3

Modular Symbols

These are 10/01/03 notes for Math 252 by William Stein.

This chapter is about how to explicitly compute the homology of modular curves using modular symbols.

We assume the reader is familiar with basic notions of algebraic topology, including homology groups of surfaces and triangulation. We also assume that the reader has read XXX about the fundamental domain for the action of $\mathrm{PSL}_2(\mathbf{Z})$ on the upper half plane, and XXX about the construction of modular curves.

Some standard references for modular symbols are [4] [3, IV], [1], and [5]. Sections 3.1–3.2 below very closely follow Section 1 of Manin's paper [4].

For the rest of this chapter, let $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$ and let G be a subgroup of Γ of finite index. Note that we do not require G to be a congruence subgroup. The quotient $X(G) = G\backslash\mathfrak{h}^*$ of $\mathfrak{h}^* = \mathfrak{h} \cup \mathbf{P}^1(\mathbf{Q})$ by G has an induced structure of compact Riemann surface. Let $\pi : \mathfrak{h}^* \to X(G)$ denote the natural projection. The matrices

$$s = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad t = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$$

together generate Γ; they have orders 2 and 3, respectively.

3.1 Modular symbols

Let $\mathrm{H}^0(X(G), \Omega^1)$ denote the complex vector space of holomorphic 1-forms on $X(G)$. Integration of differentials along homology classes defines a perfect pairing

$$\mathrm{H}_1(X(G), \mathbf{R}) \times \mathrm{H}^0(X(G), \Omega^1) \to \mathbf{C},$$

hence an isomorphism

$$\mathrm{H}_1(X(G), \mathbf{R}) \cong \mathrm{Hom}_{\mathbf{C}}(\mathrm{H}^0(X(G), \Omega^1), \mathbf{C}).$$

For more details, see [3, §IV.1].

Given two elements $\alpha, \beta \in \mathfrak{h}^*$, integration from α to β induces a well-defined element of $\mathrm{Hom}_{\mathbf{C}}(\mathrm{H}^0(X(G), \Omega^1), \mathbf{C})$, hence an element

$$\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{R}).$$

Definition 3.1.1 (Modular symbol). The homology class $\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{R})$ associated to $\alpha, \beta \in \mathfrak{h}^*$ is called the *modular symbol* attached to α and β.

Proposition 3.1.2. *The symbols $\{\alpha, \beta\}$ have the following properties:*

1. $\{\alpha, \alpha\} = 0$, $\{\alpha, \beta\} = -\{\beta, \alpha\}$, *and* $\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0$.
2. $\{g(\alpha), g(\beta)\} = \{\alpha, \beta\}$ *for all* $g \in G$
3. *If $X(G)$ has nonzero genus, then $\{\alpha, \beta\} \in \mathrm{H}_1(X(G), \mathbf{Z})$ if and only if $G(\alpha) = G(\beta)$ (i.e., the cusps α and β are equivalent).*[1] 1

Remark 3.1.3. We only have $\{\alpha, \beta\} = \{\beta, \alpha\}$ if $\{\alpha, \beta\} = 0$, so the modular symbols notation, which suggests "unordered pairs", is actively misleading.

Proposition 3.1.4. *For any $\alpha \in \mathfrak{h}^*$, the map $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ that sends g to $\{\alpha, g\alpha\}$ is a surjective group homomorphism that does not depend on the choice of α.*

Proof. If $g, h \in G$ and $\alpha \in \mathfrak{h}^*$, then

$$\{\alpha, gh(\alpha)\} = \{\alpha, g\alpha\} + \{g\alpha, gh\alpha\} = \{\alpha, g\alpha\} + \{\alpha, h\alpha\},$$

so the map is a group homomorphism. To see that the map does not depend on the choice of α, suppose $\beta \in \mathfrak{h}^*$. By Proposition 3.1.2, we have $\{\alpha, \beta\} = \{g\alpha, g\beta\}$. Thus

$$\{\alpha, g\alpha\} + \{g\alpha, \beta\} = \{g\alpha, \beta\} + \{\beta, g\beta\},$$

so cancelling $\{g\alpha, \beta\}$ from both sides proves the claim.

The fact that the map is surjective follows from general facts from algebraic topology. Let $\mathfrak{h}^0$ be the complement of $\Gamma i \cup \Gamma\rho$ in $\mathfrak{h}$, fix $\alpha \in \mathfrak{h}^0$, and let $X(G)^0 = \pi(\mathfrak{h}^0)$. The map $\mathfrak{h}^0 \to X(G)^0$ is an unramified covering of (noncompact) Riemann surfaces with automorphism group G. Thus α determines a group homomorphism $\pi_1(X(G)^0, \pi(\alpha)) \to G$. When composed with the morphism $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ above, the composition

$$\pi_1(X(G)^0, \pi(\alpha)) \to G \to \mathrm{H}_1(X(G), \mathbf{Z})$$

is the canonical map from the fundamental group of $X(G)^0$ to the homology of the corresponding compact surface, which is surjective. This forces the map $G \to \mathrm{H}_1(X(G), \mathbf{Z})$ to be surjective, which proves the claim. □

3.2 Manin symbols

We continue to assume that G is a finite-index subgroup of $\Gamma = \mathrm{PSL}_2(\mathbf{Z})$, so the set $G\backslash\Gamma = \{Gg_1, \ldots Gg_d\}$ of right cosets of G in Γ is finite.

[1] Say more about this one.

3.2.1 Using continued fractions to obtain surjectivity

Let $R = G\backslash\Gamma$ be the set of right cosets of G in Γ. Define

$$[\,] : R \to \mathrm{H}_1(X(G), \mathbf{R})$$

by $[r] = \{r0, r\infty\}$, where $r0$ means the image of 0 under any element of the coset r (it doesn't matter which). For $g \in \Gamma$, we also write $[g] = [gG]$.

Proposition 3.2.1. *Any element of* $\mathrm{H}_1(X(G), \mathbf{Z})$ *is a sum of elements of the form* $[r]$*, and the representation* $\sum n_r\{\alpha_r, \beta_r\}$ *of* $h \in \mathrm{H}_1(X(G), \mathbf{Z})$ *can be chosen so that* $\sum n_r(\pi(\beta_r) - \pi(\alpha_r)) = 0 \in \mathrm{Div}(X(G))$.

Proof. By Proposition 3.1.4, every element h of $\mathrm{H}_1(X(G), \mathbf{Z})$ is of the form $\{0, g(0)\}$ for some $g \in \mathbf{G}$. If $g(0) = \infty$, then $h = [G]$ and $\pi(\infty) = \pi(0)$, so we may assume $g(0) = a/b \neq \infty$, with a/b in lowest terms and $b > 0$. Also assume $a > 0$, since the case $a < 0$ is treated in the same way. Let

$$0 = \frac{p_{-2}}{q_{-2}} = \frac{0}{1},\ \frac{p_{-1}}{q_{-1}} = \frac{1}{0},\ \frac{p_0}{1} = \frac{p_0}{q_0},\ \frac{p_1}{q_1},\ \frac{p_2}{q_2}, \ldots, \frac{p_n}{q_n} = \frac{a}{b}$$

denote the continued fraction convergents of the rational number a/b. Then

$$p_j q_{j-1} - p_{j-1} q_j = (-1)^{j-1} \qquad \text{for } -1 \leq j \leq n.$$

If we let $g_j = \begin{pmatrix} (-1)^{j-1} p_j & p_{j-1} \\ (-1)^{j-1} q_j & q_{j-1} \end{pmatrix}$, then $g_j \in \mathrm{SL}_2(\mathbf{Z})$ and

$$\begin{aligned} \left\{0, \frac{a}{b}\right\} &= \sum_{j=-1}^{n} \left\{\frac{p_{j-1}}{q_{j-1}}, \frac{p_j}{q_j}\right\} \\ &= \sum_{j=-1}^{n} \{g_j 0, g_j \infty\}) \\ &= \sum_{j=-1}^{r} [g_j]. \end{aligned}$$

For the assertion about the divisor sum equaling zero, notice that the endpoints of the successive modular symbols cancel out, leaving the difference of 0 and $g(0)$ in the divisor group, which is 0. □

Lemma 3.2.2. *If* $x = \sum_{j=1}^{t} n_j\{\alpha_j, \beta_j\}$ *is a* $\mathbf{Z}$*-linear combination of modular symbols for* G *and* $\sum n_j(\pi(\beta_j) - \pi(\alpha_j)) = 0 \in \mathrm{Div}(X(G))$*, then* $x \in \mathrm{H}^1(X(G), \mathbf{Z})$.

Proof. We may assume that each n_j is ± 1 by allowing duplication. We may further assume that each $n_j = 1$ by using that $\{\alpha, \beta\} = -\{\beta, \alpha\}$. Next reorder the sum so $\pi(\beta_j) = \pi(\alpha_{j+1})$ by using that the divisor is 0, so every β_j must be equivalent to some $\alpha_{j'}$, etc. The lemma should now be clear. □

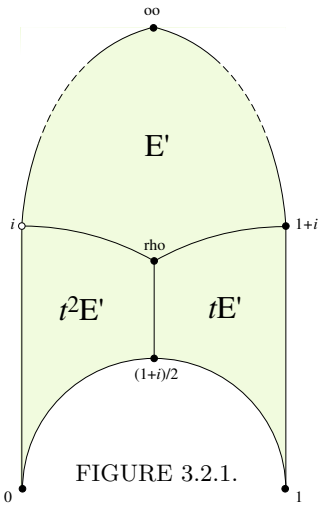

FIGURE 3.2.1.

3.2.2 Triangulating $X(G)$ to obtain injectivity

Let C be the abelian group generated by symbols (r) for $r \in G\backslash\Gamma$, subject to the relations

$$(r) + (rs) = 0, \qquad \text{and } (r) = 0 \quad if \quad r = rs.$$

For $(r) \in C$, define the boundary of (r) to be the difference $\pi(r\infty) - \pi(r0) \in \mathrm{Div}(X(G))$. Since s swaps 0 and ∞, the boundary map is a well-defined map on C. Let Z be its kernel.

Let B be the subgroup of C generated by symbols (r), for all $r \in G\backslash\Gamma$ that satisfy $r = rt$, and by $(r) + (rt) + (rt^2)$ for all other r. If $r = rt$, then $rt(0) = r(0)$, so $r(\infty) = r(0)$, so $(r) \in Z$. Also, using (3.2.1) below, we see that for any r, the element $(r) + (rt) + (rt^2)$ lies in Z.

The map $G\backslash\Gamma \to \mathrm{H}_1(X(G), \mathbf{R})$ that sends (r) to $[r]$ induces a homomorphism $C \to \mathrm{H}_1(X(G), \mathbf{R})$, so by Proposition 3.2.1 we obtain a surjective homomorphism

$$\psi : Z/B \to \mathrm{H}_1(X(G), \mathbf{Z}).$$

Theorem 3.2.3 (Manin). *The map $\psi : Z/B \to \mathrm{H}_1(X(G), \mathbf{Z})$ is an isomorphism.*

Proof. We only have to prove that ψ is injective. Our proof follows the proof of [4, Thm. 1.9] very closely. We compute the homology $\mathrm{H}_1(X(G), \mathbf{Z})$ by triangulating $X(G)$ to obtain a simplicial complex L with homology Z_1/B_1, then embed Z/B in the homology Z_1/B_1 of $X(G)$. Most of our work is spent describing the triangulation L.

Let E denote the *interior* of the triangle with vertices 0, 1, and ∞, as illustrated in Figure 3.2.1. Let E' denote the union of the interior of the region bounded by the path from i to $\rho = e^{\pi i/3}$ to $1 + i$ to ∞ with the indicated path from i to ρ, not including the vertex i.

When reading the proof below, it will be helpful to look at the following table, which illustrates what $s = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, $t = \left(\begin{smallmatrix} 1 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, and t^2 do to the vertices in

Figure 3.2.1:

1	0	1	∞	i	$1+i$	$(1+i)/2$	ρ
s	∞	-1	0	i	$(-1+i)/2$	$-1+i$	$-\overline{\rho}$
t	∞	0	1	$1+i$	$(1+i)/2$	i	ρ
t^2	1	∞	0	$(1+i)/2$	i	$1+i$	ρ

(3.2.1)

Note that each of E', tE', and t^2E' is a fundamental domain for Γ, in the sense that every element of the upper half plane is conjugate to exactly one element in the closure of E' (except for identifications along the boundaries). For example, E' is obtained from the standard fundamental domain for Γ, which has vertices ρ^2, ρ, and ∞, by chopping it in half along the imaginary axis, and translating the piece on the left side horizontally by 1.

If $(0,\infty)$ is the path from 0 to ∞, then $t(0,\infty) = (\infty, 1)$ and $t^2(0,\infty) = (1,0)$. Also, $s(0,\infty) = (\infty, 0)$. Thus each half side of E is Γ-conjugate to the side from i to ∞. Also, each 1-simplex in Figure 3.2.1, i.e., the sides that connected two adjacent labeled vertices such as i and ρ, maps homeomorphically into $X(\Gamma)$. This is clear for the half sides, since they are conjugate to a path in the interior of the standard fundamental domain for Γ, and for the medians (lines from midpoints to ρ) since the path from i to ρ is on an edge of the standard fundamental domain with no self identifications.

We now describe our triangulation L of $X(G)$:

0-cells The 0 cells are the cusps $\pi(\mathbf{P}^1(\mathbf{Q}))$ and i-elliptic points $\pi(\Gamma i)$. Note that these are the images under π of the vertices and midpoints of sides of the triangles gE, for all $g \in \Gamma$.

1-cells The 1 cells are the images of the half-sides of the triangles gE, for $g \in \Gamma$, oriented from the edge to the midpoint (i.e., from the cusp to the i-elliptic point). For example, if $r = Gg$ is a right coset, then

$$e_1(r) = \pi(g(\infty), g(i)) \in X(G)$$

is a 1 cell in L. Since, as we observed above, every half side is Γ-conjugate to $e_1(G)$, it follows that every 1-cell is of the form $e(r)$ for some right coset $r \in G\backslash\Gamma$.

Next observe that if $r \neq r'$ then

$$e_1(r) = e_1(r') \qquad \text{implies} \qquad r' = rs. \tag{3.2.2}$$

Indeed, if $\pi(g(\infty), g(i)) = \pi(g'(\infty), g'(i))$, then $ri = r'i$ (note that the endpoints of a path are part of the definition of the path). Thus there exists $h, h' \in G$ such that $hg(i) = h'g'(i)$. Since the only nontrivial element of Γ that stabilizes i is s, this implies that $(hg)^{-1}h'g' = s$. Thus $h'g' = hgs$, so $Gg' = Ggs$, so $r' = rs$.

2-cells There are two types of 2-cells, those with 2 sides and those with 3.

2-sided: The 2-sided 2-cells $e_2(r)$ are indexed by the cosets $r = Gg$ such that $rt = r$. Note that for such an r, we have $\pi(rE') = \pi(rtE') = \pi(rt^2E')$. The 2-cell $e_2(r)$ is $\pi(gE')$. The image $g(\rho, i)$ of the half median maps to a

line from the center of $e_2(r)$ to the edge $\pi(g(i)) = \pi(g(1+i))$. Orient $e_2(r)$ in a way compatible with the e_1. Since $Ggt = Gg$,

$$\pi(g(1+i), g(\infty)) = \pi(gt^2(1+i), gt^2(\infty)) = \pi(g(i), g(0)) = \pi(gs(i), gs(\infty)),$$

so

$$e_1(r) - e_1(rs) = \pi(g(\infty), g(i)) + \pi(gs(i), gs(\infty)) = \pi(g(\infty), g(i)) + \pi(g(1+i), g(\infty)).$$

Thus

$$\partial e_2(r) = e_1(r) - e_1(rs).$$

Finally, note that if $r' \neq r$ also satisfies $r't = r'$, then $e_2(r) \neq e_2(r')$ (to see this use that E' is a fundamental domain for Γ).

3-sided: The 3-sided 2-cells $e_2(r)$ are indexed by the cosets $r = Gg$ such that $rt \neq r$. Note that for such an r, the three triangles rE', rtE', and rt^2E' are distinct (since they are nontrivial translates of a fundamental domain). Orient $e_2(r)$ in a way compatible with the e_1 (so edges go from cusps to midpoints). Then

$$\partial e_2(r) = \sum_{n=0}^{2} \left(e_1(rt^n) - e_1(rt^n s)\right).$$

We have now defined a complex L that is a triangulation of $X(G)$. Let C_1, Z_1, and B_1 be the group of 1-chains, 1-cycles, and 1-boundaries of the complex L. Thus C_1 is the abelian group generated by the paths $e_1(r)$, the subgroup Z_1 is the kernel of the map that sends $e_1(r) = \pi(r(\infty), r(0))$ to $\pi(r(0)) - \pi((\infty))$, and B_1 is the subgroup of Z_1 generated by boundaries of 2-cycles.

Let C, Z, B be as defined before the statement of the Theorem 3.2.3. We have $\mathrm{H}_1(X(G), \mathbf{Z}) \cong Z_1/B_1$, and would like to prove that $Z/B \cong Z_1/B_1$.

Define a map $\varphi : C \to C_1$ by $(r) \mapsto e_1(rs) - e_1(r)$. The map φ is well defined because if $r = rs$, then clearly $(r) \mapsto 0$, and $(r) + (rs)$ maps to $e_1(rs) - e_1(r) + e_1(r) - e_1(rs) = 0$. To see that f is injective, suppose $\sum n_r(r) \neq 0$. Since in C we have the relations $(r) = -(rs)$ and $(r) = 0$ if $rs = r$, we may assume that $n_r n_{rs} = 0$ for all r. We have

$$\varphi\left(\sum n_r(r)\right) = \sum n_r(e_1(rs) - e_1(r)).$$

If $n_r \neq 0$ then $r \neq rs$, so (3.2.2) implies that $e_1(r) \neq e_1(rs)$. If $n_r \neq 0$ and $n_{r'} \neq 0$ with $r' \neq r$, then $r \neq rs$ and $r' \neq r's$, so $e_1(r), e_1(rs), e_1(r'), e_1(r's)$ are all distinct. We conclude that $\sum n_r(e_1(rs) - e_1(r)) \neq 0$, which proves that φ is injective.

Suppose $(r) \in C$. Then

$$\varphi(r) + B_1 = \psi(r) = \{r(0), r(\infty)\} \in \mathrm{H}_1(X(G), \mathbf{Z}) = C_1/B_1,$$

since

$$\varphi(r) = e_1(rs) - e_1(r) = \pi(rs(\infty), rs(i)) - \pi(r(\infty), r(i)) = \pi(r(0), r(i)) - \pi(r(\infty), r(i))$$

belongs to the homology class $\{r(0), r(\infty)\}$. Extending linearly, we have, for any $z \in C$, that $\varphi(z) + B_1 = \psi(z)$.

The generators for B_1 are the boundaries of 2-cells $e_2(r)$. As we saw above, these have the form $\varphi(r)$ for all r such that $r = rt$, and $\varphi(r) + \varphi(rt) + \varphi(rt^2)$ for the r such that $rt \neq r$. Thus $B_1 = \varphi(B) \subset \varphi(Z)$, so the map φ induces an injection $Z/B \hookrightarrow Z_1/B_1$. This completes the proof of the theorem.

□

References

[1] J. E. Cremona, *Algorithms for modular elliptic curves*, second ed., Cambridge University Press, Cambridge, 1997.

[2] F. Diamond and J. Im, *Modular forms and modular curves*, Seminar on Fermat's Last Theorem, Providence, RI, 1995, pp. 39–133.

[3] S. Lang, *Introduction to modular forms*, Springer-Verlag, Berlin, 1995, With appendixes by D. Zagier and W. Feit, Corrected reprint of the 1976 original.

[4] J. I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66. MR 47 #3396

[5] L. Merel, *Universal Fourier expansions of modular forms*, On Artin's conjecture for odd 2-dimensional representations, Springer, 1994, pp. 59–94.

[6] J-P. Serre, *A Course in Arithmetic*, Springer-Verlag, New York, 1973, Translated from the French, Graduate Texts in Mathematics, No. 7.

[7] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kan Memorial Lectures, 1.