(Classical) modular forms and modular symbols

FRANCESC GISPERT

Montréal, 12th April 2019

Abstract

These are the (extended) notes for a talk given in the graduate students seminar[1] at Concordia University. I present briefly the theory of (classical) modular forms and show how they can be computed using (classical) modular symbols. The talk is aimed at a general audience of mathematicians, so I always specialize to the simplest cases without mention and take an informal approach. The notes are a summary of various well-known results (mostly without proofs). No originality is claimed.

1 Modular forms

Consider the complex upper half-plane

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

You might have encountered $\mathbb{H}$ before in one of the following contexts:

(1) One can define a metric for which $\mathbb{H}$ is a model for the hyperbolic plane known as the Poincaré half-plane mode. We are not going to do any hyperbolic geometry, but this interpretation will be used implicitly in the pictures to talk about *lines* or *geodesics*, *triangles*, etc.

(2) By Riemann's uniformization theorem, $\mathbb{H}$ is one of the three simply connected Riemann surfaces (up to isomorphism). It is also isomorphic to the open unit disk. The automorphism group of $\mathbb{H}$ has a very concrete description. Namely, $\operatorname{Aut}(\mathbb{H}) \cong \mathrm{PSL}_2(\mathbb{R})$ via the action of matrices by fractional linear transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} z = \frac{az+b}{cz+d} \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R}) \text{ and all } z \in \mathbb{H}.$$

[1] I thank Magloire Loudegui Djimdou for organizing the seminar.

In number theory, we usually use some discrete subgroups of $\mathrm{PSL}_2(\mathbb{R})$. For the rest of this document, fix a positive integer N and define

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

Definition 1. A *weakly modular form* is a meromorphic function $f \colon \mathbb{H} \to \mathbb{P}^1(\mathbb{C})$ with the property that

$$f\Big(\frac{az+b}{cz+d}\Big) = (cz+d)^2 f(z) \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \text{ and all } z \in \mathbb{H}.$$

Remark. The condition that defines weakly modular forms is that these functions must be *almost* invariant under the action of $\Gamma_0(N)$. Therefore, they are *almost* functions on the quotient $\Gamma_0(N) \setminus \mathbb{H}$, except for the *twist* introduced by the factor $(cz+d)^2$. In fact, weakly modular forms are sections of a line bundle: they will be differential forms. Indeed, the mysterious factor $(cz+d)^2$ appears from the computation

$$d\Big(\frac{az+b}{cz+d}\Big) = \frac{a(cz+d)-(az+b)c}{(cz+d)^2}\,dz = \det\begin{pmatrix} a & b \\ c & d \end{pmatrix} \frac{1}{(cz+d)^2}\,dz.$$

This shows that the expression $f(z)\,dz$ is actually invariant under the action of $\Gamma_0(N)$.

By the previous remark, we can view weakly modular forms as meromorphic differential 1–forms on the orbit space $Y_0(N) = \Gamma_0(N) \setminus \mathbb{H}$, which turns out to be a Riemann surface (the charts are *essentially* given by the natural projection $\mathbb{H} \twoheadrightarrow Y_0(N)$, which is a local homeomorphism except at a few ramification points where one must be more careful).

Let us take a closer look at what the surface $Y_0(N)$ looks like (at least topologically). It is well-known that the set

$$\mathcal{F} = \left\{ z \in \mathbb{H} : -\frac{1}{2} \leq \mathrm{Re}(z) \leq \frac{1}{2} \text{ and } |z| \geq 1 \right\}$$

is a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$. This means that $\mathcal{F}$ contains exactly one representative of each orbit except that we have some identifications on the boundary. Figure 1 shows $\mathcal{F}$ (in grey) and some of its translates under elements of $\mathrm{SL}_2(\mathbb{Z})$. This produces a tessellation of $\mathbb{H}$ (in figure 1, the three dots mean that there should be infinitely many more *triangles* forming a fractal picture). Since $\Gamma_0(N)$ is a subgroup of $\mathrm{SL}_2(\mathbb{Z})$, we can find a system of representatives of the quotient $\mathrm{SL}_2(\mathbb{Z}) \,/\, \Gamma_0(N)$ such that the union of the transforms of $\mathcal{F}$ by these matrices is a fundamental domain for the action of $\Gamma_0(N)$ on $\mathbb{H}$.

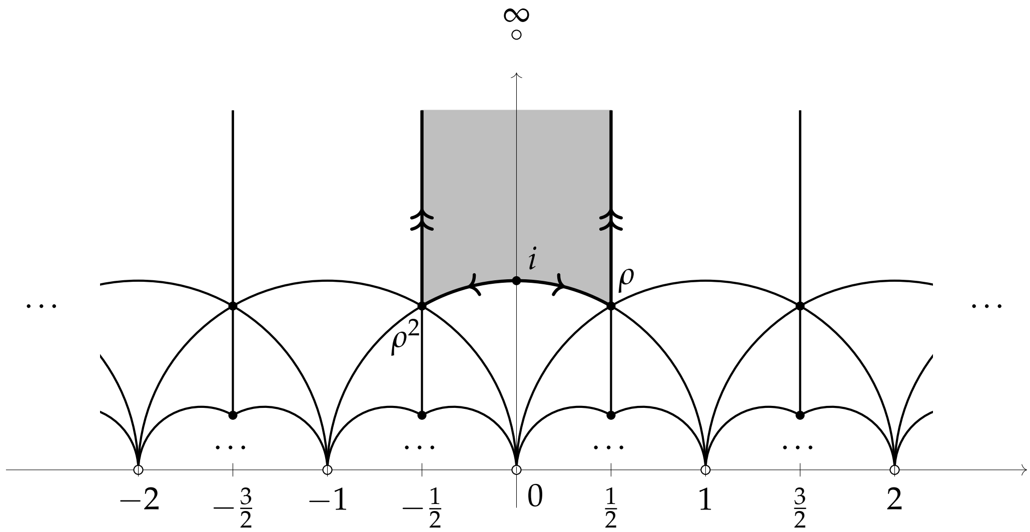

Figure 1: The fundamental domain $\mathcal{F}$ and some of its transforms.

In the particular case of $\Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z})$, the fundamental domain is a triangle minus a point: we can think of the two parallel lines $\mathrm{Re}(z) = -\frac{1}{2}$ and $\mathrm{Re}(z) = \frac{1}{2}$ as meeting at a point at infinity. That is, $\mathcal{F}$ is the triangle with vertices $\rho = e^{\pi i/3}$, ρ^2 and ∞ minus the point ∞, as shown in figure 1. Moreover, the sides from ρ to ∞ and from ρ^2 to ∞ are glued together and, similarly, the third side has its two halves from ρ to i and from ρ^2 to i glued together. Therefore, $Y_0(1)$ is homeomorphic to a sphere minus a point (see figure 2).

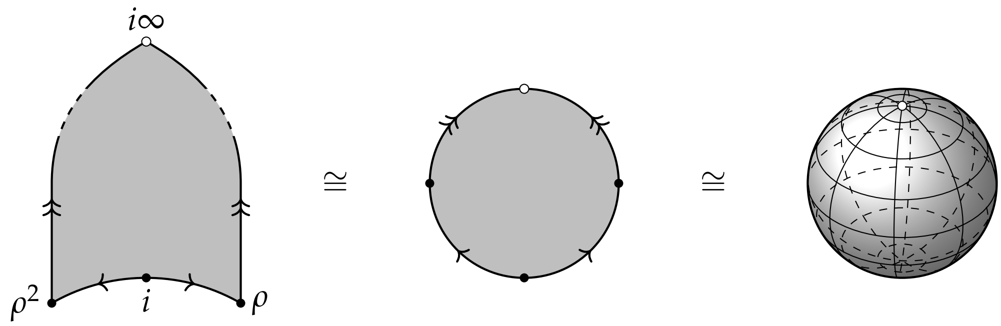

Figure 2: The (topological) surface $Y_0(1)$.

More generally, we can compactify the Riemann surface $Y_0(N)$ by adding the point ∞ and its $\mathrm{SL}_2(\mathbb{Z})$–conjugates to $\mathbb{H}$. But

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \infty = \frac{a\infty + b}{c\infty + d} = \frac{a}{c} \quad \text{for any } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$$

and every rational number can be obtained in this way: by Bézout's identity, given two coprime integers a and c, we can find integers b and d such that $ad - bc = 1$.

We define $\mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ and the orbit space $X = X_0(N) = \Gamma_0(N) \setminus \mathbb{H}^*$. The points of $\mathbb{P}^1(\mathbb{Q})$ are called the cusps (this name comes from the shape of the tessellation sketched in figure 1 at the rational numbers). We write $\pi \colon \mathbb{H}^* \twoheadrightarrow X$ for the canonical projection. It turns out that X can be endowed with the structure of a compact Riemann surface and so X is known as the *modular curve* (it is a complex curve). We want to define a subclass of weakly modular forms which correspond to holomorphic differential 1–forms on X. There is a clear notion of holomorphicity on $\mathbb{H}$, but what about holomorphicity at the cusps?

Take a weakly modular form $f \colon \mathbb{H} \to \mathbb{P}^1(\mathbb{C})$. The weak modularity condition applied to the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N)$$

is that

$$f(z+1) = f(z) \quad \text{for all } z \in \mathbb{H}.$$

That is to say, f is periodic of period 1. As f is also meromorphic, it admits a Fourier series expansion

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

Such a series is called a q–expansion.

The change of variable $z \mapsto q = e^{2\pi i z}$ defines a biholomorphism

$$\left\{ z \in \mathbb{H} : -\frac{1}{2} < \mathrm{Re}(z) \leq \frac{1}{2} \right\} \cong \{ q \in \mathbb{C} : 0 < |q| < 1 \}$$

which can be extended by $\infty \mapsto 0$. Thus, we interpret the q–expansion of f as a kind of Laurent series of f at ∞ and use it to define holomorphicity at ∞. In fact, the change of variables $z \mapsto q$ gives a chart of X at ∞ (as it identifies an open neighbourhood of ∞ with the open unit disc).

We have seen that the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{P}^1(\mathbb{Q})$ is transitive. Therefore, we could define a similar q–expansion at every $x \in \mathbb{Q}$ by translation by an appropriate element of $\mathrm{SL}_2(\mathbb{Z})$.

Definition 2. A *cusp form* is a holomorphic weakly modular form $f \colon \mathbb{H} \to \mathbb{C}$ whose q–expansions lie in $q \cdot \mathbb{C}[\![q]\!]$. Let S denote the space of cusp forms.

Remark. The condition that the q–expansions must lie in $q \cdot \mathbb{C}[\![q]\!]$ and not just in $\mathbb{C}[\![q]\!]$ might be a bit surprising at first, but a computation shows that it is indeed the right condition:

$$q = e^{2\pi i z} \implies dq = 2\pi i e^{2\pi i z}\, dz = 2\pi i q\, dz \implies dz = \frac{1}{2\pi i}\frac{dq}{q}.$$

The fact that the q–expansion of f is divisible by q makes $f(z)\, dz \in \mathbb{C}[\![q]\!]\, dq$. Thus, $S \cong \Omega^1(X)$ (i.e., cusp forms are naturally identified with holomorphic differential 1–forms on X).

There is not much arithmetic in the theory exposed up to this point. Number theorists are interested in modular forms because their q–expansions often occur as generating functions of invariants of arithmetic objects (e.g., numbers of points of elliptic curves over finite fields) and one can use the theory of modular forms to deduce properties of these arithmetic objects (e.g., by finding congruences between the coefficients of q–expansions). The main tool to do this is the extra structure provided by a family of operators on the space of cusp forms.

2 Hecke operators

Hecke operators are defined more naturally by means of the modular interpretation of $\Gamma_0(N) \setminus \mathbb{H}$. Namely, the points of this space *parametrize* elliptic curves with a cyclic subgroup of order N. However, in these notes we present an unmotivated definition of Hecke operators for brevity of exposition.

Definition 3. For each prime number p, define

$$\Delta(p) = \left\{ \begin{pmatrix} 1 & b \\ 0 & p \end{pmatrix} : b \in \mathbb{Z} \text{ and } 0 \le b < p \right\} \cup \left\{ \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

and define the *Hecke operator* T_p acting on S by

$$(\mathrm{T}_p f)(z)\, dz = \sum_{\alpha \in \Delta(p)} f(\alpha z)\, d(\alpha z).$$

We extend this definition and define *Hecke operators* T_m for all $m \in \mathbb{N}$ via the following recurrences:

(i) $\mathrm{T}_{mn} f = \mathrm{T}_m \mathrm{T}_n f$ for all $m, n \in \mathbb{N}$ with $(m, n) = 1$ and

(ii) $\mathrm{T}_p \mathrm{T}_{p^n} f = \mathrm{T}_{p^{n+1}} f + \chi_N(p) p \mathrm{T}_{p^{n-1}} f$ for all prime numbers p and all $n \in \mathbb{N}$, where χ_N is the trivial character modulo N, defined by

$$\chi_N(d) = \begin{cases} 1 & \text{if } (d, N) = 1, \\ 0 & \text{if } (d, N) > 1. \end{cases}$$

Using this definition, it is possible to compute the action of Hecke operators on q–expansions.

Proposition 4. *Let $f \in S$ and let $m \in \mathbb{N}$. If the q–expansion of f is*

$$f(z) = \sum_{n \in \mathbb{N}} a_n(f) q^n,$$

the q–expansion of $\mathrm{T}_m f$ is

$$(\mathrm{T}_m f)(z) = \sum_{n \in \mathbb{N}} a_n(\mathrm{T}_m f) q^n,$$

where

$$a_n(\mathrm{T}_m f) = \sum_{d \mid (n,m)} \chi_N(d) \cdot d \cdot a_{nm/d^2}(f) \quad \textit{for all } n \in \mathbb{N}.$$

Remarks.

1. For a prime number p, the operator T_p has a particularly simple expression in terms of q–expansions:

$$(\mathrm{T}_p f)(z) = \sum_{n \in \mathbb{Z}} \left[a_{np}(f) + \chi_N(p) p a_{n/p}(f) \right] q^n$$

(where, by convention, $a_{n/p}(f) = 0$ if $p \nmid n$).

2. Since modular forms are uniquely determined by their q–expansion, we could have defined Hecke operators directly by their action on q–expansions, which is quite simple. However, this approach has an important drawback: it is not clear from the formula why Hecke operators send cusp forms to cusp forms.

Next, we present the main properties which make Hecke operators interesting. Let $\mathbb{T}$ denote the $\mathbb{C}$–subalgebra of $\mathrm{End}_{\mathbb{C}}(S)$ generated by the Hecke operators. That is, $\mathbb{T} = \mathbb{C}[\mathrm{T}_m : m \in \mathbb{N}] \subset \mathrm{End}_{\mathbb{C}}(S)$.

Theorem 5. *The $\mathbb{C}$–algebra $\mathbb{T}$ is commutative.*

Proof. It follows from the recurrence relations defining the Hecke operators. □

It turns out that S admits a basis consisting *essentially* of eigenforms (i.e., simultaneous eigenvectors of all the Hecke operators). Indeed, one can define a hermitian inner product $\langle -, - \rangle$ on S and one can prove that *most* Hecke operators are self-adjoint with respect to $\langle -, - \rangle$. That is, we have a family of commuting self-adjoint operators acting on a finite-dimensional hermitian space and so they must be simultaneously diagonalizable by the spectral theorem of linear algebra. Therefore, in many situations it suffices to study eigenforms.

Let $f \in S$ be an eigenform. For every $m \in \mathbb{N}$, there exists $\lambda_m \in \mathbb{C}$ such that $\mathrm{T}_m f = \lambda_m f$. In terms of q–expansions, proposition 4 shows that

$$a_m(f) = a_1(\mathrm{T}_m f) = \lambda_m a_1(f) \text{ for all } m \in \mathbb{N}.$$

This motivates the following result.

Theorem 6. *The map*

$$\langle -, - \rangle \colon S \times \mathbb{T} \longrightarrow \mathbb{C}$$

given by $\langle f, T \rangle = a_1(\mathrm{T} f)$ *defines a perfect pairing of finite-dimensional* $\mathbb{C}$*–vector spaces and so induces an isomorphism* $S \cong \mathbb{T}^\vee$ *(here,* $(-)^\vee$ *means the dual space).*

The key idea that we can extract from the theorem is that, to understand S, it suffices to understand $\mathbb{T}$ instead. To study $\mathbb{T}$, we will construct a simpler space containing an isomorphic copy of S as a $\mathbb{T}$–algebra. This new $\mathbb{T}$–algebra will be the space of modular symbols.

3 Modular symbols

We go back to the geometry of the compact Riemann surface X. Recall that $g = \dim_{\mathbb{C}} \Omega^1(X)$ is the genus of X. Topologically, X is a g–holed torus. Hence, the first homology group $\mathrm{H}_1(X; \mathbb{Z})$ is a free $\mathbb{Z}$–module of rank $2g$, with two generators per hole as shown in figure 3.

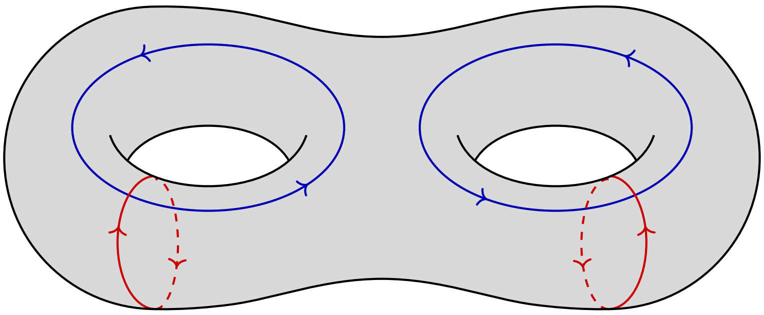

Figure 3: Generators of $\mathrm{H}_1(X; \mathbb{Z})$ (when $g = 2$).

In fact, there is a very deep connection between the two spaces $\Omega^1(X)$ and $\mathrm{H}_1(X; \mathbb{Z})$, as the following result (used to define the jacobian of a Riemann surface) indicates.

Theorem 7. *The integration pairing*

$$I \colon \mathrm{H}_1(X; \mathbb{Z}) \times \Omega^1(\Gamma) \longrightarrow \mathbb{C}$$

defined by

$$I\Big(\sum_{i=1}^{n} a_i [\varphi_i], \omega\Big) = \sum_{i=1}^{n} a_i \int_{\varphi_i} \omega$$

is non-degenerate and gives a lattice

$$\begin{array}{ccc} \mathrm{H}_1(X;\mathbb{Z}) & \hookrightarrow & \Omega^1(X)^\vee \\ \wr\| & & \wr\| \\ \mathbb{Z}^{2g} & \hookrightarrow & \mathbb{C}^g \end{array}$$

(here, $(-)^\vee$ means the dual space).

Remark. If we tensor $\mathrm{H}_1(X;\mathbb{Z})$ with $\mathbb{R}$, we get a bijection (by dimension reasons).

Corollary 8. *Integration of forms along paths induces an isomorphism of $\mathbb{R}$–vector spaces $\mathrm{H}_1(X;\mathbb{R}) \cong \Omega^1(X)^\vee$.*

Definition 9. The action of *Hecke operators* on $\mathrm{H}_1(X;\mathbb{R})$ is induced by the action on S by duality. That is, given $x \in \mathrm{H}_1(X;\mathbb{R})$ and $\mathrm{T} \in \mathbb{T}$, we define $\mathrm{T}\,x$ to be the unique element of $\mathrm{H}_1(X;\mathbb{R})$ such that

$$I(\mathrm{T}\,x,\omega) = I(x,\mathrm{T}\,\omega) \quad \text{for all } \omega \in \Omega^1(X).$$

Remark. Here, $\omega = f(z)\,dz$ for some $f \in S$ and we write $\mathrm{T}\,\omega = (\mathrm{T}\,f)(z)\,dz$ by abuse of notation.

At last we are in a position to define modular symbols. Modular symbols are essentially homology classes in disguise. Recall that our objective is to find a space with a simple (explicit) presentation with respect to which Hecke operators are easy to compute.

The elements of $\mathrm{H}_1(X;\mathbb{R})$ are *essentially* $\mathbb{R}$–linear combinations of paths on X. However, there are many such paths. We are going to choose a class of paths which arise from the construction of X as a quotient of the upper half-plane.

Definition 10. For $r,s \in \mathbb{P}^1(\mathbb{Q})$, the *modular symbol* $\{r,s\}$ is the image in $\mathrm{H}_1(X;\mathbb{R})$ of a geodesic path from r to s in $\mathbb{H}^*$.

Remarks.

1. A geodesic from r to s looks like a semicircle centred at the real axis and intersecting it at r and s (in the degenerate case in which one of r and s is ∞, the *semicircle* becomes a straight line parallel to the imaginary axis).
2. Figure 4 shows some examples of the situations that can occur. The geodesics on the upper half-plane project to paths between some points on the modular curve. Distinct cusps in $\mathbb{H}^*$ may or may not be equivalent in X, so some paths are loops and some are not.

3. In general, a path which is not a loop does not yield a homology class. A more precise definition would be the following: $\{r,s\}$ is the unique element of $\mathrm{H}_1(X;\mathbb{R})$ such that
$$I\big(\{r,s\},\omega\big) = \int_{\varphi_{rs}} \omega \quad \text{for all } \omega \in \Omega^1(X),$$
where φ_{rs} is the projection on X of the geodesic path from r to s.
4. Let p be a prime number. For any $f \in S$,
$$\begin{aligned} I\big(\mathrm{T}_p\{r,s\}, f(z)\,dz\big) &= I\big(\{r,s\}, (\mathrm{T}_p f)(z)\,dz\big) = \int_r^s \sum_{\alpha\in\Delta(p)} f(\alpha z)\,d(\alpha z) \\ &= \sum_{\alpha\in\Delta(p)} \int_{\alpha r}^{\alpha s} f(z)\,dz = I\Big(\sum_{\alpha\in\Delta(p)} \alpha\{r,s\}, f(z)\,dz\Big). \end{aligned}$$
Therefore,
$$\mathrm{T}_p\{r,s\} = \sum_{\alpha\in\Delta(p)} \alpha\{r,s\}.$$

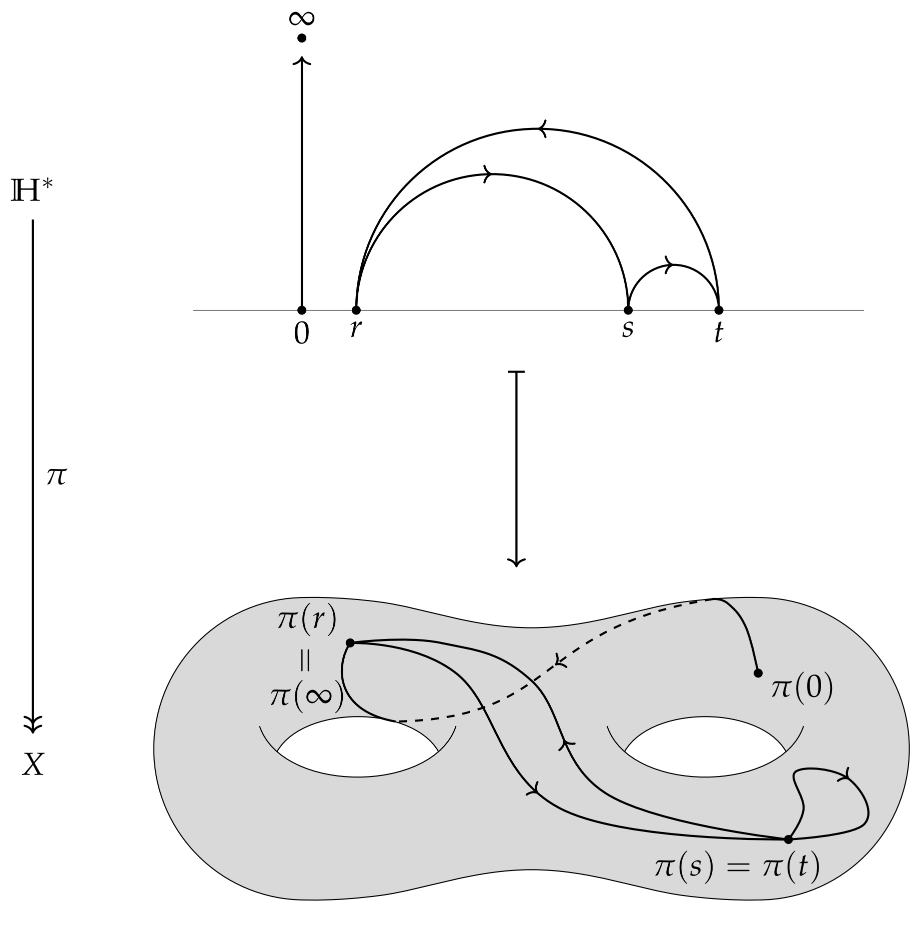

Figure 4: Paths representing the modular symbols $\{0,\infty\}$, $\{r,s\}$, $\{s,t\}$ and $\{t,r\}$.

We can deduce some easy properties from the geometric interpretation of modular symbols.

Proposition 11. *For every $r, s, t \in \mathbb{P}^1(\mathbb{Q})$ and $\gamma \in \Gamma$,*

(i) $\{r, r\} = 0$,

(ii) $\{r, s\} + \{s, r\} = 0$,

(iii) $\{r, s\} + \{s, t\} + \{t, r\} = 0$ *and*

(iv) $\gamma\{r, s\} = \{r, s\}$.

Proof.

(i) The path that goes from r to r is trivial.

(ii) The path that goes from r to s and back to r is homotopic to a trivial path.

(iii) The path that goes from r to s, then to t and finally back to r is the boundary of the triangle $\triangle\langle r, s, t\rangle$.

(iv) Γ acts trivially on $X = \Gamma \setminus \mathbb{H}^*$. □

From now on, write $\overline{}$ for the image of $-$ in $\mathrm{PSL}_2(\mathbb{R})$. (Since we consider only matrices in $\mathrm{SL}_2(\mathbb{Z})$, that means taking the quotient by $\{\pm 1\}$.)

Lemma 12. *The map*

$$\begin{aligned} \overline{\Gamma} &\longrightarrow \mathrm{H}_1(X; \mathbb{Z}) \\ \overline{\gamma} &\longmapsto \{r, \gamma r\} \end{aligned}$$

is a well-defined (i.e., independent of $r \in \mathbb{P}^1(\mathbb{Q})$) surjective morphism.

Idea of the proof. Playing a bit with properties (i) to (iv) in proposition 11, one checks that this map is independent of r and a group homomorphism.

For the surjectivity, the key observation is that $\mathbb{H}$ is simply connected. Therefore, $\pi\colon \mathbb{H}^* \twoheadrightarrow \overline{\Gamma} \setminus \mathbb{H}^* = X$ is *essentially* a universal covering (except for some ramification points) and $\overline{\Gamma}$ is *essentially* the fundamental group of X (minus some points). The result follows by Hurewicz's theorem. □

The previous lemma shows that we do not need so many modular symbols to generate the homology group of X. This leads us to focus only on modular symbols of a special form.

Definition 13. A *distinguished modular symbol* is a modular symbol of the form

$$\alpha\{0, \infty\} = \left\{\frac{b}{d}, \frac{a}{c}\right\} \quad \text{for some } \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}).$$

Theorem 14 (Manin's trick). *Consider a (right) coset decomposition*

$$\mathrm{PSL}_2(\mathbb{Z}) = \bigsqcup_{i=1}^{m} \overline{\Gamma}\overline{\alpha_i}.$$

Every class $x \in \mathrm{H}_1(X;\mathbb{Z})$ can be represented as

$$x = \sum_{i=1}^{m} \lambda_i \cdot \alpha_i\{0,\infty\}$$

with $\lambda_i \in \mathbb{Z}$ for $1 \le i \le m$ and

$$\partial x = \sum_{i=1}^{m} \lambda_i \big[\{\pi(\alpha_i\infty)\} - \{\pi(\alpha_i 0)\}\big] = 0 \quad \text{in } \mathrm{Div}^0(X).$$

(Here, $\{P\}$ denotes the (prime) divisor associated with $P \in X$.)

Idea of the proof. By lemma 12, we can restrict to the case $x = \{0,\gamma 0\} = \left\{0, \frac{p}{q}\right\}$ for some $\gamma \in \Gamma$. Expand as a continued fraction

$$\frac{p}{q} = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{\ddots + \cfrac{1}{a_n}}}}$$

and consider the successive convergents

$$\frac{p_{-2}}{q_{-2}} = \frac{0}{1}, \quad \frac{p_{-1}}{q_{-1}} = \frac{1}{0}, \quad \frac{p_0}{q_0} = \frac{a_0}{1}, \quad \frac{p_1}{q_1} = a_0 + \frac{1}{a_1}, \quad \ldots, \quad \frac{p_n}{q_n} = \frac{p}{q}$$

(all of them written in lowest terms and the first two included formally). One checks that

$$\left\{0, \frac{p}{q}\right\} = \sum_{i=-1}^{n} \left\{\frac{p_{i-1}}{q_{i-1}}, \frac{p_i}{q_i}\right\} = \sum_{i=-1}^{n} \begin{pmatrix} (-1)^{i-1}p_i & p_{i-1} \\ (-1)^{i-1}q_i & q_{i-1} \end{pmatrix} \{0,\infty\}.$$

□

With this, we already have quite simple generators of a Hecke module isomorphic to S. We can still improve it a bit (i.e., find the relations).

Let

$$\sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \tau = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix},$$

so that $\mathrm{PSL}_2(\mathbb{Z}) = \langle \overline{\sigma}, \overline{\tau} : \overline{\sigma}^2 = \overline{\tau}^3 = 1 \rangle$. It suffices to understand how these matrices act on distinguished modular symbols. We have

$$\sigma\{0,\infty\} = \{\infty, 0\} \implies \alpha\{0,\infty\} + \alpha\sigma\{0,\infty\} = 0$$

and

$$\left.\begin{aligned}\tau\{0,\infty\} &= \{\infty,1\}\\ \tau\{\infty,1\} &= \{1,0\}\end{aligned}\right\} \implies \alpha\{0,\infty\} + \alpha\tau\{0,\infty\} + \alpha\tau^2\{0,\infty\} = 0$$

for all $\alpha \in \mathrm{SL}_2(\mathbb{Z})$. In some sense, these are all the relations we need.

Definition 15. Consider a (right) coset decomposition

$$\mathrm{PSL}_2(\mathbb{Z}) = \bigsqcup_{i=1}^{m} \overline{\Gamma}\overline{\alpha_i}.$$

The *Manin symbols* are formal symbols $(\alpha_i) = (\overline{\Gamma}\overline{\alpha_i})$ for $1 \le i \le m$ (one symbol for each right coset). There is a right action of $\mathrm{SL}_2(\mathbb{Z})$ on Manin symbols given by $(\alpha)\beta = (\alpha\beta)$.

Remark. One should think of the Manin symbol (α_i) as representing the distinguished modular symbol $\alpha_i\{0,\infty\}$.

Definition 16.

(1) The group C of *Manin chains* is the free abelian group generated by Manin symbols modulo the relations $(\alpha) + (\alpha)\sigma = 0$ and modulo any torsion (i.e., if $(\alpha) = (\alpha)\sigma$, then $(\alpha) = 0$ in C).

(2) The *Manin boundary map* $\partial\colon C \to \mathrm{Div}^0(X)$ is given by

$$\partial(\alpha) = \{\pi(\alpha\infty)\} - \{\pi(\alpha 0)\}.$$

The group of *Manin cycles* is $Z = \mathrm{Ker}(\partial)$.

(3) The group of *Manin boundaries* is the subgroup B of C generated by the elements $(\alpha) + (\alpha)\tau + (\alpha)\tau^2$ and the elements (α) such that $(\alpha) = (\alpha)\tau$.

Recall that the first homology group $\mathrm{H}_1(X;\mathbb{Z})$ is defined as a quotient of (singular) cycles modulo boundaries. We can finally state the last result which yields a presentation of $\mathrm{H}_1(X;\mathbb{Z})$.

Theorem 17 (Manin). *The map*

$$\xi\colon Z\,/\,B \longrightarrow \mathrm{H}_1(X;\mathbb{Z})$$

defined by

$$\xi\Big(\sum_{i=1}^{m} \lambda_i(\alpha_i) \mod B\Big) = \sum_{i=1}^{m} \lambda_i \cdot \alpha_i\{0,\infty\}$$

is an isomorphism of groups.

Idea of the proof. One checks easily that B is a subgroup of Z and that ξ is well-defined from the relations given by σ and τ. Surjectivity follows from theorem 14.

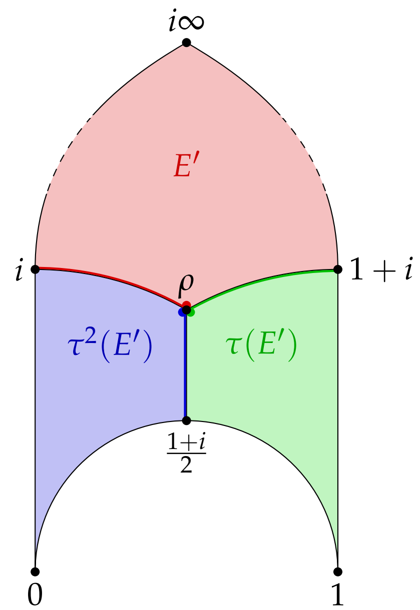

Figure 5: The basic triangle E on $\mathbb{H}^*$, used to define a cell complex for X.

The difficult part is to prove the injectivity of ξ. To do so, we describe $\mathrm{H}_1(X;\mathbb{Z})$ via a cell complex (i.e., using cellular homology).

Consider the following regions of $\mathbb{H}^*$, depicted in figure 5:

- E is the interior of the triangle $\triangle\langle 0,1,\infty\rangle$ (i.e., all the coloured area in figure 5), and
- E' is the interior of the quadrilateral $\square\langle i,\rho,1+i,\infty\rangle$ plus the line segment $\langle i,\rho\rangle$ minus its endpoint $\{\,i\,\}$ (i.e., the red area in figure 5).

Observe that $E = E' \cup \tau E' \cup \tau^2 E'$ and $E' \cap \tau E' = E' \cap \tau^2 E' = \{\,\rho\,\}$. In addition, E' is a fundamental domain for the action of $\mathrm{PSL}_2(\mathbb{Z})$ on $\mathbb{H}^*$. Indeed, $\overline{E'}$ corresponds to the right half plus a translate of the left half of the fundamental domain $\mathcal{F}$ (see figure 1). Pairs of sides of this quadrilateral are identified under the action of $\mathrm{PSL}_2(\mathbb{Z})$, but there are no identifications inside any one side. Therefore, projection by π embeds each half-side and each half-median of the triangle E in X.

Having made this key observations, we construct the following 2–dimensional cell complex (naturally homeomorphic to X).

(i) The 0–cells are the images under π of the vertices and midpoints of the sides of the triangles αE for $\overline{\alpha} \in \mathrm{PSL}_2(\mathbb{Z})$.

(ii) The 1–cells are the images under π of the half-sides, oriented from vertex to midpoint, of the triangles αE for $\overline{\alpha} \in \mathrm{PSL}_2(\mathbb{Z})$. That is, such a cell is $e_1\big((\alpha)\big) = \pi\big(\langle \alpha\infty, \alpha i\rangle\big)$.

(iii) The 2–cells are the images under π of the triangles αE for $\overline{\alpha} \in \mathrm{PSL}_2(\mathbb{Z})$. These can be of two types:

- if $(\alpha) = (\alpha)\tau$, we get a cell $e_2\big((\alpha)\big) = \pi(\alpha E')$ with boundary
$$\partial e_2\big((\alpha)\big) = e_1\big((\alpha)\big) - e_1\big((\alpha)\sigma\big),$$
and
- if $(\alpha) \neq (\alpha)\tau$, we get a cell $e_2\big((\alpha)\big) = \pi(\alpha E)$ with boundary
$$\partial e_2\big((\alpha)\big) = \sum_{j=0}^{2}\big[e_1\big((\alpha)\tau^j\big) - e_1\big((\alpha)\tau^j\sigma\big)\big].$$

One can check that the map $(\alpha) \mapsto e_1\big((\alpha)\sigma\big) - e_1\big((\alpha)\big)$ induces an isomorphism
$$Z \,/\, B \xrightarrow{\cong} Z_1^{\text{cell}}(X) \,/\, B_1^{\text{cell}}(X) \cong \mathrm{H}_1(X;\mathbb{Z}). \qquad \square$$

References

[1] Koblitz, N. *Introduction to elliptic curves and modular forms*. 2nd ed. Graduate texts in mathematics 97. New York, NY, USA: Springer–Verlag, 1993. Chap. III, pp. 98–175.

[2] Lang, S. *Introduction to modular forms*. Grundlehren der mathematischen Wissenschaften 222. Berlin, Germany: Springer–Verlag, 1995. Chap. IV, pp. 57–67. Corrected reprint of the 1976 original.

[3] Manin, Y. I. 'Parabolic points and zeta–functions of modular curves'. Trans. by Koblitz, N. In: *Math. USSR-Izv.* 6.1 (1972), pp. 19–64.

[4] Milne, J. S. *Modular functions and modular forms (v1.31)*. Course notes. 2012. Chap. I, pp. 13–85. URL: `http://www.jmilne.org/math/CourseNotes/mf.html` (visited on 11/05/2019).

[5] Serre, J.-P. *A course in arithmetic*. Graduate texts in mathematics 7. New York, NY, USA: Springer–Verlag, 1973. Chap. VII, pp. 77–111.

[6] Stein, W. A. *Modular forms, a computational approach*. Graduate studies in mathematics 79. Providence, RI, USA: American Mathematical Society, 2007. Chap. 3, pp. 35–61. URL: `http://wstein.org/books/modform/` (visited on 11/05/2019).