficients satisfy

$$(11) \quad a_n a_m = a_{nm} \quad \text{if } (n, m) = 1, \text{ and}$$

$$(12) \quad a^{p^n} = a_{p^{n-1}} a_p - p^{k-1} a_{p^{n-2}} \quad \text{for } p \text{ prime.}$$

We can also define Hecke operators for modular forms with level structure N , but we must be careful if $(n, N) \neq 1$. For T_p , if $p \mid N$ then we only use the elements $\begin{bmatrix} 1 & a \\ 0 & p \end{bmatrix} \in \mathcal{X}_p$, in other words we omit $\begin{bmatrix} p & 0 \\ 0 & 1 \end{bmatrix}$. The resulting operator is usually denoted U_p .

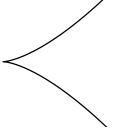
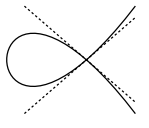
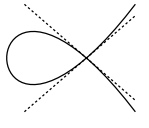
Now that we have level structure and Hecke operators, we can give an example to show how the operators reveal the arithmetic information hidden in the modular forms. Let E/\mathbb{Q} be an elliptic curve. Concretely, we can consider E to be a nonsingular plane curve defined by the equation

$$(13) \quad y^2 + a_1 xy + a_3 = x^3 + a_2 x^2 + a_4 + a_6, \quad a_i \in \mathbb{Z},$$

although in doing so we are missing one point (the point at infinity). The equation (13) can be reduced modulo any prime p since the a_i are integral, and one knows that for almost all p the resulting curve $E(\mathbb{F}_p)$ is nonsingular. Using the finitely many p for which $E(\mathbb{F}_p)$ one

You should read this book anyway, if you're interested in number theory. It's one of the greats.

TABLE 2. Singularity type and corresponding a_p value.

| a_p | Picture | Description |
|-------|---|---|
| 0 |  | cusp |
| 1 |  | node, slopes of tangency defined over \mathbb{F}_p |
| -1 |  | node, slopes of tangency defined over $\mathbb{F}_{p^2} \setminus \mathbb{F}_p$ |

can define the *conductor* of E ; it is an integer N_E such that E/\mathbb{F}_p is nonsingular if and only if $p \nmid N_E$. In general N_E is not squarefree, but there is an explicit algorithm to determine it.

Now we want to attach a Dirichlet series to E . Define a sequence $\{a_n\} \subset \mathbb{Z}$ as follows. If $p \nmid N$, put $a_p = p + 1 - \#E(\mathbb{F}_p)$ (this enumeration of points on $E \bmod p$ also includes the point at infinity). If $p \mid N$, then $a_p \in \{0, \pm 1\}$ depending on the singularity E acquires mod p . If $E(\mathbb{F}_p)$ has a cusp mod p we put $a_p = 0$. If $E(\mathbb{F}_p)$ has a node mod p , then we put $a_p = 1$ (respectively, -1) if the slopes of the two tangents to the node lie in \mathbb{F}_p (respectively, lie in $\mathbb{F}_{p^2} \setminus \mathbb{F}_p$.) We extend the definition from the a_p to all a_n via an *Euler product*:

$$(14) \quad \sum a_n n^{-s} := \prod_{p \nmid N} (1 - a_p p^{-s} + p^{1-2s})^{-1} \cdot \prod_{p \mid N} (1 - a_p p^{-s}).$$

We get a_n by expanding the factors on the right of (14) into geometric series, just as one does to prove the Euler product for the Riemann ζ -function

$$\zeta(s) = \sum n^{-s} = \prod (1 - p^{-s})^{-1}.$$

The Dirichlet series

$$L(E, s) = \sum_{n>0} a_n/n^s$$

is called the *L-function of the elliptic curve* E . For instance, if E is defined by the equation $y^2 + y = x^3 - x$, then $N_E = 37$. We have

$$L(E, s) = 1 - 2/2^s - 3/3^s + 2/4^s - 2/5^s + \dots$$

The coefficients a_2, a_3 , and a_5 are determined by counting points mod p , whereas a_1 and a_4 are determined using the Euler product (14).

Now we have the following amazing theorem. I personally consider myself extremely lucky to have been around when this theorem was proved:

Theorem 1.9. *Let E/\mathbb{Q} be an elliptic curve of conductor N_E and let $f_E = \sum a_n q^n$, where the a_n are defined as above. Then f_E is the q -expansion of a Hecke eigenform in $S_2(N_E)$.*

Why is this so amazing? **fixme:** something about this theorem and how great it is

LECTURE 2. MODULAR SYMBOLS

Modular curves. Our ultimate goal is to explain how to compute with modular forms. Now that we have defined modular forms, the first step is to learn more about the geometry of *modular curves*, which are quotients of \mathfrak{H} by congruence subgroups. This will also help us understand the statement of the growth conditions for modular forms on congruence subgroups.

Let $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ be a congruence subgroup, that is a group containing $\Gamma(N)$ for some N . Then $\Gamma \backslash \mathfrak{H}$ is an open Riemann surface, in other words topologically is an orientable surface of some genus with some punctures.

We can canonically compactify $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ by adding *cusps*. First define

$$\mathfrak{H}^* = \mathfrak{H} \cup \mathbb{Q} \cup \{\infty\} = \mathfrak{H} \cup \mathbb{P}^1(\mathbb{Q}),$$

where $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ and $\{\infty\}$ is considered to be a single point infinitely far up the imaginary axis.

We need to put a topology on \mathfrak{H}^* . We do this by first extending the action of $\mathrm{SL}_2(\mathbb{Z})$ on \mathfrak{H} to an action on \mathfrak{H}^* . There are two ways to think about this:

- (1) We can act directly on $\mathbb{P}^1(\mathbb{Q})$. For $z \in \mathbb{Q} \subset \mathbb{P}^1(\mathbb{Q})$, we put

$$z \mapsto \frac{az + b}{cz + d},$$

where we use the convention that $z \mapsto \infty$ if $z = -d/c$. In other words, we act directly on fractions where the “fraction” $1/0$ is considered to be the point at infinity in $\mathbb{P}^1(\mathbb{Q})$.

- (2) We can convert to integral vectors and then act: the fraction $\frac{m}{n}$, written in reduced terms, is converted to the vector $\begin{bmatrix} m \\ n \end{bmatrix} \in \mathbb{Z}^2$, with ∞ corresponding to $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then the action of $\mathrm{SL}_2(\mathbb{Z})$ is just by matrix multiplication:

$$\begin{bmatrix} m \\ n \end{bmatrix} \mapsto \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} m \\ n \end{bmatrix}.$$

Now we define a topology on \mathfrak{H}^* . For a basis of open sets of ∞ we take the sets

$$B_c := \{z \in \mathfrak{H} : \mathrm{Im}(z) > c\}.$$

The $\mathrm{SL}_2(\mathbb{Z})$ -translates are open disks tangent to the rational points of the real axis (cf. Figure 1). This gives a system of neighborhoods of $\partial\mathfrak{H}^* = \mathfrak{H}^* \setminus \mathfrak{H}$. This induces a topology called the *Satake topology* on $\Gamma \backslash \mathfrak{H}^*$. With this topology, the quotient $\Gamma \backslash \mathfrak{H}^*$ is now a compact Riemann surface.

Definition 2.1. The Γ -orbits in $\mathbb{P}^1(\mathbb{Q})$, and their images in the quotient $\Gamma \backslash \mathfrak{H}^*$, are called *cusps*.

Why are these points called cusps? The quotient $\Gamma \backslash \mathfrak{H}$ is more than just a topological surface. It has an induced metric, since the standard hyperbolic metric on \mathfrak{H} is Γ -invariant. The metric on $\Gamma \backslash \mathfrak{H}^*$ degenerates to 0 as one approaches a cusp, and in fact the surface appears metrically to be a sharp horn in a neighborhood of a cusp. In other words, a cusp looks metrically like a cusp!

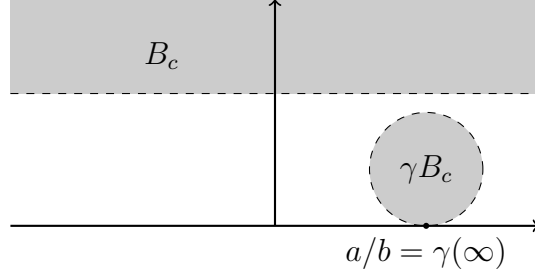


FIGURE 1. Open neighborhoods in the Satake topology on \mathfrak{H} .

Now is also a good time to explain the connection between the cusps and the growth condition for modular forms with level. As we said before, for a general finite-index subgroup Γ of $SL_2(\mathbb{Z})$ there is more than one way to go to infinity on the quotient $\Gamma \backslash \mathfrak{H}$. The different ways correspond exactly (surprise) to the cusps, and our growth condition is effectively ensuring that the image of f doesn't blow up as one approaches a cusp on $\Gamma \backslash \mathfrak{H}^*$. However, there is a subtlety lurking here: since f is not *invariant* under the left action of Γ on \mathfrak{H} , f does not induce a function on the quotient $\Gamma \backslash \mathfrak{H}$. However, it is the section of a certain line bundle on $\Gamma \backslash \mathfrak{H}$, so the growth condition guarantees that this section extends over the cusps.

When Γ is one of our special congruence subgroups, we will use the following notation for its quotients:

| Γ | $\Gamma \backslash \mathfrak{H}$ | $\Gamma \backslash \mathfrak{H}^*$ |
|---------------|----------------------------------|------------------------------------|
| $\Gamma(N)$ | $Y(N)$ | $X(N)$ |
| $\Gamma_0(N)$ | $Y_0(N)$ | $X_0(N)$ |
| $\Gamma_1(N)$ | $Y_1(N)$ | $X_1(N)$ |

Here are some examples.

Example 2.2.

$$Y(1) \simeq \mathbb{P}^1 \setminus \{\text{pt}\}, \quad X(1) \simeq \mathbb{P}^1,$$

so there is only one cusp. This is not hard to show directly: one checks that the group $SL_2(\mathbb{Z})$ acts transitively on $\mathbb{P}^1(\mathbb{Q})$.

Example 2.3.

$$Y(3) \simeq \mathbb{P}^1 \setminus \{4 \text{ pts}\}, \quad X(3) \simeq \mathbb{P}^1.$$

See Figure 2. Note that the four cusps correspond exactly to the four points of $\mathbb{P}^1(\mathbb{F}_3)$.

Example 2.4.

$$Y(7) \simeq C_3 \setminus \{24 \text{ pts}\}, \quad X(7) \simeq C_3, \text{ (a surface of genus 3).}$$

This time there are more cusps than points in $\mathbb{P}^1(\mathbb{F}_7)$, which has order 6.

Example 2.5.

$$Y_0(11) = C_1 \setminus \{2 \text{ pts}\}, \quad X_0(11) \simeq C_1, \text{ (surface of genus 1).}$$

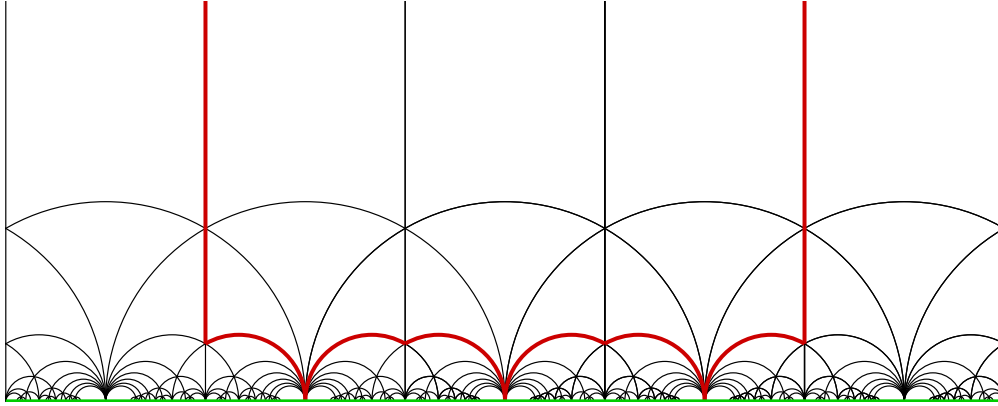


FIGURE 2. A fundamental domain for $\Gamma(3)$ is outlined in red. The four cusps are the three shown on the real axis and ∞ . The edge identifications are the obvious ones that yield $X(3) \simeq \mathbb{P}^1$.

Weight 2 modular symbols. Finally, we can start talking about modular symbols. Let's focus on weight 2 for now. Suppose $f \in S_2(\Gamma)$. Then f is not a function on X_Γ , as we said before, but $f dz$ is a holomorphic 1-form on X_Γ . To see why, first look at how the product $f dz$ transforms under Γ :

$$f\left(\frac{az+b}{cz+d}\right) d\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z) \frac{ad-bc}{(cz+d)^2} dz = f dz.$$

Thus the modularity of f implies that the differential form $f dz$ is invariant under Γ . One then needs to see that $f dz$ is actually holomorphic on X_Γ . This is a straightforward computation in local coordinates; the only tricky parts are checking what happens at ramified points of the map $\mathfrak{H}^* \rightarrow X_\Gamma$. In particular, one can check that $f dz$ is actually holomorphic at the cusps, which means that any weight two cusp form determines a holomorphic 1-form on X_Γ . Conversely, any holomorphic 1-form on X_Γ can be written as $f(z)dz$ for $f \in S_2(\Gamma)$. The theory of Riemann surfaces shows that $\dim(S_2(\Gamma)) = g(X_\Gamma)$, where g the genus of X_Γ as an orientable topological surface.

Now suppose that α and β are cusps that are equivalent mod Γ . We can use them to construct a homology class: we take any reasonable oriented path between α and β on \mathfrak{H} , say the geodesic directed from α to β , and then take the image mod Γ . Since α and β are equivalent mod Γ , the image becomes a closed oriented 1-curve on X_Γ , i.e. a 1-cycle. Thus we get a class in $H_1(X_\Gamma; \mathbb{Z})$. Let us denote this class by $\{\alpha, \beta\}$. Note that this notation looks a lot like the set $\{\alpha, \beta\}$, but it's not: it really represents an *ordered* pair, since if we change the roles of α and β we reverse the orientation on the cycle and thus get the opposite class: $\{\beta, \alpha\} = -\{\alpha, \beta\}$. This can be confusing, but the notation is traditional.

Now consider the pairing $S_2(\Gamma) \times H_1(X_\Gamma; \mathbb{Z}) \rightarrow \mathbb{C}$ given by integration

$$(15) \quad (f, \{\alpha, \beta\}) \mapsto 2\pi i \int_\alpha^\beta f(z) dz := \langle \{\alpha, \beta\}, f \rangle.$$

It is important to take f to be a cusp form here. In fact, the (omitted) computation in local coordinates shows that if f is nonzero at a cusp, then the differential form $f dz$ will have a pole of order 1 there. This is caused by the effect of the nontrivial stabilizer of a cusp in Γ on the local coordinates. See Milne for details.

This is independent of the path between α and β since f is holomorphic (essentially this boils down to Cauchy's theorem from complex analysis). Note also that f has to be a cusp form for the integral make sense; if f is nonvanishing at the cusp, say when f is an Eisenstein series, the integral diverges.

We can extend (15) from integral homology to real homology to get a pairing

$$S_2(\Gamma) \times H_1(X_\Gamma; \mathbb{R}) \rightarrow \mathbb{C}.$$

This is done in the obvious way. First choose an integral basis of $H_1(X_\Gamma; \mathbb{Z})$. Any class in $H_1(X_\Gamma; \mathbb{R})$ can be written as a linear combination of this basis with real coefficients, so we can extend the pairing using linearity.

Now recall that

$$\dim_{\mathbb{C}}(S_2(\Gamma)) = g,$$

from our discussion about weight 2 cusp forms and holomorphic 1-forms. Thus as a real vector space, we have

$$\dim_{\mathbb{R}}(S_2(\Gamma)) = 2g,$$

which is the same as $\dim_{\mathbb{R}}(H_1(X_\Gamma; \mathbb{R}))$. This is not a coincidence:

Claim 2.6. *The pairing $S_2(\Gamma) \times H_1(X_\Gamma; \mathbb{R}) \rightarrow \mathbb{C}$ perfect, and identifies the dual $S_2(\Gamma)^\vee$ of $S_2(\Gamma)$ with $H_1(X_\Gamma; \mathbb{R})$.*

In fact, the truth of this claim has nothing to do with modular forms. It's really a combination of Poincaré duality and the Hodge theorem. There is a slight subtlety in that the differentiable structure of X_Γ is more complicated at some points, namely those whose preimages in \mathfrak{H}^* have nontrivial stabilizers, but nevertheless everything works out.

Now we want to extend the notation $\{\alpha, \beta\}$ to include cusps that aren't necessarily equivalent mod Γ . This is done by integration: we can still integrate f along the geodesic from α to β , which produces a number. Thus these two cusps determine a linear form on $S_2(\Gamma)$, and so define an element of $S_2(\Gamma)^\vee = H_1(X_\Gamma; \mathbb{R})$. Thus again $\{\alpha, \beta\}$ gives a class in $H_1(X_\Gamma; \mathbb{R})$.

Definition 2.7. The *modular symbol* attached to the pair of cusps α, β is the real homology class $\{\alpha, \beta\} \in H_1(X_\Gamma; \mathbb{R})$.

Here are some basic properties of modular symbols:

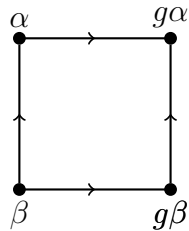
- (1) $\{\alpha, \beta\} = -\{\beta, \alpha\}$ (2-term relation)
- (2) $\{\alpha, \beta\} = \{\alpha, \gamma\} + \{\gamma, \beta\}$ (3-term relation)
- (3) $\{g\alpha, g\beta\} = \{\alpha, \beta\}$ for all $g \in \Gamma$ (Γ -action)
- (4) $\{\alpha, g\alpha\} \in H_1(X_\Gamma; \mathbb{Z})$
- (5) $\{\alpha, g\alpha\} = \{\beta, g\beta\}$

These are all easy to verify. The 2-term relation just says that reversing the limits of integration introduces a minus sign. The 3-term relation says that we can divide an integral into two integrals by introducing a common new endpoint. Perhaps the last is the most complicated. It can be proved by considering the square in Figure 3.

Properties (4) and (5) imply that we have constructed a map

$$\begin{aligned} \Gamma &\longrightarrow H_1(X_\Gamma; \mathbb{Z}) \\ g &\longmapsto \{\alpha, g\alpha\} \end{aligned}$$

that is independent of α .

FIGURE 3. $\{\alpha, g\alpha\} = \{\beta, g\beta\}$

By the way, our construction of modular symbols means that all we can say a priori is that $\{\alpha, \beta\} \in H_1(X_\Gamma; \mathbb{R})$, i.e. $\{\alpha, \beta\}$ is a real homology class. However, the theorem of Manin–Drinfeld tells us that this class often lies in the rational homology $H_1(X_\Gamma; \mathbb{Q}) = H_1(X_\Gamma; \mathbb{Z}) \otimes \mathbb{Q}$:

Theorem 2.8 (Manin–Drinfeld). *If Γ is a congruence subgroup, and α, β are cusps of Γ , then $\{\alpha, \beta\} \in H_1(X_\Gamma; \mathbb{Q})$.*

Why is this important? I.e., why should it matter whether a homology class lives in the real homology or the rational homology? The point is, this theorem has arithmetic consequences. For instance, suppose f has q -expansion $\sum a_n q^n$. We can make an L -function from f as in the previous section using the Dirichlet series built from the a_n :

$$L(f, s) = \sum a_n / n^s.$$

(The discussion before might lead one to believe that the a_n need to be Hecke eigenvalues, since there we were connecting modular forms to elliptic curves. But this is not true.) We can make a more direct connection between f and its L -function using the Mellin transform of f . We have

$$(16) \quad L(f, s) = \frac{(2\pi)^s}{\Gamma(s)} \int_0^{i\infty} (-iz)^s f(z) \frac{dz}{z}.$$

Now the L -function of a modular form satisfies many properties, the most important of which is the existence of a functional equation taking s into $2 - s$ (the 2 comes from f having weight 2). The central point $s = 1$ is especially important for many applications (cf. our discussion of the BSD **fixme**: write this). Evaluating (16) at $s = 1$, we have

$$L(f, 1) = -2\pi i \int_0^{i\infty} f(z) dz = -\langle \{0, \infty\}, f \rangle.$$

Thus the fact that the modular symbol $\{0, \infty\}$ is a rational homology class means that the special value $L(f, 1)$ is a rational multiple of a period of f . In other words, the Manin–Drinfeld theorem implies that the quantity $L(f, 1)$, which is a priori an extremely complicated transcendental number, actually lies in the rational span of certain other numbers, still transcendental to be sure, but nevertheless more tractable.

Let’s go back to the general discussion. At this point we’ve written almost all the relations needed to reconstruct H_1 from our symbols. Specifically, let $\mathcal{M}_2(\Gamma)$ denote the \mathbb{Q} -vector space generated by the $\{\alpha, \beta\}$, modulo the 2-term and 3-term relations and Γ -action. Then we have the following result of Manin relating modular symbols to a *relative* homology group. Such groups can be unfamiliar to some, so we take a moment to recall them. Suppose X is

a space with a nice subspace Y . We have the chain complexes $C_*(X)$, $C_*(Y)$ that can be used to compute their homology. We have an inclusion $C_*(Y) \rightarrow C_*(X)$ and can form the quotient chain complex $C_*(X)/C_*(Y)$. Then the relative homology of the pair (X, Y) is the homology of this complex. We denote relative homology by $H_*(X, Y; \mathbb{Z})$. Intuitively, the difference between $H_*(X)$ and $H_*(X, Y)$ is that in the latter, we consider a chain to be a cycle not only if its boundary vanishes, but also if its boundary lies in $C_*(Y)$. Now we can state Manin's key theorem:

Theorem 2.9 (Manin). *We have*

$$\mathcal{M}_2(\Gamma) \xrightarrow{\sim} H_1(X_\Gamma, \partial X_\Gamma; \mathbb{Q}).$$

For example, recall that the modular curve $X_0(11)$ has genus 1 and has 2 cusps. Thus $X_0(11)$ is topologically a torus, and as one learns in topology class the usual homology group $H_1(X_0(11); \mathbb{Q})$ has dimension 2. We claim the relative homology $H_1(X_0(11), \partial X_0(11); \mathbb{Q})$ is 3-dimensional. Indeed, we still have the two closed 1-cycles giving our 2 dimensions from before, and now there is an additional class, which can be represented by a path from one cusp to the other. See Figure 5.

The space $\mathcal{M}_2(\Gamma)$ is a good start, but we are primarily interested in part of the homology relevant for studying the cusp forms, in other words $H_1(X_\Gamma; \mathbb{Q})$. But it is easy to identify the subspace of $\mathcal{M}_2(\Gamma)$ mapping onto the usual homology. From our example above, it's clear that we don't want relative classes that have boundary in the cusps. Instead we want those relative classes with vanishing boundary at the cusps. Formally, let $\mathcal{B}_2(\Gamma)$ be the \mathbb{Q} -vector space generated by the cusps of X_Γ , equipped with the obvious Γ -action. Define

$$\partial: \mathcal{M}_2(\Gamma) \longrightarrow \mathcal{B}_2(\Gamma),$$

by

$$\{\alpha, \beta\} \longmapsto \beta - \alpha.$$

A moments thought shows that this definition makes sense (the point is one has to think about the relations defining $\mathcal{M}_2(\Gamma)$ and make sure that the map is well-defined modulo them.) Put $\mathcal{S}_2(\Gamma) = \ker(\partial)$. It is clear that this is the subspace we want. Classes in $\mathcal{S}_2(\Gamma)$ are called *cuspidal modular symbols*. Manin proved that cuspidal modular symbols exactly capture the homology of X_Γ :

Theorem 2.10 (Manin). *We have an isomorphism*

$$(17) \quad \mathcal{S}_2(\Gamma) \xrightarrow{\sim} H_1(X_\Gamma; \mathbb{Q}).$$

After tensoring with \mathbb{R} , it follows from (17) that we have an isomorphism

$$(18) \quad \mathcal{S}_2(\Gamma) \otimes \mathbb{R} \xrightarrow{\sim} \mathcal{S}_2(\Gamma)^\vee,$$

and thus have a topological model of the vector space of cusp forms.

Hecke operators and unimodular symbols. At this point we have found a way to connect the topology of the modular curve X_Γ to weight 2 modular forms on Γ . This is great but isn't good enough for number theory. The point is, we have Hecke operators acting on modular forms, and unless we can incorporate them into our model, it doesn't do us much good. But amazingly, the pairing between cusp forms and cycles, and the identification (18), are compatible with the Hecke action. Namely, there exists an action of the Hecke operators

directly on the modular symbols: given a symbol $\{\alpha, \beta\}$ and an n , we can define a new (sum of) symbol(s) $T_n\{\alpha, \beta\}$, and we have the fundamental relation

$$(19) \quad \langle T_n\{\alpha, \beta\}, f \rangle = \langle \{\alpha, \beta\}, T_n f \rangle.$$

Furthermore, the action on symbols is simple to describe. We can use the matrices \mathcal{X}_n from before that we used to define the Hecke action on modular forms. Let's take the set \mathcal{X}_p , where p is a prime not dividing the level. Then we define

$$(20) \quad T_p\{\alpha, \beta\} = \sum_{g \in \mathcal{X}_p} \{g\alpha, g\beta\}.$$

The same conditions on \mathcal{X}_p that guarantee that the Hecke image of a modular form is modular also guarantee that the right of (20) is a well-defined modular symbol. The relation (19) implies that if we can find eigenclasses and eigenvalues in $\mathcal{S}_2(\Gamma)$, then we can recover eigenclasses and eigenvalues in $S_2(\Gamma)$. This is great news, but unfortunately there's a catch: in its present form, our model for $\mathcal{S}_2(\Gamma)$ is not computable. The problem is that the current definitions give infinite presentations of $\mathcal{M}_2(\Gamma)$ and $\mathcal{S}_2(\Gamma)$ (as spaces spanned by infinitely many symbols divided by infinitely many relations).

To address this, we want to identify a *finite* generating set of $\mathcal{M}_2(\Gamma)$. To this end, we introduce *unimodular symbols*. These are the symbols given by the pairs of cusps corresponding to the edges of the *Farey tessellation* of \mathfrak{H} (Figure 4). To make this picture, take the ideal triangle in \mathfrak{H} with vertices at the cusps $\{0, 1, \infty\}$. Then the $\mathrm{SL}_2(\mathbb{Z})$ -translates of this triangle fill out all of \mathfrak{H} . The edges are the $\mathrm{SL}_2(\mathbb{Z})$ -translates of the geodesic connecting 0 to ∞ .

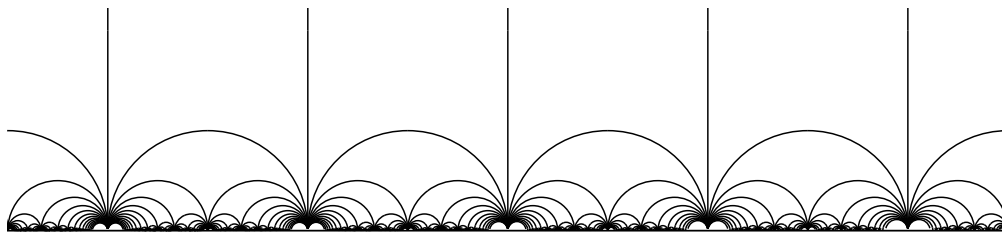


FIGURE 4. Farey tessellation of \mathfrak{H} . The edges are the $\mathrm{SL}_2(\mathbb{Z})$ -translates of the geodesic from 0 to ∞ .

Since Γ has finite-index in $\mathrm{SL}_2(\mathbb{Z})$, there are only finitely many unimodular symbols mod Γ . Thus the unimodular symbols yield a computable version of $\mathcal{S}_2(\Gamma)$, at least potentially: we of course need to know that $\mathcal{S}_2(\Gamma)$ is spanned by them, and that all the relations needed to cut out $\mathcal{S}_2(\Gamma)$ can be written using unimodular symbols (this is actually a separate question). We also have the problem that the Hecke operators can't possibly preserve unimodularity. This is clear from the definition (20); in general a symbol of the form $\{g \cdot 0, g \cdot \infty\}$ won't correspond to an edge of the tessellation.

We solve these difficulties in one stroke.

If $\Gamma = \Gamma_0(N)$ or $\Gamma_1(N)$, then we just mean $p \nmid N$. For a general congruence subgroup Γ , we can just fix N minimal such that $\Gamma(N) \subset \Gamma$, and then our discussion applies to $p \nmid N$.

Theorem 2.11 (Manin’s trick, a.k.a. the modular symbol algorithm). *For cusps α and β , we have the relation*

$$\{\alpha, \beta\} = \sum \{\alpha_i, \beta_i\},$$

where each term is unimodular.

Proof. Without loss of generality, assume

$$\{\alpha, \beta\} = \left\{0, \frac{p}{q}\right\}.$$

Make simple continued fraction for $\frac{p}{q}$

$$\frac{p}{q} = a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \frac{1}{\ddots + \frac{1}{a_r}}}} = \llbracket a_1, a_2, \dots, a_r \rrbracket.$$

We get convergents

$$\frac{p_k}{q_k} := \llbracket a_1, a_2, \dots, a_k \rrbracket.$$

Then $\left\{\frac{p_k}{q_k}, \frac{p_{k+1}}{q_{k+1}}\right\}$ is unimodular, and our desired relation is

$$\left\{0, \frac{p}{q}\right\} = \{0, \infty\} + \left\{\infty, \frac{p_1}{q_1}\right\} + \left\{\frac{p_1}{q_1}, \frac{p_2}{q_2}\right\} + \dots + \left\{\frac{p_{r-1}}{q_{r-1}}, \frac{p_r}{q_r}\right\}.$$

□

Example 2.12. Let’s express $\left\{0, \frac{71}{31}\right\}$ as a sum of unimodular symbols. We have

$$\frac{71}{31} = \llbracket 2, 3, 2, 4 \rrbracket.$$

Then the convergents are

$$\llbracket 2 \rrbracket = 2, \quad \llbracket 2, 3 \rrbracket = \frac{7}{3}, \quad \text{and} \quad \llbracket 2, 3, 2 \rrbracket = \frac{16}{7}.$$

Thus

$$\left\{0, \frac{71}{31}\right\} = \{0, \infty\} + \{\infty, 2\} + \left\{2, \frac{7}{3}\right\} + \left\{\frac{7}{3}, \frac{16}{7}\right\} + \left\{\frac{16}{7}, \frac{71}{31}\right\}.$$

OK, now we have

- a finite, computable model of $\mathcal{M}_2(\Gamma)$ and $\mathcal{S}_2(\Gamma)$,
- an algorithm for to compute Hecke operators.

To go further, we specialize to $\Gamma = \Gamma_0(N)$ (actually just $\Gamma_0(p)$ right now). We also introduce another trick that’s even faster than the modular symbol algorithm for Hecke operator computations.

Proposition 2.13. *We have a bijection*

$$\Gamma \backslash \mathrm{SL}_2(\mathbb{Z}) \xrightarrow{\sim} \mathbb{P}^1(\mathbb{F}_p)$$

given by the bottom row map

$$\Gamma \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto (c : d).$$

Proof. The group $\mathrm{SL}_2(\mathbb{Z})$ acts transitively on $\mathbb{P}^1(\mathbb{F}_p)$, and the stabilizer of $(0 : 1)$ is Γ . \square

Thus we can identify cosets in $\Gamma \backslash \mathrm{SL}_2(\mathbb{Z})$ with $\mathbb{P}^1(\mathbb{F}_p)$. This implies unimodular symbols mod Γ are in bijection with $\mathbb{P}^1(\mathbb{F}_p)$.

What about the relations? We need 2-term and 3-term relations. Let $S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$, and let $R = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$.

Claim 2.14. *The relations are those of the form*

$$(21) \quad (c : d) + (c : d)S = 0$$

$$(22) \quad (c : d) + (c : d)R + (c : d)R^2 = 0.$$

We have (21) from the orientation reversing identity (2-term relation). We have (22) because the boundary of a triangle is zero (3-term relation). Why? Lift $(c : d)$ to a matrix in $\mathrm{SL}_2^\pm(\mathbb{Z})$ to get a unimodular symbol. Then (21) visibly flips the orientation, and (22) finds one of the two Farey triangles with this as an edge (we get the other for another choice of lift). Note we use the \pm here because the determinant of a Farey edge is either ± 1 .

Computing, we get

$$(23) \quad (c : d) + (-d : c)0$$

$$(24) \quad (c : d) + (-d : c + d) + (-c - d : c) = 0.$$

Theorem 2.15 (*M*-symbols). *The \mathbb{Q} -vector space generated by $\mathbb{P}^1(\mathbb{F}_p)$ modulo (23) and (24) is isomorphic to $\mathcal{M}_2(\Gamma_0(p))$.*

An example. It's time to actually compute something. Let's take $p = 11$ and figure out what's happening. The finite projective space $\mathbb{P}^1(\mathbb{F}_{11})$ has $12 = 11 + 1$ points. We take

$$(0 : 1), (1 : 0), (1 : 1), (1 : 2), \dots, (1 : A)$$

as representatives, and because we're lazy we abbreviate $(c : d)$ to cd . What are the relations? We start with a 12-dimensional \mathbb{Q} -vector space. The 2-term relation gives

$$\begin{array}{ll} 10 = -01 & 13 = -17 \\ 11 = -1A & 14 = -18 \\ 12 = -15 & 16 = -19. \end{array}$$

This cuts us down to a 6-dimensional space. The 3-term relation gives

$$\begin{array}{ll} 10 + 01 + 1A = 0 & 12 + 14 + 17 = 0 \\ 11 + 19 + 15 = 0 & 13 + 16 + 18 = 0 \end{array}$$

which implies

$$\begin{array}{ll} 10 + 01 - 11 = 0 & 12 + 14 - 13 = 0 \\ 11 - 16 - 12 = 0 & 13 + 16 - 14 = 0. \end{array}$$

Combining these with the 2-term relations cuts the space down to a 3-dimensional space. Namely, we get everything in terms of 10, 12, and 14.

This is what we expect. See Figure 5. Obviously $\mathcal{S}_2(\Gamma)$ is spanned by 12 and 14. **fixme: why is this obvious?**

There are two choices:

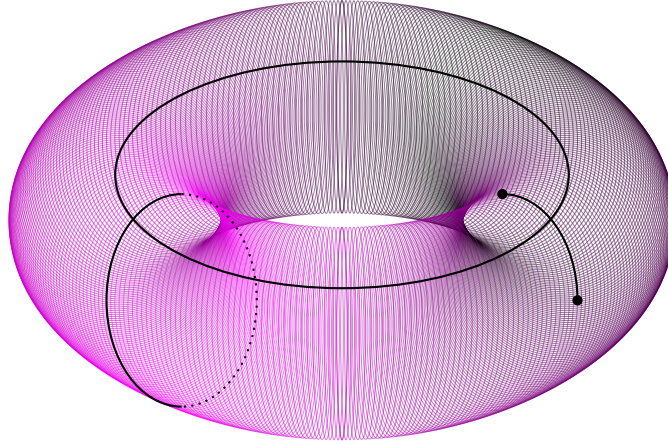


FIGURE 5. The modular curve $X_0(11)$. The two cusps are given as black dots. Representatives of the three nontrivial homology classes in $H_1(X_0(11), \partial X_0(11); \mathbb{Q})$ are shown. In addition to the two obvious ones, there is a class joining the two cusps.

- (1) Lift M -symbols to modular symbols, and do the modular symbol algorithm.
- (2) Work directly with M -symbols (Mazur, Merel, Manin).

Definition 2.16. Let \mathcal{Y}_n be the set of integral matrices

$$\mathcal{Y}_n = \left\{ g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} : \det(g) = n, a > b \geq 0, d > c \geq 0 \right\}.$$

Note that $\#\mathcal{Y}_n$ is finite.

Claim 2.17. If $\ell \neq p$ is prime, then

$$T_\ell = \sum_{g \in \mathcal{Y}_\ell} (c : d)g.$$

Example 2.18. Consider $\ell = 2$.

$$\mathcal{Y}_2 = \left\{ \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix} \right\}.$$

Let's compute T_2 on $\mathcal{M}_2(11)$

$$\begin{aligned} (1 : 0)T_2 &= 10 + 10 + 10 + 16 \\ &= 3 \cdot 10 - 12 \end{aligned}$$

$$\begin{aligned} (1 : 2)T_2 &= 14 + 15 + 11 + 17 \\ &= 14 - 12 - 12 - 14 \\ &= -2 \cdot 12 \end{aligned}$$

$$\begin{aligned} (1 : 4)T_2 &= 18 + 16 + 12 + 18 \\ &= 14 - 12 + 12 - 14 \\ &= -2 \cdot 14. \end{aligned}$$

The matrix of T_2 is

$$T_2 = \begin{bmatrix} 3 & 0 & 0 \\ -1 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix}.$$

The eigenvalues are 3, -2 , -2 . This is what we expect. From [Cre97, page 110] we get an equation for this curve

$$y^2 + y = x^3 - x^2 - 10x - 20.$$

From [Cre97, page 265], we get the Hecke eigenvalues. This elliptic curve has $a_2 = -2$. (Note: $a_p := p + 1 - \#E(\mathbb{F}_p)$.) We get it twice. The other eigenvalue is coming from the the Eisenstein series. Similarly, we find that the matrix of T_3 is

$$T_3 = \begin{bmatrix} 4 & 0 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

What are the eigenvectors? They are 12, 14 and $10 - \frac{1}{5} \cdot 12$. The first two are integral, but the third is rational. The denominator of this class is interesting because it is an Eisenstein homology class. We have $\mathcal{M}_2(\Gamma)$ is really dual to $S_2(\Gamma) \oplus \text{Eis}_2(\Gamma)$.

To generalize this to $\Gamma_0(N)$ with N not necessarily prime, we use $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$ instead of $\mathbb{P}^1(\mathbb{F}_p)$. This means we consider tuples $(a, b) \bmod N$, where $\gcd(a, b, N) = 1$ modulo the action of $(\mathbb{Z}/N\mathbb{Z})^*$. e.g., For $N = 4$, we can choose representatives

$$\mathbb{P}^1(\mathbb{Z}/4\mathbb{Z}) = \{(1, 0), (0, 1), (1, 1), (1, 3), (1, 2), (2, 1)\}.$$

LECTURE 3. HIGHER WEIGHT

OK that was great, but what about higher weight? We need to enlarge the coefficients. We must be careful because now the discrete group acts. Let

$$\begin{aligned} \mathcal{M}_2 &:= \mathbb{Q}\text{-vector space on symbols } \{\alpha, \beta\} \\ &\quad \text{modulo 2-term and 3-term relations,} \\ \mathcal{M}_k &:= \mathbb{Q}[X, Y]_{k-2} \otimes_{\mathbb{Q}} \mathcal{M}_2, \end{aligned}$$

where $\mathbb{Q}[X, Y]_{k-2}$ is the space of homogeneous polynomials in x and y of degree $k-2$. Then Γ acts on \mathcal{M}_k since it acts on $\mathbb{Q}[X, Y]_{k-1}$ and \mathcal{M}_2 . To ease the notation, we omit \otimes and just write $P \otimes \{\alpha, \beta\}$ as $P\{\alpha, \beta\}$. Let $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \Gamma$. Then for $P \in \mathbb{Q}[X, Y]_{k-1}$ and $\{\alpha, \beta\} \in \mathcal{M}_2$,

$$\begin{aligned} (gP)(X, Y) &:= P\left(g^{-1} \begin{bmatrix} X \\ Y \end{bmatrix}\right) = P(dX - bY, -cX + aY), \\ g\{\alpha, \beta\} &:= \{g\alpha, g\beta\}, \end{aligned}$$

and so

$$g(P\{\alpha, \beta\}) = gP\{g\alpha, g\beta\}.$$

We can now define higher weight modular symbols and cuspidal modular symbols just as we did for weight 2.

Definition 3.1.

$$\mathcal{M}_k(\Gamma) := \mathcal{M}_k / (P\{\alpha, \beta\} - g(P\{\alpha, \beta\}))$$

To get $\mathcal{S}_k(\Gamma)$ we can take the kernel of a boundary map as we did before. Let

$$\begin{aligned} \mathcal{B}_2 &:= \mathbb{Q}\text{-vector space on symbols } \{\alpha\} \text{ for } \alpha \in \mathbb{P}^1(\mathbb{Q}), \\ \mathcal{B}_k &:= \mathbb{Q}[X, Y]_{k-2} \otimes_{\mathbb{Q}} \mathcal{B}_2, \\ \mathcal{B}_k(\Gamma) &:= \mathcal{B}_k / (x - gx). \end{aligned}$$

Then the boundary map $\partial: \mathcal{M}_k(\Gamma) \rightarrow \mathcal{B}_2(\Gamma)$ is given by

$$\partial(P\{\alpha, \beta\}) = P\{\beta\} - P\{\alpha\}.$$

Then

$$\mathcal{S}_k(\Gamma) := \ker(\partial).$$

Pairing. As before, we have a pairing of cuspforms and modular symbols. Let

$$\begin{aligned} S_k(\Gamma) &:= \mathbb{C}\text{-vector space of weight } k \text{ holomorphic cuspforms} \\ \bar{S}_k(\Gamma) &= \mathbb{C}\text{-vector space of weight } k \text{ antiholomorphic cuspforms} \\ &= \{\bar{f}: f \in S_k(\Gamma)\}. \end{aligned}$$

The integration pairing is now

$$\begin{aligned} S_k(\Gamma) \oplus \bar{S}_k(\Gamma) \times \mathcal{M}_k(\Gamma) &\rightarrow \mathbb{C} \\ \langle (f_1, f_2), P\{\alpha, \beta\} \rangle &= \int_{\alpha}^{\beta} f_1(z)P(z, 1) dz + \int_{\alpha}^{\beta} f_2(z)P(\bar{z}, 1) d\bar{z}. \end{aligned}$$

Theorem 3.2 (Shokurov). *The pairing*

$$\langle \cdot, \cdot \rangle: S_k(\Gamma) \oplus \bar{S}_k(\Gamma) \times \mathcal{S}_k(\Gamma) \otimes_{\mathbb{Q}} \mathbb{C} \rightarrow \mathbb{C}$$

is a nondegenerate pairing of \mathbb{C} -vector spaces.

One application is computing special values of L-functions at critical integers. The L-function $L(f, s)$ has functional equation of shape $s \mapsto k - s$. The integers $j = 1, \dots, k - 1$ are called *critical* [Del79]. They are analogues of $s = 1$ in the weight 2 case.

We have

$$L(f, j) = \frac{(-2\pi i)^j}{(j-1)!} \langle f, X^{j-1}Y^{k-2-(j-1)}\{0, \infty\} \rangle.$$

M-symbols for $\Gamma_0(N)$. As before, we work with points in $\mathbb{P}^1(\mathbb{Z}/N\mathbb{Z})$. Coefficients are now homogeneous polynomials in two variables of degree $k - 2$. Let

$$S = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}, \quad \text{and} \quad J = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}.$$

(We need J because the action on coefficients is nontrivial!) Define a right action of $\text{SL}_2(\mathbb{Z})$ on \mathbb{P}^1 by **fixme: be more explicit about what this \mathbb{P}^1 actually is?**

$$(P(c : d))g = (g^{-1}P)((c : d)g).$$

Then $\mathcal{M}_k(N)$ is the \mathbb{Q} -vector space generated by $x = X^i Y^{k-2-i}(c : d) \in \mathbb{P}^1$, modulo

$$\begin{aligned}x + xS &= 0, \\x + xR + xR^2 &= 0, \\x - xJ &= 0,\end{aligned}$$

for all x as above.

REFERENCES

- [Cre97] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997. MR 1628193 (99e:11068)
- [Del79] P. Deligne, *Valeurs de fonctions L et périodes d'intégrales*, Automorphic forms, representations and L -functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2, Proc. Sympos. Pure Math., XXXIII, Amer. Math. Soc., Providence, R.I., 1979, With an appendix by N. Koblitz and A. Ogus, pp. 313–346. MR 546622 (81d:12009)
- [Ser73] J.-P. Serre, *A course in arithmetic*, Springer-Verlag, New York-Heidelberg, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7. MR 0344216 (49 #8956)

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF MASSACHUSETTS, AMHERST, MA 01003-9305

E-mail address: gunnells@math.umass.edu

URL: <http://www.math.umass.edu/~gunnells/>