Modular Symbols.

Kevin Buzzard

April 26, 2012

Abstract

We summarise various people's work/writings on modular symbols.

V1 (Cremona, Merel) written 7/8/07

1 Cremona's book.

Cremona explains modular symbols in weight 2 from a computational approach, in the introduction to his book of tables. The idea is that weight 2 cusp forms of level N (as a complex vector space) are isomorphic to singular cohomology of the compact Riemann surface $X_1(N)$ (with real coefficients) and hence are dual to the singular homology of $X_1(N)$. Modular Symbols provide a way of computing this singular homology in the sense that they give explicit $\mathbf{Q}$-vector spaces Z and B with $Z/B = H_1(X_1(N), \mathbf{Q})$.

Here's a summary of the explanation given in Cremona's book of tables. Let Γ be a congruence subgroup and let X_Γ be the associated compactified modular curve.

The first observation is that if $\alpha \in \mathbf{P}^1(\mathbf{Q})$ and $g \in \Gamma$ then a path in the upper half plane linking α to $g.\alpha$ gives rise to a loop in X_Γ, and hence a homology class $\{\alpha, g.\alpha\} \in H_1(X_\Gamma, \mathbf{Z})$. Cremona says that

(a) The class only depends on g (easy proof: a path $p(t)$ from α to β in the extended upper half plane gives us a homotopy from $\{p(0), g.p(0)\}$ to $\{p(1), g.p(1)\}$.)

(b) The resulting map $\Gamma \to H_1(X_\Gamma, \mathbf{Z})$ is surjective (this is not clear, as far as I can see, but will become clear when we think about things in terms of group cohomology).

If $\alpha \in \mathbf{P}^1(\mathbf{Q})$ and $g, h \in \Gamma$ then $\{\alpha, gh\alpha\} = \{\alpha, g\alpha\} + \{\alpha, h\alpha\}$ (this is clear) and hence if we fix α then we actually get a *group homomorphism* $\Gamma \to H_1(X_\Gamma, \mathbf{Z})$ which is independent of α (this is clear) and surjective (this is not so clear). The kernel contains all the elliptic elements (because they have finite order and homology of a compact Riemann surface is torsion-free) and also all the parabolic elements (because they fix a cusp and the resulting loop is then zero). In fact the kernel is the normal subgroup generated by these elements (this will also become clear later on).

To give another take on what is going on here we recall Shimura's definition of parabolic group cohomology on p224 of his book, specialised to the case where his coefficient module is $\mathbf{Z}$ with the trivial action. Then $C^1(\Gamma, \mathbf{Z})$ is just all the (set-theoretic) maps $\Gamma \to \mathbf{Z}$, $Z^1(\Gamma, \mathbf{Z})$ are those which happen to be group homomorphisms (note that Γ is finitely-generated and hence $Z^1(\Gamma, \mathbf{Z})$ is a finitely-generated free abelian group), $B^1 := B^1(\Gamma, \mathbf{Z}) = 0$, $Z^1_P := Z^1_P(\Gamma, \mathbf{Z})$ are the group homomorphisms in $Z^1(\Gamma, \mathbf{Z})$ which vanish on the set of parabolic elements, and $H^1_P(\Gamma, \mathbf{Z})$ is by definition Z^1_P/B^1 and so it's just the group homorphisms $\Gamma \to \mathbf{Z}$ vanishing on the parabolic elements. This cohomology group is presumably isomorphic to $H^1(X_\Gamma, \mathbf{Z})$ and hence the construction above is giving us a map from Γ to the space of linear maps $Z^1_P \to \mathbf{Z}$ and this map is surely just evaluation, something which explains all of Cremona's assertions above.

A mild generalisation: if $\alpha, \beta \in \mathbf{P}^1(\mathbf{Q})$, not necessarily in the same Γ-equivalence class, then one can integrate a weight 2 cusp form from α to β and get a complex number, and in this way we get an element of the complex dual of the space of weight 2 cusp forms and hence an element $\{\alpha, \beta\}$ in $H_1(X_\Gamma, \mathbf{R})$. Richard Thomas points out to me that this construction is not topological; it relies on the complex structure on X_Γ. The Manin-Drinfeld theorem says that in fact $\{\alpha, \beta\} \in H_1(X_\Gamma, \mathbf{Q})$ (note that I think that this is the first time that we assume that Γ is a congruence subgroup, rather than just some finite index subgroup of $\mathrm{SL}_2(\mathbf{Z})$!). The proof of this is combinatorial and slightly delicate; I read the idea in a Seminaire Bourbaki article by Mazur (1972). Of particular interest is the symbol $\{0, \infty\}$, which is related to L-values.

We triangulate the extended upper-half plane by the line from 0 to $i\infty$ and the images of this line under $\mathrm{SL}_2(\mathbf{Z})$. Note that any two elements of $\mathbf{P}^1(\mathbf{Q})$ are connected by a finite union of such lines, by the continued fraction algorithm. This triangulation contains the "basic triangle" T_0 with vertices at 0, 1 and ∞; every triangle is the image of T_0 under some element of $\mathrm{SL}_2(\mathbf{Z})$, and this element is unique modulo a group of order 3 in $\mathrm{PSL}_2(\mathbf{Z})$ generated by TS (where T and S are the usual suspects; the element TS has order 3 and permutes the vertices of T_0).

If (γ) is the image of the oriented path $\gamma.\{0, \infty\}$ in X_Γ then one checks easily that $(\gamma) + (\gamma.TS) + (\gamma.TSTS) = 0$ (TS is that element of order 3, so we have a triangle) and $(\gamma) + (\gamma.S) = 0$ (S reflects the line) in $H_1(X_\Gamma, \mathbf{Q})$ and $(\delta.\gamma) = (\delta)$ if $\delta \in \Gamma$.

The fact that $(\delta.\gamma) = (\gamma)$ means that there are naturally e:=(index of Γ) of these paths to think about, so we get a subspace of $H_1(X_\Gamma, \mathbf{Q})$ spanned by these e paths, and arguments given already show that it's a surjection. In fact let C denote the e-dimensional vector space over $\mathbf{Q}$ spanned by these e formal symbols; we get a linear map $C \to H_1(X_\Gamma, \mathbf{Q})$ and, even better, if we only consider the subspace Z of C spanned by formal linear combinations which become loops in X_Γ, rather than just paths (i.e. the kernel of the map from C to the $\mathbf{Q}$-vector space spanned by Γ-orbits of cusps sending γ to $\gamma.\infty - \gamma.0$ then we get a surjection from Z to $H_1(X_\Gamma, \mathbf{Q})$ and the kernel of this map contains, and can be shown to be equal to, the $\mathbf{Q}$-subspace B of C spanned by $(\gamma) + (\gamma.S)$ and $(\gamma) + (\gamma.TS) + (\gamma.TSTS)$ (note that these things are obviously all in Z).

2 Merel's thesis.

I wonder whether what Merel does is exactly the same as above but in some kind of more abstract language. Merel starts off with relative homology and cohomology, so let's review that. If X is a topological space then its homology is defined as the homology groups of the chain complex $C_*(X) = \cdots \to C_2(X) \to C_1(X) \to C_0(X) \to 0$, which in degree n is the abelian group with basis the continuous maps from the standard n-simplex to X, and whose boundary maps are the usual ones. Now if Y is a subspace of X then there's an exact sequence of chain complexes

$$0 \to C_*(Y) \to C_*(X) \to C_*(X)/C_*(Y) \to 0$$

and the homology of the rightmost complex is (by definition) the relative homology of X with respect to Y. Richard Thomas told me to think of it as measuring cycles in X with boundaries in Y. The short exact sequence of cycles above gives a long exact sequence of homology groups

$$\cdots \to H_1(Y) \to H_1(X) \to H_1(X, Y) \to H_0(Y) \to H_0(X) \to H_0(X, Y) \to 0$$

and the connecting homomorphism can be interpreted thus: given an n-cycle in X with boundary in Y which contributes to $H_n(X, Y)$, then by definition the n-cycle in X, when you d it, becomes an $n-1$-cycle living completely in Y and this is the boundary map.

Example: if X is a compact Riemann surface, if A and B are finite disjoint non-empty subsets of X, and if we consider the relative homology of $X - A$ with respect to $Y = B$, we get

$$0 = H_2(X-A) \to H_2(X-A, B) \to H_1(B) = 0 \to H_1(X-A) \to H_1(X-A, B) \to H_0(B) \to H_0(X-A) \to H_0(X-A, B)$$

so $H_2(X-A, B) = 0$ and $0 \to H_1(X-A) \to H_1(X-A, B) \to \mathbf{Z}^{\#B} \to \mathbf{Z} \to 0$ and $H_0(X-A, B) = 0$. In particular $H_1(X-A, B)$ is a beefing-up of $H_1(X-A)$ by a group of rank $\#B-1$ and furthermore $H_1(X-A, B)$ is a free abelian group (the exact sequence of length 4 breaks into two short exact sequences, both of which split).

Notation: $H_A^B := H_1(X-A, B)$. "Paths start and end at the top index".

Merel says that the cup product induces a perfect pairing between H_B^A and H_A^B.

Now let Γ be a congruence subgroup containing -1, let C denote the cusps in $X := X_\Gamma$, let I denote the stuff that maps to i in the j-line, and let R denote the stuff that maps to ρ in the j-line. Merel proves in his thesis that the map sending $g \in \mathrm{SL}_2(\mathbf{Z})$ to the path from gi to $g.\infty$ induces an isomorphism

$$\mathbf{Z}^{(\Gamma\backslash \mathrm{SL}_2(\mathbf{Z}))} \to H_R^{C\cup I}$$

and the map sending g to the path from ω (the cube root of unity in the upper half plane) to $-\omega^2$ induces an isomorphism

$$\mathbf{Z}^{(\Gamma\backslash \mathrm{SL}_2(\mathbf{Z}))} \to H_{C\cup I}^R.$$

Furthermore the cup product pairing becomes identified with the obvious pairing on $\mathbf{Z}^{(\Gamma\backslash \mathrm{SL}_2(\mathbf{Z}))}$.

What one is however really after is a description of $H_1(Y_\Gamma)$. As we've seen above, $H_1(X-C)$ injects into $H_1(X-C,R)$ which is canonically a quotient of $H_1(X-C-I,R) = H_{C+I}^R$. Merel explicitly writes down the pre-image of $H_1(X-C)$ in H_{C+I}^R and one imagines that he's perhaps doing much the same as Cremona here. This is in section 1 of his thesis. Later on he considers the path from i to ρ and the same construction gives an isomorphism

$$\mathbf{Z}^{(\Gamma\backslash \mathrm{SL}_2(\mathbf{Z}))} \to H_C^{R+I}$$

and Merel locates the image of H_C in this space.

3 Higher weight: Ash-Stevens.

I have no idea whether all this is due to Ash-Stevens or whether it was known for ages. Ash-Stevens is certainly a reference!

Let $\mathcal{D}_0$ denote the divisors of degree zero on $\mathbf{P}^1(\mathbf{Q})$. The group $\mathcal{D}_0$ has a natural action of $\mathrm{GL}_2(\mathbf{Q})$. If Γ is a congruence subgroup and E is a Γ-module (that is, an abelian group with a left action of Γ) then the *E-valued modular symbols* are just $\mathrm{Hom}_\Gamma(\mathcal{D}_0, E)$, the Γ-linear maps from $\mathcal{D}_0$ to E. If $\mathcal{D}$ is all divisors on $\mathbf{P}^1(\mathbf{Q})$ then $0 \to \mathcal{D}_0 \to \mathcal{D} \to \mathbf{Z} \to 0$ is a short exact sequence of $\mathrm{GL}_2(\mathbf{Q})$-modules (with $\mathbf{Z}$ having the trivial action). In section 4 of Ash-Stevens "modular forms in characteristic ℓ" they assert that if furthermore if E is a $\mathbf{Z}[1/t]$-module for t the lowest common multiple of the orders of the torsion elements in Γ (so t divides 12) then $H_c^1(\Gamma, E) = \mathrm{Hom}_\Gamma(\mathcal{D}_0, E)$, and moreover Stevens asserts that the map from right to left sends the modular symbol $\phi : \mathcal{D}_0 \to E$ to the 1-cocycle $g \mapsto \phi(gx - x)$ for any $x \in \mathbf{P}^1(\mathbf{Q})$.

Ash and Stevens then go on to show that if f is a cuspidal eigenform of weight k for Γ then the special values of $L(f \otimes \chi, s)$ at $s = 1, 2, \ldots, k-1$ (here χ is a Dirichlet character) can be put together to give an element of $\mathrm{Symm}^{k-2}(\mathbf{C}^2)$ which turns out to be an explicit linear combination of values of a certain modular symbol—the $\mathrm{Symm}^{k-2}(\mathbf{C}^2)$-valued one naturally coming from f via integration (its value at $y - x$ is just the integral of a certain differential from x to y). For some reason this should be trivial but the reference in Ash-Stevens seems to be unpublished :-/