MODULAR SYMBOLS AND THE COMPUTATION OF MODULAR ELLIPTIC CURVES

J.E.CREMONA

1. Introduction

Modular symbols are used to compute the homology of certain manifiolds which arise in number theory, related to modular forms and elliptic curves. The idea originated with Birch, but Manin was the first to develop them systematically. Now they occur in many different guises, from the relatively concrete to the relatively abstract: for example, they are also used in K-theory.

In these lecture notes we will show how modular symbols may be used in the computation of certain modular forms, and hence modular elliptic curves (over the rationals $\mathbb{Q}$). Everything here will be completely explicit; the methods have all been implemented by the author as a set of programs (in `C++`) which will, in principle, output for any given positive integer N, a complete set of elliptic curves over $\mathbb{Q}$ of conductor N, up to isogeny.

It is impractical to make these notes completely self-contained. Prerequisite background knowledge expected of the reader consists of: basic theory of holomorphic modular forms for the modular group $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ and its congruence subgroups; a few results from Riemann surface theory. We will also need a little basic algebraic topology (mainly the definition of homology groups in terms of triangulations of a surface); this will be outlined very briefly in section 2.2 below.

References for modular symbols include: Lang's book [14], the articles [15] and [16] and the author's book [6]. A good reference which also includes a detailed account of the theoretical construction of modular elliptic curves is Knapp's book [10]. These notes follow Chapter 2 of [6] closely, with certain details added or omitted in places.

2. Modular curves and homology

2.1. The upper half-plane, the modular group and cusp forms.

Let $\mathcal{H}$ denote the upper half-plane

$$\mathcal{H} = \{z = x + iy \in \mathbb{C} \mid y > 0\},$$

and $\mathcal{H}^* = \mathcal{H} \cup \mathbb{Q} \cup \{\infty\}$ the extended upper half-plane, obtained by including the cusps $\mathbb{Q} \cup \{\infty\}$. The group $PSL_2(\mathbb{R})$ acts on $\mathcal{H}^*$ via linear fractional transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az+b}{cz+d};$$

these are the isometries of the hyperbolic geometry on $\mathcal{H}$, for which geodesics are either half-lines perpendicular to the real axis $\mathbb{R}$, or semicircles perpendicular to $\mathbb{R}$.

The modular group $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ is a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$ (in the topology induced from $\mathrm{SL}(2,\mathbb{R}) \subset M_2(\mathbb{R}) \cong \mathbb{R}^4$), and acts discontinuously on $\mathcal{H}$, in the sense that for each $z \in \mathcal{H}$ the orbit $\Gamma.z$ is discrete. Note that the cusps $\mathbb{Q} \cup \{\infty\}$ form a complete Γ-orbit.

Γ is generated by the matrices $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} : z \mapsto -1/z$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} : z \mapsto z+1$. This fact, and the related fact that a fundamental region for the action of Γ on $\mathcal{H}$ is given by the set

$$\mathcal{F} := \{z = x + iy \in \mathcal{H} \mid |x| \le \frac{1}{2}, |z| \ge 1\}, \tag{2.1.1}$$

are standard and will not be proved here. Both results depend essentially on the fact that $\mathbb{Z}$ is Euclidean.

Let G be a subgroup of Γ of finite index e. Then G also acts discretely on $\mathcal{H}$. A fundamental region for G on $\mathcal{H}$ is given by $\cup M_i.\mathcal{F}$, where the M_i, (for $1 \le i \le e$), are right coset representatives for G in Γ.

Let $X_G = G\backslash\mathcal{H}^*$ denote the quotient space; this may be given the structure of a compact Riemann surface. Around most points the local parameter is just z, but more care is needed about the "parabolic points" or cusps, and the "elliptic points" which have non-trivial stabilisers in the Γ-action. These elliptic points for G (if any) are in the orbits of i (stabilised by S of order 2), and of $\rho = (1+\sqrt{-3})/2$ (stabilised by TS of order 3). See the books of Lang [14], Shimura [19] or Knapp [10] for details of the Riemann surface construction.

Let g denote the genus of the surface X_G; as a real manifold, X_G is a g-holed torus. (Strictly speaking, X_G is not a manifold unless G has no elements of finite order, because of the branching over the elliptic points. However this makes no difference in practice and we may safely ignore it.) We will be concerned with the explicit computation of the 1-homology $H_1(X_G, \mathbb{Z})$, which is a free $\mathbb{Z}$-module of rank $2g$. (See §2.2 below for a brief review of homology). This homology will be expressed in terms of "modular symbols", defined below. We must also explain the connection between homology, modular forms, and elliptic curves. First we review the definition of cusp forms.

The space of *cusp forms of weight 2* for G will be denoted by $S_2(G)$. These cusp forms are holomorphic functions $f(z)$ (for $z \in \mathcal{H}$) which satisfy

(1) $f \mid M = f$ for all $M \in G$, where

$$(f \mid \begin{pmatrix} a & b \\ c & d \end{pmatrix})(z) = (cz+d)^{-2} f(\frac{az+b}{cz+d}).$$

Thus, since $(cz+d)^{-2} = (d/dz)(M(z))$ for $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, we have

$$f(M(z))d(M(z)) = f(z)dz \qquad \forall M \in G.$$

(2) $f(z)$ behaves nicely at the cusps. The significance of this condition is that, by (1), a cusp form of weight 2 for G is the pull-back of a (holomorphic) differential on the Riemann surface $G\backslash\mathcal{H}$, of which X_G is the compactification after adding the (finitely many) G-inequivalent cusps, and we want this differential to be holomorphic on the whole of X_G. In future we will identify cusp forms of weight 2 for G with holomorphic differentials on X_G. From standard Riemann surface theory, we then know that $S_2(G)$ is a complex vector space of dimension g.

We can make explicit the condition that $f(z)dz$ is holomorphic at the cusp ∞ (for the other cusps, see one of the references on the theory of modular forms). The stabiliser of ∞ in Γ is the infinite cyclic subgroup generated by T; if h is the least positive integer such that $T^h \in G$, then clearly we have

$$\mathrm{Stab}(\infty) \cap G = \langle T^h \rangle,$$

and every $f \in S_2(G)$ has a Fourier expansion of the form

$$f(z) = \sum_{n=1}^{\infty} a_n e^{2\pi i n z/h} \tag{2.1.2}$$

with coefficients $a_n \in \mathbb{C}$. The integer h is called the *width* of the cusp ∞; for $G = \Gamma_0(N)$, the case we will be most interested in, we have $h = 1$, since $T \in \Gamma_0(N)$.

2.2. The duality between cusp forms and homology.

The basis for our method is the explicit computation of the homology (specifically, the rational 1-homology) of the Riemann surface X_G. This is useful for various reasons. On the one hand, this gives us a very explicit vector space on which Hecke operators act, which is isomorphic (or more strictly, dual) to the space of cusp forms. Thus by computing homology and the Hecke action on it, we are indirectly also able to obtain information about the space of cusp forms. The Fourier coefficients of the cusp forms are determined by their Hecke eigenvalues (see §5.2), so we obtain these indirectly as eigenvalues of Hecke operators acting on homology. Secondly, in order to actually compute the elliptic curves attached to these cusp forms, we need to know their periods, which are obtained by integrating the corresponding differentials around closed paths on the surface X_G; since two paths give the same period (for all forms) if and only if they are homologous (essentially by Cauchy's Theorem on X_G), it is clear that to determine the whole period lattice we will also require an explicit knowledge of the homology of X_G.

The integral homology $H_1(X_G, \mathbb{Z})$ is most easily defined geometrically: it is the abeilan group obtained by taking as generators all closed paths on X_G, and factoring out by the relation that two closed paths are equivalent (or *homologous*) if one can be continuously deformed into the other. If the genus of the surface X_G is g, this gives a free abelian group of rank $2g$: roughly speaking, the surface is a g-holed torus, and there are two generating loops around each hole. To determine this homology group in practice, one triangulates the surface, so that every path is homologous to a path along the edges of the triangulation. Now the generators are the directed edges of the triangulation, modulo relations given by the sum of the edges around each triangle being homologous to zero. A typical element of $H_1(X_G, \mathbb{Z})$ will then be given as a $\mathbb{Z}$-linear combination of these directed edges. In §2.6 below, we will make this very explicit: there will be one edge of the triangulation for each coset of G in Γ, and the triangle relations will be expressed algebraically in terms of the coset action of Γ. This description will entirely algebraicise the situation, in a way which is then easy to implement on a computer.

For any other ring R, the homology with coefficients in R is obtained simply by tensoring with R:

$$H_1(X_G, R) = H_1(X_G, \mathbb{Z}) \otimes_{\mathbb{Z}} R.$$

Explicitly, one just takes R-linear combinations of the $2g$ generators of the $\mathbb{Z}$-module $H_1(X_G, \mathbb{Z})$ ("extension of scalars"); the result is then an R-module. In what follows we will only need to take $R = \mathbb{Q}$, $R = \mathbb{R}$, and $R = \mathbb{C}$.

Let $H_1(X_G, \mathbb{R}) = H_1(X_G, \mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{R}$, which is a real vector space of dimension $2g$. Abstractly, this space is obtained by formal extension of scalars from $H_1(X_G, \mathbb{Z})$; but we can be more concrete, if we introduce the notion of modular symbols.

First let α, $\beta \in \mathcal{H}^*$ be points equivalent under the action of G, so that $\beta = M(\alpha)$ for some $M \in G$. Any smooth path (for instance, a geodesic path) from α to β in $\mathcal{H}^*$ projects to a closed path in the quotient space X_G, and hence determines an integral homology class in $H_1(X_G, \mathbb{Z})$, which depends only on α and β and not on the path chosen, because $\mathcal{H}^*$ is simply connected. (In fact, the class depends only on M: see (5) in Lemma 2.4.1 below). We denote this homology class by the *modular symbol* $\{\alpha, \beta\}_G$, or simply $\{\alpha, \beta\}$ when the group G is clear from the context.

Conversely, every integral homology class $\gamma \in H_1(X_G, \mathbb{Z})$ can be represented by such a modular symbol $\{\alpha, \beta\}$. Also, if $f \in S_2(G)$ then the integral

$$\int_\gamma 2\pi i f(z) dz = 2\pi i \int_\alpha^\beta f(z) dz$$

is well-defined, since $f(z)$ is holomorphic, and will be denoted $\langle \gamma, f \rangle$. The (complex) value of such an integral is called a *period* of the cusp form f, or of the associated differential.

Let f_1, f_2, $\dots$, f_g be a fixed basis for $S_2(G)$, so that the differentials $2\pi i f_j(z) dz$ are a basis for the holomorphic differentials on X_G. Also let γ_1, γ_2, $\dots$, γ_{2g} be a fixed $\mathbb{Z}$-basis for the integral homology $H_1(X_G, \mathbb{Z})$. Then we may form the $2g \times g$ complex *period matrix*

$$\Omega = (\omega_{jk}) = (\langle \gamma_j, f_k \rangle) .$$

By standard Riemann surface theory, the $2g$ rows of Ω are linearly independent over $\mathbb{R}$, and so their $\mathbb{Z}$-span is a lattice (discrete subgroup) Λ of rank $2g$ in $\mathbb{C}^g$. The quotient $J(G) = \mathbb{C}^g/\Lambda$ is the Jacobian of X_G: it is an abelian variety of dimension g.

The symbols $\{\alpha, \beta\}$ give $\mathbb{C}$-linear functionals $S_2(G) \to \mathbb{C}$ via $f \mapsto \langle \{\alpha, \beta\}, f \rangle = 2\pi i \int_\alpha^\beta f(z) dz$. We may identify $H_1(X_G, \mathbb{R})$ with the space of all $\mathbb{C}$-linear functionals on $S_2(G)$ as follows: given an element $\gamma \in H_1(X_G, \mathbb{R})$ we can write γ uniquely in the form

$$\gamma = \sum_{j=1}^{2g} c_j \gamma_j$$

with coefficients $c_j \in \mathbb{R}$. Define $\langle \gamma, f \rangle = \sum_{j=1}^{2g} c_j \langle \gamma_j, f \rangle$. Then the corresponding functional is

$$f \mapsto \langle \gamma, f \rangle .$$

Conversely, given a functional $\omega \colon S_2(G) \to \mathbb{C}$, the vector $(\omega(f_1), \omega(f_2), \dots, \omega(f_g)) \in \mathbb{C}^g$ may be expressed uniquely as an $\mathbb{R}$-linear combination of the rows of Ω, so there exist real scalars c_j $(1 \le j \le 2g)$ such that

$$\omega(f) = \sum_{j=1}^{2g} c_j \langle \gamma_j, f \rangle$$

for all $f \in S_2(G)$; then $\omega(f) = \langle \gamma, f \rangle$ where $\gamma = \sum_{j=1}^{2g} c_j \gamma_j \in H_1(X_G, \mathbb{R})$.

In particular, let α, $\beta \in \mathcal{H}^*$ be arbitrary (not necessarily in the same G-orbit); then the functional

$$f \mapsto \int_\alpha^\beta 2\pi i f(z) dz$$

corresponds to a unique element $\gamma = \sum_{j=1}^{2g} c_j \gamma_j \in H_1(X_G, \mathbb{R})$, and we *define* the modular symbol $\{\alpha, \beta\}_G \in H_1(X_G, \mathbb{R})$ to be this element. Clearly this definition agrees with the earlier one in the special case where $\beta = M(\alpha)$ for some $M \in G$; indeed, this case holds if and only if all $c_j \in \mathbb{Z}$.

By the *field of definition* of an element $\gamma \in H_1(X_G, \mathbb{R})$ we mean the field generated over $\mathbb{Q}$ by its coefficients c_j (with respect to the $\mathbb{Z}$-basis for the integral homology, as above). For example, γ is rational (has field of definition $\mathbb{Q}$) if and only if $\gamma \in H_1(X_G, \mathbb{Q})$.

We now have an $\mathbb{R}$-bilinear pairing

$$S_2(G) \times H_1(X_G, \mathbb{R}) \longrightarrow \mathbb{C} \tag{2.2.1}$$

given by

$$(f, \gamma) \mapsto \langle \gamma, f \rangle = \int_\gamma 2\pi i f(z) dz$$

which gives an exact duality between the two spaces on the left if we view $S_2(G)$ as a real vector space of dimension $2g$.

To interpret this as a duality over $\mathbb{C}$, we can give $H_1(X_G, \mathbb{R})$ the structure of a vector space over $\mathbb{C}$ (of dimension g) as follows. Given $\gamma \in H_1(X_G, \mathbb{R})$ and $c \in \mathbb{C}$, we define $c\gamma$ to be that element of $H_1(X_G, \mathbb{R})$ which satisfies $\langle c\gamma, f \rangle = \langle \gamma, cf \rangle$ for all $f \in S_2(G)$; in other words, $c\gamma$ is the element corresponding to the functional $f \mapsto c\langle \gamma, f \rangle$. Now the map $(f, \gamma) \mapsto \langle \gamma, f \rangle$ is $\mathbb{C}$-bilinear, and our dual pairing (2.2.1) between homology and cusp forms is an exact duality over $\mathbb{C}$.

2.3. Real Structure.

For suitable groups G we can restrict the duality described above to a duality between real vector spaces of dimension g. This has important implications for explicit computations, where a halving of the dimension (from $2g$ to g) gives a significant saving of effort.

Let $J = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. We say that a subgroup G of Γ is *of real type* if J normalises G. Explicitly, let $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$; then $M^* := J^{-1} M J = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$, and G is of real type when $M \in G \iff M^* \in G$. This will be true, in particular, for the congruence subgroups $\Gamma_0(N)$ and $\Gamma_1(N)$ of most interest to us.

For $z \in \mathcal{H}$ set $z^* = -\overline{z}$. Then a trivial calculation shows that $w = M(z) \iff w^* = M^*(z^*)$; it follows that, for G of real type, the map $z \mapsto z^*$ induces a well-defined map on the quotient X_G, and hence also on homology, via $\{\alpha, \beta\} \mapsto \{\alpha^*, \beta^*\}$. Clearly this is an $\mathbb{R}$-linear involution on $H_1(X_G, \mathbb{R})$. Hence we obtain a decomposition into $+1$ and -1 eigenspaces for $*$:

$$H_1(X_G, \mathbb{R}) = H_1^+(X_G, \mathbb{R}) \oplus H_1^-(X_G, \mathbb{R}).$$

Remark. The involution $*$ also acts on the integral homology $H_1(X_G, \mathbb{Z})$, and we may set $H_1^{\pm}(X_G, \mathbb{Z}) = H_1^{\pm}(X_G, \mathbb{R}) \cap H_1(X_G, \mathbb{Z})$. However the direct sum $H_1^{+}(X_G, \mathbb{Z}) \oplus H_1^{-}(X_G, \mathbb{Z})$ will in general have finite index in $H_1(X_G, \mathbb{Z})$.

We now define dually an involution, also denoted $*$, on the space $S_2(G)$ where G is of real type. For a holomorphic function f on $\mathcal{H}$, we set $f^*(z) = \overline{f(z^*)}$. The following facts are now easily verified:

(1) If f has Fourier expansion $f(z) = \sum a_n q^n$ (where $q = \exp(2\pi i z/h)$ as in (2.1.2) above), then $f^*(z) = \sum \overline{a_n} q^n$. In other words, the Fourier coeffients of f^* are the conjugates of those of f.

(2) For $M \in \Gamma$, we have $f^* \mid M = (f \mid M^*)^*$.

(3) $\langle \gamma^*, f^* \rangle = \overline{\langle \gamma, f \rangle}$ for all f, γ.

As a formal consequence of fact (2), we immediately see that, for G of real type, the map $f \mapsto f^*$ is an $\mathbb{R}$-linear map from $S_2(G)$ to itself, which is an involution. Denote by $S_2(G)_{\mathbb{R}}$ the $\mathbb{R}$-subspace of $S_2(G)$ fixed by this involution, which by fact (1) consists of those cusp forms with real Fourier coefficients. Then $\dim_{\mathbb{R}}(S_2(G)_{\mathbb{R}}) = g$, and $S_2(G)_{\mathbb{R}}$ spans $S_2(G)$ over $\mathbb{C}$.

For nonzero $f \in S_2(G)_{\mathbb{R}}$ we have (from fact (3)):

$$\langle \gamma, f \rangle \in \mathbb{R} \iff \gamma \in H_1^{+}(X_G, \mathbb{R}),$$

and also

$$\langle \gamma, f \rangle \in i\mathbb{R} \iff \gamma \in H_1^{-}(X_G, \mathbb{R}).$$

Moreover, multiplication by i on $H_1(X_G, \mathbb{R})$ interchanges the "real" and "pure imaginary" eigenspaces $H_1^{\pm}(X_G, \mathbb{R})$ since

$$\begin{aligned} \gamma \in H_1^{+} &\iff \langle \gamma, f \rangle \in \mathbb{R} \qquad \forall f \in S_2(G)_{\mathbb{R}} \\ &\iff \langle i\gamma, f \rangle \in i\mathbb{R} \qquad \forall f \in S_2(G)_{\mathbb{R}} \\ &\iff i\gamma \in H_1^{-}. \end{aligned}$$

It follows that $\dim H_1^{+}(X_G, \mathbb{R}) = \dim H_1^{-}(X_G, \mathbb{R}) = g$.

Also, since the period $\langle \gamma, f \rangle$ is real for $\gamma \in H_1^{+}$ and $f \in S_2(G)_{\mathbb{R}}$, the duality over $\mathbb{C}$ we had earlier now restricts to a duality over $\mathbb{R}$:

$$S_2(G)_{\mathbb{R}} \times H_1^{+}(X_G, \mathbb{R}) \longrightarrow \mathbb{R}. \tag{2.3.1}$$

It follows that the real vector spaces $S_2(G)_{\mathbb{R}}$ and $H_1^{+}(X_G, \mathbb{R})$ are dual to each other. We will exploit this duality (which also respects the action of Hecke and other operators, see below), as we will compute $H_1(X_G, \mathbb{R})$ explicitly in order to gain information about the cusp forms in $S_2(G)$. Also, this duality is crucial in the definition of modular elliptic curves.

2.4. Modular symbol formalism.

The following simple properties of the modular symbols $\{\alpha, \beta\}$ are now immediate.

LEMMA 2.4.1. *Let $\alpha, \beta, \gamma \in \mathcal{H}^*$, and let M, M_1, $M_2 \in G$. Then*
(1) $\{\alpha, \alpha\} = 0$;
(2) $\{\alpha, \beta\} + \{\beta, \alpha\} = 0$;
(3) $\{\alpha, \beta\} + \{\beta, \gamma\} + \{\gamma, \alpha\} = 0$;
(4) $\{M\alpha, M\beta\}_G = \{\alpha, \beta\}_G$;
(5) $\{\alpha, M\alpha\}_G = \{\beta, M\beta\}_G$;
(6) $\{\alpha, M_1 M_2 \alpha\}_G = \{\alpha, M_1\alpha\}_G + \{\alpha, M_2\alpha\}_G$;
(7) $\{\alpha, M\alpha\}_G \in H_1(X_G, \mathbb{Z})$.

PROOF. Only (5) and (6) are not quite obvious. For (5), write $\{\alpha, M\alpha\} = \{\alpha, \beta\} + \{\beta, M\beta\} + \{M\beta, M\alpha\}$, using (2) and (3); now the first and third terms cancel by (4). For (6), we have $\{\alpha, M_1 M_2\alpha\} = \{\alpha, M_1\alpha\} + \{M_1\alpha, M_1 M_2\alpha\} = \{\alpha, M_1\alpha\} + \{\alpha, M_2\alpha\}$ using (4). □

COROLLARY 2.4.2. *The map $M \mapsto \{\alpha, M\alpha\}_G$ is a surjective group homomorphism $G \to H_1(X_G, \mathbb{Z})$, which is independent of $\alpha \in \mathcal{H}^*$.*

The kernel of this homomorphism contains all commutators and elliptic elements (since the latter have finite order, and the image is a torsion-free abelian group), and also all parabolic elements: for if $M \in G$ is parabolic, it is a conjugate of T and hence fixes some $\alpha \in \mathbb{Q} \cup \{\infty\}$, so $M \mapsto \{\alpha, M\alpha\} = 0$. In fact, the kernel is generated by these elements, but we will not prove that here.

2.5. Rational Structure and the Manin-Drinfeld Theorem.

We have seen that every element γ of $H_1(X_G, \mathbb{Z})$ has the form $\{\alpha, M\alpha\}$ with $M \in G$ and $\alpha \in \mathcal{H}^*$ arbitrary; usually we take α to be a cusp, so that γ is a path between G-equivalent cusps. It is not clear in general what is the field of definition of a modular symbol $\{\alpha, \beta\}_G$ for which α and β are both cusps. However, when G is a congruence subgroup, this is answered by the Manin-Drinfeld Theorem.

A *congruence subgroup* of Γ is a subgroup G such that membership of G is determined by means of congruence conditions on the entries of a matrix in Γ. A moment's thought shows that this is equivalent to the condition that for some positive integer N, G contains the *principal congruence subgroup* $\Gamma(N)$, which is defined to be the subgroup of Γ consisting of matrices congruent to the identity modulo N. The least such N is called the *level* of G.

The most important congruence subgroups are $\Gamma(N)$ itself;

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \mid c \equiv 0 \pmod{N} \right\};$$

and

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \mid c \equiv 0, a \equiv 1 \pmod{N} \right\}.$$

We can now state the Manin-Drinfeld Theorem.

THEOREM 2.5.1. *(Manin, Drinfeld) Let G be a congruence subgroup of the modular group Γ. Then for all pairs of cusps α, $\beta \in \mathcal{H}^*$ we have*

$$\{\alpha, \beta\}_G \in H_1(X_G, \mathbb{Q}).$$

In particular, the modular symbol $\{0, \infty\}_G$ is rational; the denominator of this element is very important in many ways.

Thus the rational homology $H_1(X_G, \mathbb{Q})$ is generated by paths between cusps (since it is generated by the integral homology), and conversely every path between cusps is rational. We will see how to use this fact to develop an algorithm for computing the rational homology in the next section.

The proof of the Manin-Drinfeld Theorem involves the use of Hecke operators (see Lang's book [14], for example). Using Hecke operators we can also prove that the space of cusp forms $S_2(G)$ has a $\mathbb{Q}$-structure, namely a basis consisting of forms with rational Fourier coefficients (when G is a congruence subgroup). This is related to the fact that the modular curve X_G, which as a Riemann surface is certainly an algebraic curve over $\mathbb{C}$, can in fact be given the structure of an algebraic curve over the field of algebraic numbers $\overline{\mathbb{Q}}$ (and even over the N'th cyclotomic field, if G has level N). This rational structure is crucial to the construction of modular elliptic curves, to ensure that we obtain elliptic curves defined over $\mathbb{Q}$.

The duality between cusp forms and homology does not descend entirely to $\mathbb{Q}$, however, because even if $f \in S_2(G)$ has rational Fourier coefficients and $\gamma \in H_1(X_G, \mathbb{Q})$, the period $\langle \gamma, f \rangle$ will not be rational. Rationality questions for periods of modular forms have been studied extensively, notably by Manin, but we will not go into this further here.

2.6. Triangulations and homology.

From now on we will assume that G is a congruence subgroup, so that the rational homology of X_G is precisely the homology generated by paths between cusps.

We will compute the homology of X_G by first triangulating the upper half-plane $\mathcal{H}^*$, using a tessellation of hyperbolic triangles, and then passing to the quotient. This will give us a very explicit triangulation of the surface X_G, using which we can write down explicit generators and relations for its 1-homology.

For $M \in \Gamma$ let $(M) := \{M(0), M(\infty)\}$, viewed as a path in $\mathcal{H}^*$, the image under M of the imaginary axis $\{0, \infty\}$. These geodesic paths form the oriented edges of a triangulation of $\mathcal{H}^*$ whose vertices are the cusps $\mathbb{Q} \cup \{\infty\}$. Explicitly, there is an edge from ∞ to n for all $n \in \mathbb{Z}$, and an edge between rational numbers a/c and b/d such that $ad - bc = 1$. The triangles of the triangulation are images under $M \in \Gamma$ of the basic triangle with vertices at 0, 1 and ∞ and edges (I), (TS), $((TS)^2)$. We denote by $\langle M \rangle$ the image of this triangle under M, which has vertices $M(0)$, $M(1)$ and $M(\infty)$ and edges (M), (MTS) and $(M(TS)^2)$. This representation of the triangles is unique except for the relation

$$\langle M \rangle = \langle MTS \rangle = \left\langle M(TS)^2 \right\rangle.$$

Also, triangles $\langle M \rangle$ and $\langle MS \rangle$ meet along the edge (M) since $(MS) = -(M)$ (the negative sign indicating reverse orientation).

We will use the symbol $(M)_G$ to denote the image of the path (M) in the quotient X_G, and also its image in the rational homology. The geometric observations of the previous paragraph now give us the following relations between these symbols:

$$\begin{aligned} (M)_G + (MTS)_G + (M(TS)^2)_G &= 0 \\ (M)_G + (MS)_G &= 0. \end{aligned}$$

We also clearly have the relations

$$(M'M)_G = (M)_G \tag{2.6.1}$$

for all $M' \in G$, so we may use as generators of the rational homology the finite set of symbols $(M_i)_G$, $(1 \le i \le e)$, where as before $M_1, \ldots, M_e$ are a set of right coset representatives for G in Γ.

Let $C(G)$ be the $\mathbb{Q}$-vector space with basis the formal symbols $(M)_G$ for each M in Γ, identified by the relations (2.6.1), so that $\dim(C(G)) = e = [\Gamma : G]$.

Let $B(G)$ be the subspace of $C(G)$ spanned by all elements of the form

$$(M)_G + (MS)_G,$$
$$(M)_G + (MTS)_G + (M(TS)^2)_G.$$

Let $C_0(G)$ be the $\mathbb{Q}$-vector space spanned by the G-cusps $[\alpha]_G$ for $\alpha \in \mathbb{Q} \cup \{\infty\}$ (so that $[\alpha]_G = [\beta]_G \iff \beta = M(\alpha)$ for some $M \in G$). Define the boundary map $\delta: C(G) \to C_0(G)$ by

$$\delta((M)_G) = [M(\infty)]_G - [M(0)]_G$$

and set $Z(G) = \ker(\delta)$. Note that $B(G) \subseteq Z(G)$, by a trivial calculation using the facts that S transposes 0 and ∞ while TS cycles 0, 1 and ∞.

Finally we define $H(G) = Z(G)/B(G)$. The crucial result, due in this form to Manin [15, Theorem 1.9], is that this formal construction does in fact give us the rational homology of X_G:

THEOREM 2.6.1. *$H(G)$ is isomorphic to $H_1(X_G, \mathbb{Q})$, the isomorphism being induced by*

$$(M)_G \mapsto \{M(0), M(\infty)\}_G.$$

We may thus use the symbol $(M)_G$ either as an abstract symbol obeying certain relations, or to denote an element of $H_1(X_G, \mathbb{Q})$, without confusion. In future, as the subgroup G will be fixed, we will omit the subscript on these symbols and blur the distinction between (M) as a path in the upper half-plane and $(M)_G$ as representing an element of the rational 1-homology of X_G.

Note that the form of the relations between the generating symbols (M) does not depend at all on the specific group G. In particular we do not have to consider explicitly the shape of a fundamental region for the action of G on $\mathcal{H}^*$, or how the edges of such a region are identified. This represents a major simplification compared with earlier approaches, such as that used by Tingley [22]. In order to develop this result into an explicit algorithm for computing homology, we need to have a specific set of right coset representatives for the subgroup G of Γ, and also to have a test for G-equivalence of cusps. These are purely algebraic problems which can easily be solved for arithmetically defined subgroups G such as congruence subgroups. Observe that from this point on, we do not have to do any geometry at all.

One final remark before we specialise to the case $G = \Gamma_0(N)$: every path between cusps may be expressed as a finite sum of paths of the form (M) with $M \in \Gamma$.

Writing $\{\alpha,\beta\} = \{0,\beta\} - \{0,\alpha\}$, it suffices to do this for modular symbols of the form $\{0,\alpha\}$. Let

$$\frac{p_{-2}}{q_{-2}} = \frac{0}{1}, \frac{p_{-1}}{q_{-1}} = \frac{1}{0}, \frac{p_0}{1} = \frac{p_0}{q_0}, \frac{p_1}{q_1}, \frac{p_2}{q_2}, \dots, \frac{p_k}{q_k} = \alpha$$

denote the continued fraction convergents of the rational number α. Then, as is well-known,

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

Hence

$$\{0,\alpha\} = \sum_{j=-1}^{k} \left\{ \frac{p_{j-1}}{q_{j-1}}, \frac{p_j}{q_j} \right\} = \sum_{j=-1}^{k} \{M_j(0), M_j(\infty)\} = \sum_{j=-1}^{k} (M_j)$$

where $M_j = \begin{pmatrix} (-1)^{j-1}p_j & p_{j-1} \\ (-1)^{j-1}q_j & q_{j-1} \end{pmatrix}$.

These conversions are easy to implement using the Euclidean algorithm.

3. M-symbols and $\Gamma_0(N)$

We now specialise to the case $G = \Gamma_0(N)$:

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \mid c \equiv 0 \pmod{N} \right\}.$$

The index of $\Gamma_0(N)$ in Γ is given (see [19, Proposition 1.43]) by

$$[\Gamma : \Gamma_0(N)] = N \prod_{p \mid N} \left(1 + p^{-1}\right).$$

Define $H(N) = H(\Gamma_0(N))$ and $X_0(N) = X_G$. After Theorem 2.6.1, we will identify $H(N)$ with $H_1(X_0(N), \mathbb{Q})$ by identifying (M) with $\{M(0), M(\infty)\}$.

The next Lemma is needed to determine right coset representatives for $\Gamma_0(N)$ in Γ.

LEMMA 3.1.1. *For $i = 1,2$ let $M_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} \in \Gamma$. The following are equivalent.*

(1) *The right cosets $\Gamma_0(N)M_1$ and $\Gamma_0(N)M_2$ are equal;*
(2) $c_1 d_2 \equiv c_2 d_1 \pmod{N}$*;*
(3) *There exists u with $\gcd(u,N) = 1$ such that $c_1 \equiv u c_2$ and $d_1 \equiv u d_2 \pmod{N}$.*

On the set of ordered pairs $(c,d) \in \mathbb{Z}^2$ such that $\gcd(c,d,N) = 1$ we now define the relation $\sim$, where

$$(c_1, d_1) \sim (c_2, d_2) \iff c_1 d_2 \equiv c_2 d_1 \pmod{N}. \tag{3.1.1}$$

By Lemma 3.1.1, this is an equivalence relation. The equivalence class of (c,d) will be denoted $(c:d)$, and such symbols will be called M-symbols (after Manin, who introduced them in [15]). The set of these M-symbols modulo N is $P^1(N) = P^1(\mathbb{Z}/N\mathbb{Z})$, the projective line over the ring of integers modulo N.

Notice that in an M-symbol $(c:d)$, c and d are only determined modulo N, and that we can always choose them such that $\gcd(c,d) = 1$.

Lemma 3.1.1 now implies the following.

PROPOSITION 3.1.2. *There exist bijections*

$$P^1(N) \longleftrightarrow [\Gamma : \Gamma_0(N)] \longleftrightarrow \{(M) : M \in [\Gamma : \Gamma_0(N)]\}$$

given by

$$(c : d) \leftrightarrow M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \leftrightarrow (M) = \{b/d, a/c\} \tag{3.1.2}$$

where $a, b \in \mathbb{Z}$ are chosen so that $ad - bc = 1$. $\square$

Note that a different choice of a, b in (3.1.2) has the effect of multiplying M on the left by a power of T, which does not change the right coset of M, or the symbol (M), since $T \in \Gamma_0(N)$ for all N.

The right coset action of Γ on $[\Gamma : \Gamma_0(N)]$ induces an action on $P^1(N)$:

$$(c : d) \begin{pmatrix} p & q \\ r & s \end{pmatrix} = (cp + dr : cq + ds). \tag{3.1.3}$$

The boundary map δ now takes the form

$$\delta\colon \quad (c : d) \mapsto [a/c] - [b/d]. \tag{3.1.4}$$

In order to compute $\ker(\delta)$, we must be able to determine when two cusps are $\Gamma_0(N)$-equivalent. This is achieved by the following result.

LEMMA 3.1.3. *For $i = 1, 2$ let $\alpha_i = p_i/q_i$ be cusps written in lowest terms. The following are equivalent:*

(1) $\alpha_2 = M(\alpha_1)$ *for some* $M \in \Gamma_0(N)$*;*

(2) $q_2 \equiv uq_1 \pmod{N}$ *and* $up_2 \equiv p_1 \pmod{\gcd(q_1, N)}$, *with* $\gcd(u, N) = 1$.

(3) $s_1q_2 \equiv s_2q_1 \pmod{\gcd(q_1q_2, N)}$, *where s_j satisfies $p_js_j \equiv 1 \pmod{q_j}$.*

Henceforth, we can therefore assume that $H(N)$ is given explicitly in terms of M-symbols. Certain symbols will be generators, and each M-symbol $(c : d)$ will be expressed as a $\mathbb{Q}$-linear combination of these generating symbols, by means of the 2-term relations

$$(c : d) + (-d : c) = 0 \tag{3.1.5}$$

and 3-term relations

$$(c : d) + (c + d : -c) + (d : -c - d) = 0. \tag{3.1.6}$$

When we search for Hecke eigenspaces (see next section) we will need to work in the smaller space $H^+(N)$ (the $+1$-eigenspace for the $*$-involution). This can be done by including extra relations

$$(c : d) = (-c : d) \tag{3.1.7}$$

between M-symbols, and also identifying the cusps $[\alpha]$ and $[\alpha^*]$. Effectively we are viewing $H^+(N)$ as the quotient $H(N)/H^-(N)$ when we do this. One can also view this space $H^+(N)$ as the homology of the quotient of $\mathcal{H}^*$ by the group generated by $\Gamma_0(N)$ and the matrix J of determinant -1, provided that one defines the action of matrices with negative determinant appropriately. Working in $H^+(N)$ has several advantages: we have half the number of generators (identifying M-symbols in pairs via (3.1.7)), and can also halve the number of 3-term relations between them, as the 3-term relations for $(c : d)$ and $(d : c)$ now give the same information. Geometrically, we have folded the upper half-plane in half along the imaginary half-axis, and the triangles of our triangulation then coincide in pairs.

4. Hecke operators

For each prime p not dividing N, the Hecke operator T_p acts on modular symbols $\{\alpha,\beta\}$ via

$$T_p\colon\quad \{\alpha,\beta\}\mapsto\{p\alpha,p\beta\}+\sum_{k \bmod p}\left\{\frac{\alpha+k}{p},\frac{\beta+k}{p}\right\}. \tag{4.1.1}$$

This action induces a linear map from $H(N)$ to itself, provided that p does not divide N, which we again denote by T_p.

There are also Hecke operators, which we also denote T_p, acting on the space $S_2(N)=S_2(\Gamma_0(N))$ of cusp forms of weight 2 for $\Gamma_0(N)$, defined for all primes p not dividing the level N. See one of the modular form references for the definition. The notation for the action of T_p on $S_2(N)$ is $f\mapsto f\,|T_p$.

These actions on $S_2(N)$ and $H(N)$ are compatible, in the sense that they respect the duality between cusp forms and homology:

$$\langle\{\alpha,\beta\},f\,|T_p\rangle=\langle T_p\{\alpha,\beta\},f\rangle.$$

Secondly, for each prime q dividing N there is an involution operator W_q acting on $H(N)$ and $S_2(N)$. The product of all the W_q for q dividing N is the Fricke involution W_N, coming from the transformation $z\mapsto -1/Nz$, with matrix $\begin{pmatrix}0&-1\\N&0\end{pmatrix}$.

The operators T_p for primes p not dividing N and W_q for primes q dividing N together generate a commutative $\mathbb{Q}$-algebra, called the Hecke algebra and denoted $\mathbb{T}$. Moreover, each operator is self-adjoint with respect to the so-called Petersson inner product on $S_2(N)$, and so there exist bases for $S_2(N)$ consisting of simultaneous eigenforms for all the T_p and W_q, with real eigenvalues. (See [1, Theorem 2] or [14, Corollary 2 to Theorem 4.2].) Similarly, the action of $\mathbb{T}$ on $H(N)\otimes\mathbb{R}$ can also be diagonalised.

Finally, recall from section 2.3 that the transformation $z\mapsto z^*=-\overline{z}$ on $\mathcal{H}$ also induces an involution on $H(N)$ denoted $*$. This operator commutes with all the T_p and W_q, which thus preserve the eigenspaces $H^+(N)$ and $H^-(N)$. Moreover, $H^+(N)$ and $H^-(N)$ are isomorphic as modules for the Hecke algebra $\mathbb{T}$. It follows that in order to compute eigenvalues of Hecke operators, we can restrict our attention to $H^+(N)$. This has some practical significance, as indicated above.

In order to compute the action of the Hecke operators T_p, there are several ways to proceed. We may assume that we have an explicit basis for the homology $H(N)$ given in terms of M-symbols. One approach is to convert these generating M-symbols to modular symbols $\{\alpha,\beta\}$, apply T_p to each of these, giving linear combinations of modular symbols, and then using the continued fraction method to reconvert all of these back to M-symbols. Alternatively, it is possible to compute the action of T_p directly on the M-symbols. This method is based on so-called Brandt matrices: see [17] for details. The Brandt matrix method is faster in practice, particularly if we precompute the Brandt matrices for all small primes p. Lastly, once we have found a one-dimensional eigenspace in $H^+(N)$ corrsponding to a newform f, there are tricks one can use to compute many more Hecke eigenvalues of f without having to compute the complete action of T_p on the whole space $H^+(N)$. These methods are vital in practice, where we might have to compute eigenvalues of T_p for several thousand primes p: see the subsection 5.3 and [6] for details.

5. Rational newforms and modular elliptic curves

5.1. Rational newforms.

The space $S_2(N)$ of cusp forms of weight 2 for $\Gamma_0(N)$ is a complex vector space of dimension g (the genus of $X_0(N)$); it also has a rational structure, namely a basis consisting of forms with rational Fourier coefficients. We are interested here in "rational newforms" f: that is, forms f which have rational Fourier coefficients $a(n,f)$, are simultaneous eigenforms for all the Hecke operators, and which are also "newforms" in the sense of Atkin and Lehner (see [1]). We briefly recall the definition.

For each proper divisor M of N and each $g \in S_2(M)$, the forms $g(Dz)$ for divisors D of N/M are in $S_2(N)$. The subspace $S_2^{\text{old}}(N)$ of $S_2(N)$ spanned by all such forms is called the space of *oldforms*. There is also an inner product on $S_2(N)$, called the Petersson inner product, with respect to which the Hecke operators are self-adjoint (Hermitian). Define $S_2^{\text{new}}(N)$ to be the orthogonal complement in $S_2(N)$ of $S_2^{\text{old}}(N)$ with respect to the Petersson inner product. The restriction of the Hecke algebra $\mathbb{T}$ to $S_2^{\text{new}}(N)$ is semisimple; $S_2^{\text{new}}(N)$ has a basis consisting of simultaneous eigenforms, and these eigenforms are called *newforms*.

Because of the rational structure on $S_2(N)$, the Hecke operators can be expressed explicitly as matrices with rational entries; even with integral entries, since each T_p maps the integral homology $H_1(X_0(N),\mathbb{Z})$ to itself. We can compute these matrices explicitly, and find their eigenvalues and eigenvectors using standard linear algebra.

In general, newforms come in conjugate sets of $d \geq 1$ forms with eigenvalues generating an algebraic number field of degree d. The periods of such a set of conjugates $\{f\}$ form a lattice Λ of rank $2d$ in $\mathbb{C}^d$, and hence an abelian variety $A_f = \mathbb{C}^d/\Lambda$, which is defined over $\mathbb{Q}$. Here we will only be interested in the case $d = 1$, where the Hecke eigenvalues and hence Fourier coefficients of f are rational (in fact integers, being eigenvalues of integral matrices and hence algebraic integers). We will call such a form f a *rational newform*. Thus a rational newform f has an associated period lattice Λ_f:

$$\Lambda_f = \{\langle\{\alpha,\beta\},f\rangle \mid \alpha,\beta \in \mathcal{H}^*, \alpha \equiv \beta \pmod{\Gamma_0(N)}\}$$

which is a discrete rank 2 subgroup of $\mathbb{C}$. Then $E_f = \mathbb{C}/\Lambda_f$ is an elliptic curve, the modular elliptic curve attached to f. Moreover it is known that E_f is defined over $\mathbb{Q}$, has conductor N, and has L-series $L(E_f,s) = \sum a(n,f)n^{-s}$ where $f = \sum a(n,f)\exp(2\pi i n z)$. (See [20], [19], [3].)

5.2. Fourier coefficients and L-series.

The Fourier coefficients $a(n,f)$ of a newform $f(z) = \sum a(n,f)\exp(2\pi i n z)$ are obtained from the Hecke eigenvalues of f as follows (see [1]). Firstly, for a newform f we always have $a(1,f) \neq 0$, and we normalise so that $a(1,f) = 1$. Then:

If p is a prime not dividing N, and $f\,|T_p = a_p f$, then $a(p,f) = a_p$. By standard estimates, the integer a_p satisfies $|a_p| < 2\sqrt{p}$.

If q is a prime dividing N, and $f\,|W_q = \varepsilon_q f$ with $\varepsilon_q = \pm 1$, then

$$a(q,f) = \begin{cases} -\varepsilon_q & \text{if } q^2 \nmid N, \\ 0 & \text{if } q^2 | N. \end{cases} \tag{5.2.1}$$

For prime powers, we have the recurrence relation

$$a(p^{r+1},f) = a(p,f)a(p^r,f) - \delta_N(p)pa(p^{r-1},f) \qquad (r \geq 1) \tag{5.2.2}$$

where

$$\delta_N(p) = \begin{cases} 1 & \text{if } p \nmid N, \\ 0 & \text{if } p | N. \end{cases}$$

Finally, for composite indices we have multiplicativity: $a(mn,f) = a(m,f)a(n,f)$ when m and n are coprime.

The sign of the functional equation of the L-series $L(f,s)$ attached to the newform f is *minus* the eigenvalue ε_N of the Fricke involution W_N; in the above notation, it is $-\prod_{q|N}\varepsilon_q$. Explicitly, we have

$$\Lambda(f, 2-s) = -\varepsilon_N \Lambda(f,s),$$

where

$$\Lambda(f,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s) = \int_0^\infty f(iy/\sqrt{N})y^{s-1}dy.$$

Also, for $\Re(s) > 3/2$ we have the Dirichlet series and Euler product expansions of $L(f,s)$:

$$\begin{aligned} L(f,s) &= \sum_{n=1}^{\infty} \frac{a(n,f)}{n^s} \\ &= \prod_{p\nmid N}\left(1 - a(p,f)p^{-s} + p^{1-2s}\right)^{-1} \prod_{q|N}\left(1 - a(q,f)q^{-s}\right)^{-1}. \end{aligned}$$

The sign of the functional equation tells us the parity of the *analytic rank* of the elliptic curve E_f, which is (by definition) the order of $L(f,s)$ at $s=1$. By the Birch–Swinnerton-Dyer conjecture (BSD), this should equal the rank of the Mordell-Weil group $E_f(\mathbb{Q})$. If the analytic rank is odd, then obviously $L(f,1) = 0$, and we expect $E_f(\mathbb{Q})$ to be infinite: indeed, this has been proved in the rank 1 case by Kolyvagin *et al.* (see [11]). In the next subsection we will see how to determine whether or not $L(f,1) = 0$ in all cases.

5.3. The ratio $L(f,1)/\Omega(f)$.

The real periods of the rational newform f are the periods $\langle\gamma, f\rangle$ where $\gamma \in H_1^+(X_0(N),\mathbb{Z})$; these are all integer multiples of the least positive real period $\Omega_0(f)$. We set

$$\Omega(f) = \begin{cases} 2\Omega_0(f) & \text{if the period lattice of } f \text{ is rectangular,} \\ \Omega_0(f) & \text{otherwise.} \end{cases}$$

Thus $\Omega(f)/\Omega_0(f)$ is the number of components of the real locus of the elliptic curve E_f. Also note that in each case, $\Omega(f)$ is twice the least real part of a period of f. This is useful since, when we are working in $H^+(N)$, we can only (at that stage) determine the projection of the period lattice Λ_f onto the real axis.

A simple calculation with modular symbols shows that, for a prime p not dividing N, we have

$$(1+p-T_p)\cdot\{0,\infty\}=\sum_{k=0}^{p-1}\{0,k/p\}. \tag{5.3.1}$$

Integrating $2\pi i f(z)dz$ along both sides of (5.3.1) gives

$$(1+p-a_p)\cdot L(f,1)=-\sum_{k=0}^{p-1}\langle\{0,k/p\},f\rangle\,. \tag{5.3.2}$$

where $a_p=a(p,f)$, and we have the used the fact that $L(f,1)=-\langle\{0,\infty\},f\rangle$. Since p does not divide N, each modular symbol $\{0,k/p\}$ on the right of (5.3.1) is *integral*: that is, in $H_1(X_0(N),\mathbb{Z})$. Thus the right-hand side of (5.3.2) is a period of f. It is even a real period, since the right-hand side of (5.3.1) is $*$-invariant, hence equal to $n(p,f)\Omega(f)/2$ for some integer $n(p,f)$. Hence we have the relation

$$\frac{L(f,1)}{\Omega(f)}=\frac{n(p,f)}{2(1+p-a_p)}. \tag{5.3.3}$$

Note that $1+p-a_p>0$, since $|a_p|<2\sqrt{p}$.

In practice the integer $n(p,f)$ is computed as follows. Corresponding to the rational newform f there is a one-dimensional eigenspace in $H^+(N)$ (which we may identify with the column space $\mathbb{Q}^g$ after fixing a basis). Let v_f be a *dual* eigenvector with the same eigenvalues of f, *i.e.*, a row vector which is a left eigenvector for all the operator matrices; normalise v_f to lie in $\mathbb{Z}^g$ and have coprime entries (this determines v_f uniquely up to sign). Let u_p be a column vector in $\mathbb{Z}^g$ expressing (minus) the right-hand side of (5.3.1) in terms of our basis for $H^+(N)$. Then $n(p,f)$ is simply the dot product $v_f\cdot u_p$. If necessary, we may replace v_f by $-v_f$ in order to make both sides of (5.3.3) positive.

Formula (5.3.3) is significant in several ways. On the one hand, let E_f be the modular elliptic curve attached to f as above. Then $L(E_f,1)=L(f,1)$, and $\Omega_0(f)=\Omega_0(E_f)$, the least positive real period of E_f. Thus, once we know a_p and $n(p,f)$ for a single prime p, we can evaluate the rational number $L(E_f,1)/\Omega(E_f)$, whose value is predicted by the Birch–Swinnerton-Dyer conjecture (BSD) for E_f. We stress that we have obtained this rational number exactly, algebraically; in particular, we know for certain whether or not $L(f,1)=0$, without any approximation necessary.

Secondly, having computed the right-hand side of (5.3.3) for a single prime p, we may (if $L(f,1)\neq 0$) use the fact that $n(p,f)/(1+p-a_p)$ is independent of p to compute the eigenvalue a_p quickly for other p, by computing $n(p,f)$. This is how we compute a large number of Hecke eigenvalues (and hence Fourier coefficients) in practice. A variation of this trick can be used when $L(f,1)=0$: see [6] for details.

5.4. Finding newforms.

With this background we may now make more precise what we mean by "computing the modular elliptic curves of conductor N". We do the following:

(1) Compute the space $H^+(N)$ in terms of M-symbols and their relations.

(2) Compute the action of sufficiently many Hecke operators W_q and T_p on $H^+(N)$ to determine the one-dimensional eigenspaces with rational eigenvalues; by duality, we now know the rational newforms in $S_2(N)$. (Oldforms can be recognised, since in any systematic computation we will have already found them at some lower level M dividing N.)
(3) Find a $\mathbb{Z}$-basis for the period lattice Λ_f, for each rational newform f, computing the generating periods to high precision.
(4) Given a $\mathbb{Z}$-basis for Λ_f, compute the coefficients of an equation for the attached elliptic curve E_f.
(5) As well as the period lattice of the curves E_f, we can also compute the rational number $L(E_f, 1)/\Omega(E_f)$ (exactly) and the real value $L(E_f, 1)$ (approximately). Also, when $L(E_f, 1) = 0$ we can also determine the order of vanishing of $L(E_f, s)$ at $s = 1$, giving the analytic rank r of E_f, and the value of the derivative $L^{(r)}(E_f, 1)$, which is important in the Birch–Swinnerton-Dyer conjecture; we will not discuss the latter computation here, except briefly to give the relevant formulas in a later section.

We would like to make completely explicit exactly what we mean when we say that for a given level N we "know" the rational newforms in $S_2(N)$. Our knowledge is both algebraic and analytic; also, it is important to understand exactly how much information we can obtain from working in $H^+(N)$, and how much requires the full space $H(N)$ to be available.

The first phase of our computations works in $H^+(N)$, and is entirely algebraic. In this phase, we find all the 1-dimensional eigenspaces in $H^+(N)$ for all the Hecke operators (which only involve computing the full action of all the W_q and of the T_p for a few small primes p), from which we know how many rational newforms there are, and hence how many modular elliptic curves of conductor N there are, up to isogeny. For each newform, we also know the first few Euler factors of its L-series $L(f, s)$, the sign of its functional equation and also whether or not the L-series vanishes at $s = 1$. This tells us partial information about the analytic rank of the associated curve E_f: we know its parity and whether or not it is zero. (The first case where the rank is not equal to the minimal value consistent with this information is at level 5077, where we know from these algebraic considerations that the rank is odd, but where the curve turns out to have rank 3. Determining that the analytic rank is 3 rather than 1 in this case is not a purely algebraic computation, as it involves proving that the first derivative of the L-series vanishes at $s = 1$.)

In this first phase, we can also compute approximations to the value $L(f, 1)$ (when it is non-zero) and to the period $\Omega(f)$, though we do not know at this stage whether $\Omega(f)/\Omega_0(f) = 1$ or 2. In other words, we can compute the projection of the period lattice Λ_f onto the real axis. Of course, this is insufficient information from which to construct the curve E_f.

These "analytic" quantities (periods) will necessarily be computed approximately, by summing certain infinite series whose coefficients involve the Fourier coefficients of f (see the next section). In order to achieve sufficient accuracy we may have to compute many thousands of these Fourier coefficients, and it is therefore necessary to have efficient ways of doing this, such as the method described above (in section 5.3).

To determine the full period lattice for each newform f, it is necessary to compute the full $2g$-dimensional space $H(N)$. This is what we call the second phase of the computation. In $H(N)$ we find, for each rational newform f, the two-dimensional eigenspace associated to f (which is split into two one-dimensional eigenspaces by the $*$-involution). Once we have this two-dimensional eigenspace, it is a simple matter of linear algebra to determine elements $\gamma_1, \gamma_2 \in H_1(X_0(N), \mathbb{Z})$ such that the periods $\omega_i = \langle \gamma_i, f \rangle$ are a $\mathbb{Z}$-basis for the period lattice Λ_f. Then we may compute these generating periods, again approximately. From ω_1 and ω_2 we can compute approximations to the c_4 and c_6 invariants of the elliptic curve E_f. These are known to be rational (since E_f is defined over $\mathbb{Q}$). Moreover, recent work of Edixhoven [9] on the Manin constant implies that these computed values of c_4 and c_6 are *integral*, which is of course crucial if we are to recognise them from floating point approximations. Given the integers c_4 and c_6, it is then easy to find a standard Weierstrass equation for the curve E_f.

For large N (say over 3000), we have found this second phase (working in $H(N)$) to be very costly in computation time and in storage requirements. Hence we have been led to develop an indirect approach to determining the period lattice of each rational newform f. This new approach is extremely fast, but does have one critical disadvantage: we can (in most cases) determine a period lattice Λ'_f which is a sublattice of finite index in the full period lattice Λ_f, and such that $\mathbb{C}/\Lambda'_f$ is an elliptic curve E'_f of conductor N with $L(E'_f, s) = L(f, s)$, but we do not know that $\Lambda'_f = \Lambda_f$, so that E'_f might only be isogenous to the curve E_f. Secondly, we can only use this method in case N is not a perfect square, since we have to be able to choose primes $l \equiv \pm 1 \pmod 4$ with $\left(\frac{-N}{l}\right)$ of prescribed sign (see next section).

This new method relies on the fact that the values $L(f \otimes \chi, 1)$ for quadratic characters χ (with conductor coprime to N) are, on the one hand, easy to compute to high precision (given enough Fourier coefficients of the newform f); and on the other hand, they are simple multiples of integral periods of f. In our original approach we used modular symbols in $H(N)$ to determine these multiples exactly and explicitly, so that two computations of $L(f \otimes \chi, 1)$ for suitable characters χ were enough to determine the period lattice. We will discuss this situation, and the computation of steps (3) and (4), in greater detail in section 7 below.

6. Analytic formulas

In this section we collect together all the analytic formulas we need for evaluation of periods and special values of L-series. For brevity, proofs will be sketched or omitted; see [6] for details.

We keep the notation of previous sections: f will always be a rational newform for $\Gamma_0(N)$, with Fourier coefficients $a(n, f)$, normalised so that $a(1, f) = 1$.

6.1. Computing $L(f, 1)$ and $L(f \otimes \chi, 1)$.

Recall from section 5.2 that the eigenvalue ε_N of the Fricke involution is minus the sign of the functional equation of $L(f, s)$, so that when $\varepsilon_N = +1$ we have $L(f, 1) = 0$ trivially (in this case the analytic rank is odd, hence trivially positive).

When $\varepsilon_N = -1$, we use the formula

$$L(f,1) = 2\sum_{n=1}^{\infty} \frac{a(n,f)}{n} \exp(-2\pi n/\sqrt{N}). \tag{6.1.1}$$

Essentially, $L(f,1)$ is the integral of f along the imaginary axis; when $\varepsilon_N = -1$ this integral is expressed as double the integral from $i/\sqrt{N}$ to $i\infty$.

The formula (5.3.3) we derived above, which expresses $L(f,1)$ as an explicit multiple of the real period $\Omega(f)$, can be generalised to useful effect. Let l be an odd prime not dividing N, and χ the quadratic character modulo l. Define

$$(f \otimes \chi)(z) = \sum_{n=1}^{\infty} \chi(n)a(n,f)\exp(2\pi i n z).$$

Then $f \otimes \chi$ is a newform in $S_2(Nl^2)$, whose Fricke eigenvalue $\varepsilon_{Nl^2}(f\otimes\chi)$ is equal to $\chi(-N)\varepsilon_N(f)$. Hence for $\chi(-N) = \varepsilon_N$ we have $L(f\otimes\chi,1) = 0$ trivially, while if $\chi(-N) = -\varepsilon_N$ we have

$$L(f\otimes\chi,1) = 2\sum_{n=1}^{\infty} \frac{\chi(n)a(n,f)}{n} \exp(-2\pi n/l\sqrt{N}). \tag{6.1.2}$$

Set $l^* = \chi(-1)l \equiv 1 \pmod 4$, and $P(l,f) = \sqrt{l^*}L(f\otimes\chi,1)$. The generalization of (5.3.3) we need is

$$P(l,f) = \langle \gamma_l, f\rangle$$

where

$$\gamma_l = \sum_{k=0}^{l-1} \chi(k)\,\{0, k/l\}\,.$$

Since l does not divide N, the cycle γ_l is in the integral homology, so $P(l,f)$ is an integral period for each l. Also, it is easy to see that $\gamma_l^* = \chi(-1)\gamma_l$, so that this period $P(l,f)$ is real or imaginary according as $\chi(-1) = +1$ or -1, *i.e.*, according as $l \equiv \pm 1 \pmod 4$. Moreover, by expressing the cycle γ_l in terms of our fixed basis for $H_1(X_0(N),\mathbb{Z})$, we may determine explicit integers $m^{\pm}(l,f)$ (where $\pm 1 = \chi(-1)$) such that if x and y denote the least positive real and imaginary parts of periods of f respectively, then $P(l,f) = m^+(l,f)x$ or $m^-(l,f)yi$.

It is now clear that we will be able to express the period lattice for f in terms of the quantities $P(l,f)$, provided that we can find auxiliary primes l of either sign such that $\chi(-N) = -\varepsilon_N$. This is always possible in practice (and guaranteed theoretically by a theorem of Murty and Murty [18]), provided that N is not a perfect square. In this exceptional case, we compute the periods of f more directly (see subsection 6.3 below).

6.2. Computing $L^{(r)}(f,1)$ for $r \ge 1$.

Let f be a newform in $S_2(N)$ with W_N-eigenvalue ε_N, and suppose that the order of $L(f,s)$ at $s=1$ is at least r, where $\varepsilon_N = (-1)^{r-1}$. Then

$$L^{(r)}(f,1) = 2(r!)\sum_{n=1}^{\infty} \frac{a(n,f)}{n} G_r\left(\frac{2\pi n}{\sqrt{N}}\right) \tag{6.2.1}$$

where
$$G_r(x) = \frac{1}{(r-1)!}\int_1^\infty e^{-xy}(\log y)^{r-1}\frac{dy}{y}.$$

The functions $G_r(x)$ which appear here satisfy the functional equations $G_r'(x) = (-1/x)G_{r-1}(x)$, with $G_0(x) = e^{-x}$. For $r = 1$, the function $G_1(x) = \int_1^\infty e^{-xy}dy/y$, the exponential integral. For good ways of evaluating this, and for higher values of r, see [6] or the paper [2], where this formula was used to evaluate $L^{(3)}(f, 1)$ for the newform $f \in S_2(5077)$ attached to the elliptic curve of rank 3 and conductor 5077.

6.3. Computing periods directly.

To compute integral periods of a newform f, we may use essentially the same method as in [22]. It suffices to compute $\langle\gamma, f\rangle$ for a single cycle γ such that the real and imaginary parts are both nonzero; then by taking real and imaginary parts, and by expressing γ as a linear combination of our basis for $H_1(X_0(N), \mathbb{Z})$, we may determine the generating periods ω_1 and ω_2. (In some cases it may give better accuracy in practice to use two different cycles, one for the real period and one for the imaginary period, but for simplicity we will assume that this is not the case.)

Let $M = \begin{pmatrix} a & b \\ Nc & d \end{pmatrix} \in \Gamma_0(N)$, with $c > 0$. Set $y_0 = 1/(Nc)$, $x_1 = -dy_0$, and $x_2 = ay_0$, and let $\gamma = \{0, M(0)\} \in H(N)$. Then

$$\langle\gamma, f\rangle = \sum_{n=1}^{\infty} \frac{a(n, f)}{n} \exp(-2\pi n y_0)\left(\exp(2\pi i n x_2) - \exp(2\pi i n x_1)\right). \tag{6.3.1}$$

In practice we will try to choose M so that c is as small (and hence y_0 as large) as possible, to speed convergence in the series (6.3.1).

Observe that the convergence in (6.3.1) will not be so fast as in (6.1.1), since we now have a power series in $\exp(-2\pi/(Nc))$ for an integer $c \geq 1$, instead of in $\exp(-2\pi/\sqrt{N})$. Similarly, (6.1.2) will usually be better than (6.3.1), unless the auxiliary prime l is greater than $c\sqrt{N}$.

7. From newform to lattice to curve

7.1. Finding the period lattice.

Let f be, as above, a rational newform in $S_2(N)$. We wish to determine a $\mathbb{Z}$-basis for the period lattice Λ_f. We work in the $2g$-dimensional space $H(N)$, identified with the column space $\mathbb{Q}^{2g}$ by fixing a basis for $H(N)$ which is also a $\mathbb{Z}$-basis for the integral homology $H(N, \mathbb{Z}) := H_1(X_0(N), \mathbb{Z})$.

Let v^+ and v^- be dual eigenvectors corresponding to f, i.e. row vectors in the row-space $\mathbb{Q}^{2g}$ which are left eigenvectors for the Hecke algebra $\mathbb{T}$ with the same eigenvalues as f; we also suppose that $v^\pm$ are eigenvectors for the $*$-involution with eigenvalues ± 1 respectively. We may normalise $v^\pm$ so that each is in $\mathbb{Z}^{2g}$ and primitive (that is, with coprime entries); then each is uniquely determined up to sign.

There exist positive real numbers x and y such that for each cycle γ in $H(N, \mathbb{Z})$, expressed as a column vector in $\mathbb{Z}^{2g}$, we have

$$\langle\gamma, f\rangle = (v^+\gamma)x + (v^-\gamma)yi.$$

Here the products $v^{\pm}\gamma$ are scalar products (row vector times column vector). There are two possibilities for the period lattice Λ_f: if $v^+ \equiv v^- \pmod{2}$, then $\Lambda_f = \langle 2x, x+yi\rangle$; otherwise $\Lambda_f = \langle x, yi\rangle$. We call these two cases "Type 1" and "Type 2" respectively.

The "direct" method of computing the lattice Λ_f is to compute $\langle\gamma, f\rangle$ for a cycle γ such that $v^{\pm}\gamma$ are both non-zero, using the formula (6.3.1). This is the only method we can use when N is a perfect square; otherwise we obtain better precision by obtaining the periods indirectly, using special values of twisted L-series.

Recall from section 6.1 that for each odd prime l not dividing the level N we have defined integral periods

$$P(l,f) = \sqrt{l^*}L(f\otimes\chi,1) = \langle\gamma_l, f\rangle,$$

where $l^* = \pm l \equiv 1 \pmod{4}$ and $\gamma_l = \sum_{a \pmod{l}} \chi(a)\{0, a/l\}$. If $l \equiv 1 \pmod{4}$ then $v^-\gamma_l = 0$, since $\gamma_l \in H^+(N)$, and now the integer $m^+(l,f) := v^+\gamma_l$ satisfies

$$P(l,f) = m^+(l,f)x;$$

Secondly, when $l \equiv -1 \pmod{4}$ we have $v^+\gamma_l = 0$, since $\gamma_l \in H^-(N)$, and the integer $m^-(l,f) := v^-\gamma_l$ satisfies

$$P(l,f) = m^-(l,f)yi.$$

Hence to compute x and y, and so determine the period lattice, we simply have to find primes $l^{\pm} \equiv \pm 1 \pmod{4}$ such that the integers $m^{\pm} = v^{\pm}\gamma_{l^{\pm}}$ are non-zero, and set $x = P(l^+,f)/m^+$ and $yi = P(l^-,f)/m^-$.

We stress that, apart from the numerical evaluation of the periods $P(l^{\pm},f)$ (using the series for $L(f\otimes\chi,1)$ in section 6.1), all these computations are purely algebraic: we express the cycles γ_l in terms of our homology basis using continued fractions, and take the dot products of the resulting column vectors with our dual eigenvectors $v^{\pm}$ to obtain the integers $m^{\pm}$.

The result of this algebraic computation consists of the following data for each rational newform f: primes $l^{\pm}$ congruent respectively to ± 1 modulo 4; nonzero integers $m^{\pm}$; and the type (1 or 2) of the lattice. To compute the lattice from this data set of five integers, we compute the periods $P(l^{\pm},f)$ using formula (6.1.2), divide by $m^{\pm}$ respectively to obtain x and y, and take Λ_f to be the lattice with $\mathbb{Z}$-basis $2x$, $x+yi$ (if type 1) or x, yi (if type 2). In practice we store just these five integers (with the Hecke eigenvalues from which we can obtain the Fourier coefficients), and recompute the periods when we need them. In particular, if at the first attempt we are unable to compute the integer invariants c_4, c_6 of the curve E_f to sufficient precision to recognise them, then we will return to $H^+(N)$ in order to compute more Hecke eigenvalues, and then recompute the periods to greater precision without having to recompute $H(N)$.

In fact, the data l^+ and m^+ can be computed earlier in the first $H^+(N)$ phase, since they only depend on the real projection of the period lattice. Hence we can already compute the real period x from the data we have from the first phase. Moreover, it is easy to find a suitable prime l^- once we know the Hecke eigenvalues of f, by numerically computing $P(l, f)$ for several primes $l \equiv -1 \pmod 4$ until we find a value which is clearly non-zero.

It follows that the only purpose of the extremely expensive second phase of the computation, working in $H(N)$, is to determine the integer factor m^- and the type of the lattice. Our new method, which we have used systematically for larger levels ($N > 3200$), is simply to guess the value of m^- by trying each positive integer m in turn. For each $m \geq 1$ we set $yi = P(l^-, f)/m$ and test the two possible lattices (one of each type). If either lattice has approximate integer invariants c_4 and c_6, and the rounded integral values are valid invariants of an elliptic curve over $\mathbb{Q}$, and the resulting curve has conductor N, then we store for later use the successful value m^- of m and the type, and consider the curve E'_f we have found as a possible candidate for the actual modular elliptic curve E_f.

As we pointed out above, the curves E_f and E'_f are certainly isogenous; they even have the same real period. In many cases, the curve E'_f has no rational isogenies; in such a case we can conclude that $E_f = E'_f$ with no ambiguity. In any case, we can compute the isogeny class of curves isogenous to E'_f via rational isogenies, and the only loss is that we do not always know exactly which curve in the class is the "strong Weil curve" E_f. (A further disadvantage is that we cannot compute the degree of the modular parametrisation of E_f, as this requires knowledge of $H(N)$: see section 10 below.)

The huge advantage of this method is that in only a few seconds computation time, as soon as we have a rational newform, we can (almost always) write down an associated curve E'_f; before this was implemented, it could take many hours of computation time to determine $H(N)$, find the eigenvectors $v^\pm$, and hence determine the factor m^- and the lattice type, before we could compute E_f.

For an example of this method (with $N = 11$) see the next section.

Finally we discuss some variants of the trick just described.

1. We may use same trick to find l^+ and m^+ if we have not computed them earlier. Then we are obtaining the period lattice and equation of the curve using *only*: the Fourier coefficients of f (i.e. the coefficients of the L-series of the curve); the sign of the functional equation; and the conductor N. No modular symbol information at all is needed in this case. In fact, one may even guess the sign of the functional equation and compute the conductor if all one has is the L-series.

2. Let l_1 and l_2 be two primes $\equiv -1 \pmod 4$ for which $-N$ has the correct quadratic character, so that $P(l_1, f)$ and $P(l_2, f)$ are both not trivially zero. We may compute the periods $P(l_j, f)$; assume that these are nonzero (or use different primes l_j). We know that there exist nonzero integers m_j such that $P(l_j, f) = m_j yi$ for $j = 1, 2$. Therefore $P(l_2, f)/P(l_1, f) = m_2/m_1$, and we may compute a floating point approximation to this rational number. In practice (provided we have many Fourier coefficients, and the primes l_j are small) we will be able to recognise this rational number using continued fractions. Its denominator is then a factor of the unknown integer m_1. If we do this for several different values of l_2 (with the same l_1) then the least common multiple of the denominators will often give us a useful factor of m_1, and then in our search for the exact value we may restrict to multiples of this factor. This is useful in practice.

3. Another possibility, which we have not yet implemented, is to compute $H^-(N)$ in order to determine m^- exactly, as we do m^+ from $H^+(N)$. This would be no harder than the original computation of $H^+(N)$, and in fact it would be easier to find the eigenvector corresponding to each newform f, since we already know its eigenvalues. The result would be that we would have computed exactly all the data we need in a shorter time than would be required for computing $H(N)$, *except* for the type of the lattice. Then the only ambiguity is that we would not know the type, and would have to try both types to see which results in a curve with integral coefficients. If both types succeeded (as does happen), we would only know the curve E_f up to a 2-isogeny.

7.2. Finding the equation of the curve.

So far we have described how to compute, to a certain precision, the periods ω_1 and ω_2 which generate the period lattice Λ_f of the modular curve $E_f = \mathbb{C}/\Lambda_f$ attached to each rational newform f in $S_2(N)$. Now we turn to the question of finding an equation for E_f.

Set $\tau = \omega_1/\omega_2$. Interchanging ω_1 and ω_2 if necessary, we may assume that $\mathrm{Im}(\tau) > 0$. By applying the well-known algorithm for moving a point in the upper half-plane $\mathcal{H}$ into the usual fundamental region for $SL(2,\mathbb{Z})$ we may assume that $|\mathrm{Re}(\tau)| \le 1/2$ and $|\tau| \ge 1$, so that $\mathrm{Im}(\tau) \ge \sqrt{3}/2$. One merely replaces (ω_1, ω_2) by $(\omega_1 - n\omega_2, \omega_2)$ for suitable $n \in \mathbb{Z}$ and (ω_1, ω_2) by $(-\omega_2, \omega_1)$ until both conditions are satisfied. In practice one must be careful about rounding errors, as it is quite possible to have both $|\tau| < 1$ and $|-1/\tau| < 1$ after rounding, which is liable to prevent the algorithm from terminating.

Set $q = \exp(2\pi i \tau)$. Then the lattice invariants c_4 ($= 12g_2$) and c_6 ($= 216g_3$) are given by

(7.2.1)

$$c_4 = \left(\frac{2\pi}{\omega_2}\right)^4 \left(1 + 240\sum_{n=1}^{\infty} \frac{n^3 q^n}{1-q^n}\right) \quad \text{and} \quad c_6 = \left(\frac{2\pi}{\omega_2}\right)^6 \left(1 - 504\sum_{n=1}^{\infty} \frac{n^5 q^n}{1-q^n}\right)$$

(see, for example, [13, p.47]). Since $|q| = \exp(-2\pi\mathrm{Im}(\tau)) \le \exp(-\pi\sqrt{3}) < 0 \cdot 005$, these series converge extremely rapidly. Thus, assuming that ω_1 and ω_2 are known to sufficient precision, we can compute c_4 and c_6 as precisely as required.

Since E_f is defined over $\mathbb{Q}$, the numbers c_4 and c_6 are rational, but there is no *a priori* reason why they should be integral. Fortunately, a result of Edixhoven [9] shows that they are integral. Hence, provided that we have computed the periods and then c_4 and c_6 to sufficient precision, we will be able to recognise the corresponding exact integer values.

This only presents practical difficulties when c_4 and c_6 are large, since to date our programs use standard double precision, which yields around 16 decimal places. In several cases this meant that we could recognise c_4, but the last digit or digits of c_6 were undetermined. In these situations, we are helped by the fact that we know that c_4 and c_6 are the invariants of an elliptic curve of conductor N. This implies the following congruence conditions (see [12], [5] or [6]):

(1)

$$c_4^3 - c_6^2 = 1728\Delta$$

where Δ is a non-zero integer divisible by the primes dividing N;

(2) for primes $p \geq 5$ which divide N, we have

$$p \mid c_4 \iff p \mid c_6 \iff p^2 \mid N;$$

(3) $c_6 \not\equiv 9 \pmod{27}$;
(4) either $c_6 \equiv -1 \pmod 4$, or $c_4 \equiv 0 \pmod{16}$ and $c_6 \equiv 0, 8 \pmod{32}$.

Note that in condition (1) we should not assume that Δ is only divisible by the "bad primes" which divide N, since we do not know that c_4 and c_6 are the invariants of a minimal model. However, Edixhoven's result does bound the non-minimality, and in practice all the equations of curves we have constructed are minimal, verifying the conjecture (Manin's "$c = 1$" conjecture) that this should always be the case. Conditions (2)–(4) do assume minimality at the relevant primes.

Since c_4 is smaller than c_6, the usual situation is that we know c_4, but we use the above congruence conditions to help us find c_6 in case it has more than 16 digits.

Given integral invariants c_4, c_6 satisfying (1), (3) and (4) above the coefficients $[a_1, a_2, a_3, a_4, a_6]$ of a standard Weierstrass equation for the curve may be obtained as follows (see [6]):

$$\begin{aligned}
b_2 &= -c_6 \mod 12 \in \{-5, \dots, 6\};\\
b_4 &= (b_2^2 - c_4)/24;\\
b_6 &= (-b_2^3 + 36b_2b_4 - c_6)/216;\\
a_1 &= b_2 \mod 2 \in \{0, 1\};\\
a_3 &= b_6 \mod 2 \in \{0, 1\};\\
a_2 &= (b_2 - a_1^2)/4;\\
a_4 &= (b_4 - a_1a_3)/2;\\
a_6 &= (b_6 - a_3^2)/4;
\end{aligned}$$

where all divisions are exact. We may then compute the conductor of the curve $E = [a_1, a_2, a_3, a_4, a_6]$ using Tate's algorithm ([21], [6]), to check that it is equal to N. We may also compute its rank to compare with the analytic rank. Finally, we can compute all curves isogenous to E over $\mathbb{Q}$: see [6] for one way to do this. This final list of curves will, according to the Shimura-Taniyama-Weil conjectures, contain all elliptic curves defined over $\mathbb{Q}$ with conductor N (up to isomorphism). At the time of writing (April 1995) this has been proved (by Wiles and Taylor, following Ribet, Frey and others) provided that N is divisible by neither 4 nor 25.

8. Worked example: $N = 11$

We give here a worked example of the methods described in the preceding section. The level chosen is $N = 11$, which is the first non-trivial level. As 11 is prime, this is rather special: the M-symbols are easy to work with (there are only 12 of them), there are no oldforms, the value of $L(f, 1)$ is non-zero so that mass production of Hecke eigenvalues is easy, and as the level is not square our short cut for finding the period lattice will work. Nevertheless, it is instructive to see how much information it is possible to obtain here, with only a page or two of essentially hand computation (with the help of a pocket calculator at the end).

For simplicity we will only work in $H(11)$, rather than the smaller quotient space $H^+(11)$. The M-symbols for $N = 11$ are $(c : 1)$ for c modulo 11 and $(1 : 0)$, which we abbreviate as (c) and (∞) respectively, with $|c| \le 5$. (Similarly with other prime levels). The 2-term and 3-term relations are as follows.

$$\begin{aligned} (0) + (\infty) &= 0 \\ (1) + (-1) &= 0 \\ (2) + (5) &= 0 \\ (-2) + (-5) &= 0 \\ (3) + (-4) &= 0 \\ (-3) + (4) &= 0 \end{aligned} \qquad \begin{aligned} (0) + (\infty) + (-1) &= 0 \\ (1) + (-2) + (5) &= 0 \\ (2) + (4) + (-4) &= 0 \\ (3) + (-5) + (-3) &= 0 \end{aligned}$$

Solving these equations we can express all 12 symbols in terms of $A = (2)$, $B = (3)$ and $C = (0)$:

$$\begin{aligned} (0) &= C \\ (\infty) &= -C \\ (1) = (-1) &= 0 \\ (2) = (-2) &= A \\ (5) = (-5) &= -A \end{aligned} \qquad \begin{aligned} (3) &= B \\ (-3) &= A - B \\ (4) &= B - A \\ (-4) &= -B \end{aligned}$$

There are two classes of cusps, $[0]$ and $[\infty]$, with $[a/b] = [0]$ if $11 \nmid b$ and $[a/b] = [\infty]$ if $11 \mid b$. Hence $\delta((c)) = \delta(\{0, 1/c\}) = [1/c] - [0] = 0$ for $c \not\equiv 0$. It follows that

$$H(11) = \ker(\delta) = \langle A, B \rangle ,$$

with $2g = \dim H(11) = 2$, so that the genus is 1. There is therefore one newform f. This makes the rest of the calculation simpler, as we do not have to find and split off eigenspaces.

The conjugation $*$ involution maps $(c) \mapsto (-c)$, so $A^* = A$ and $B^* = A - B$. This has matrix $\begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$ with respect to the basis A, B. The $+1$- and -1-eigenspaces are generated by A and $A - 2B$ respectively, and we have left eigenvectors $v^+ = (2, 1)$ and $v^- = (0, 1)$. Thus the period lattice is non-rectangular, and $\Omega(f) = \Omega_0(f) = \langle A, f \rangle$.

If we had worked in $H^+(11)$, viewed as the quotient $H(11)/H^-(11)$, by including relations $(c) = (-c)$, the effect would be to identify (c) and $(-c)$. This gives a 1-dimensional space generated by $\overline{B}$ with $\overline{A} = 2\overline{B}$, where the bars denote the projections to the quotient. Notice that although $\overline{B}$ is a generator here, the integral of f over B is not a real period; its real part is half the real period. However we do still have $\Omega(f) = \langle B + B^*, f \rangle = 2\mathrm{Re}\,\langle B, f \rangle$, so we could compute $\Omega(f)$ in this context without actually knowing whether it was 1 or 2 times the smallest real period.

To compute Hecke eigenvalues we may work in the subspace $\langle A \rangle$; since this subspace is conjugation invariant (being the $+1$-eigenspace) we will have $T_p(A) =$

a_pA for all $p \neq 11$. We first compute T_2 explicitly, converting the M-symbol $A = (2:1)$ to the modular symbol $\{0, 1/2\}$:

$$\begin{aligned}
T_2(A) = T_2\left(\left\{0, \frac{1}{2}\right\}\right) &= \{0,1\} + \left\{0, \frac{1}{4}\right\} + \left\{\frac{1}{2}, \frac{3}{4}\right\} \\
&= \{0,1\} + \left\{0, \frac{1}{4}\right\} + \left\{\frac{1}{2}, 1\right\} + \left\{1, \frac{3}{4}\right\} \\
&= (1:1) + (4:1) + (1:2) + (-4:1) \\
&= (0) + (4) + (-5) + (-4) \\
&= 0 + (B - A) + (-A) + (-B) = -2A,
\end{aligned}$$

so that $a_2 = -2$. Now $(1 + 2 - a_2)L(f,1) = \langle \{0, 1/2\}, f\rangle = \langle A, f\rangle = \Omega(f)$, giving

$$\frac{L(f,1)}{\Omega(f)} = \frac{1}{5}.$$

For all primes $p \neq 11$ we will evaluate $\mu_p = \sum_{a=0}^{p-1} \{0, a/p\} = n_pA$ for a certain integer n_p, since then also $1/5 = n_p/(1 + p - a_p)$, giving

$$a_p = 1 + p - 5n_p.$$

At this stage we already know that the corresponding elliptic curve has rank 0, and that $1+p-a_p \equiv 0 \pmod 5$ for all $p \neq 11$, so that it will possess a rational 5-isogeny.

To save time, we can use the fact that $\{0, a/p\}^* = \{0, -a/p\}$; thus for odd p we need only evaluate half the sum, say

$$\mu_p' = \sum_{a=1}^{(p-1)/2} \left\{0, \frac{a}{p}\right\},$$

and then set $\mu_p = \mu_p' + (\mu_p')^*$.

For $p = 3$, we have $\mu_3' = \{0, 1/3\} = (3:1) = (3) = B$, so $\mu_3 = B + B^* = A$, giving $n_3 = 1$ and $a_3 = 1 + 3 - 5n_3 = -1$.

For $p = 5$ we compute:

$$\begin{aligned}
\left\{0, \frac{1}{5}\right\} &= (5:1) = (5) = -A; \\
\left\{0, \frac{2}{5}\right\} &= \left\{0, \frac{1}{2}\right\} + \left\{\frac{1}{2}, \frac{2}{5}\right\} \\
&= (2:1) + (-5:2) = (2) + (3) = A + B; \\
\mu_5' &= (-A) + (A + B) = B; \\
\mu_5 &= B + B^* = A, \qquad \text{so that } n_5 = 1; \\
a_5 &= 1 + 5 - 5n_5 = 1.
\end{aligned}$$

Similarly, with $p = 7$ we have $n_7 = 2$, so that $a_7 = 1 + 7 - 5n_7 = -2$, and with $p = 13$ we have $n_{13} = 2$ so that $a_{13} = 4$.

For the 'bad prime' $q = 11$ we compute the involution W_{11} induced by the action of the matrix $\begin{pmatrix} 0 & -1 \\ 11 & 0 \end{pmatrix}$:

$$\begin{aligned} W_{11}(A) = \begin{pmatrix} 0 & -1 \\ 11 & 0 \end{pmatrix} \left\{0, \frac{1}{2}\right\} &= \left\{\infty, \frac{-2}{11}\right\} \\ &= \{\infty, 0\} + \left\{0, \frac{-1}{5}\right\} + \left\{\frac{-1}{5}, \frac{-2}{11}\right\} \\ &= (1:0) + (-5:1) + (11:5) \\ &= (\infty) + (-5) + (0) \\ &= -A, \end{aligned}$$

so that the eigenvalue ε_{11} of W_{11} is -1. In fact, this was implicit earlier, since $L(f, 1) \neq 0$ implies that the sign of the functional equation is $+1$, which is minus the eigenvalue of the Fricke involution W_{11}.

The Fourier coefficients $a(n) = a(n, f)$ for $1 \leq n \leq 16$ are now

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$a(n)$	1	-2	-1	2	1	2	-2	0	-2	-2	1	-2	4	4	-1	-4

where we have used multiplicativity, and:

$$\begin{aligned} a(11) &= -\varepsilon_{11} = +1; \\ a(4) &= a(2)^2 - 2a(1) = 2; \\ a(8) &= a(2)a(4) - 2a(2) = 0; \\ a(16) &= a(2)a(8) - 2a(4) = -4; \\ a(9) &= a(3)^2 - 3a(1) = -2. \end{aligned}$$

We know that the period lattice Λ_f has a $\mathbb{Z}$-basis of the form $[\omega_1, \omega_2] = [2x, x+iy]$, where $\omega_1 = \langle A, f \rangle$ and $\omega_2 = \langle B, f \rangle$. We can compute the real period $\omega_1 = \Omega(f) = 5L(f, 1)$ by computing $L(f, 1)$:

$$L(f, 1) = 2 \sum_{n=1}^{\infty} \frac{a(n)}{n} t^n$$

where $t = \exp(-2\pi/\sqrt{11}) = 0.15\ldots$. Using the first 16 terms which we have already gives this to 13 decimal places:

$$L(f, 1) = 0.2538418608559\ldots;$$

thus

$$\omega_1 = \Omega(f) = 1.269209304279\ldots.$$

For the imaginary period y we twist with a prime $l \equiv 3 \pmod 4$. Here $l = 3$ will do, since

$$\gamma_3 = \left\{0, \frac{1}{3}\right\} - \left\{0, \frac{-1}{3}\right\} = (3) - (-3) = 2B - A \neq 0.$$

To project onto the minus eigenspace we take the dot product of this cycle (expressed as a vector $(-1,2)$) with $v^- = (0,1)$ to get $m^-(3) = 2$. Hence $yi = P(3,f)/2$, or $y = \sqrt{3}L(f \otimes 3, 1)/2$. Summing the series for $L(f \otimes 3, 1)$ to 16 terms gives only 4 decimals:

$$L(f \otimes 3, 1) = 1.6845\ldots.$$

This is less accurate than $L(f,1)$, since this is a power series in $\exp(-2\pi/3\sqrt{11}) = 0.53\ldots$, compared with $0.15\ldots$. Hence $y = 1.4588\ldots$, so that

$$\omega_2 = 0.634604652139\ldots + 1.4588\ldots i.$$

So far we have only used the Hecke eigenvalues a_p for $p \le 13$, and only 16 terms of each series. If we use these approximate values for the period lattice generators ω_1 and ω_2 we already find the approximate values $c_4 = 495.99$ and $c_6 = 20008.09$ which round to the integer values $c_4 = 496$ and $c_6 = 20008$. Taking the first 25 a_p and the first 100 terms of the series gives

$$c_4 = 495.9999999999954\ldots \qquad \text{and} \qquad c_6 = 20008.0000000085.$$

The exact values $c_4 = 496$ and $c_6 = 20008$ are the invariants of an elliptic curve of conductor 11, which is therefore the modular curve E_f:

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

This is the first curve in our tables, with code 11A1 (or Antwerp code 11B). The value $L(f,1)/\Omega(f) = 1/5$ agrees with the value predicted by the Birch–Swinnerton-Dyer conjecture for $L(E_f,1)/\Omega(E_f)$, provided that E_f has trivial Tate–Shafarevich group Ш. A 2-descent shows that the 2-primary component of Ш is indeed trivial.

Lastly, we illustrate the method explained at the end of the previous section, where we "guess" the imaginary period and lattice type. Having computed $P(3,f) = 2.9176\ldots i \ne 0$, we consider the lattices $\langle x, yi\rangle$ and $\langle 2x, x+yi\rangle$, where $x = 1.2692\ldots$ (from above) and $yi = P(3,f)/m^-$, for $m^- = 1, 2, 3, \ldots$. When $m^- = 1$ we do not find integral invariants, but for $m^- = 2$ and lattice type 1 we find the curve $E_f = [0,-1,1,-10,-20]$ given above.

Using the first variant of the method, where we do not even know x, we can take $l^+ = 5$ since $P(5,f) = 6.346\ldots \ne 0$. The correct value of m^+ here is 10; if we do not know this, but try $m^+ = 1, 2, 3\ldots$ in a double loop with m^-, the first valid lattice we come across is with $(m^+, m^-) = (2,2)$ and type 1, which leads to the curve $E' = [0,-1,1,0,0]$ of conductor 11; this is 5-isogenous to the "correct" curve E_f, which comes from $(m^+, m^-) = (10,2)$ and type 1.

We may also consider the ratios $P(l,f)/P(3,f)$ for other primes $l \equiv 3 \pmod 4$; we restrict to those l satisfying $\left(\frac{-11}{l}\right) = \left(\frac{l}{11}\right) = +1$, since otherise $P(l,f)$ is trivially 0 (since the sign of the functional equation for the corresponding $L(f \otimes \chi, s)$ is then -1). We find the following table of values (rounded: they are only computed approximately):

l	3	23	31	47	59	67	71	103	163	179	191	199	223	251
$\dfrac{P(l,f)}{P(3,f)}$	1	1	1	0	1	9	1	0	4	25	1	4	1	1

The zero values for $l = 47$ and $l = 103$ indicate that the corresponding twists of the newform f have positive even analytic rank (and we have checked that the corresponding twists of the curve E_f do have rank 2). As all these values are integral here (*a priori* they are only known to be rational) we do not find any nontrivial divisor of m^- (which we know in fact equals 2). The fact that all the integers are perfect squares is an amusing observation, but has a simple explanation in terms of the numbers appearing in the Birch–Swinnerton-Dyer conjecture for the twists of E_f.

Finally, there is one other curve E'' isogenous to E_f in addition to E' (found above). If the period lattice of $E_f = [0, -1, 1, -10, -20]$ is $[2x, yi]$ with $x = 0.6346...$ and $y = 1.4588...$, then $E' = [0, -1, 1, 0, 0]$ has period lattice $[10x, 5x + yi]$, and $E'' = [0, -1, 1, -7820, 263580]$ has lattice $[x/5, 2x/5 + yi]$. These curves are linked by 5-isogenies $E \leftrightarrow E'$ and $E \leftrightarrow E''$.

9. Results for $N \le 5077$

We have carried out the computations described above for all levels N up to 5077. (The reason for stopping at 5077 rather than, say, 5000, was to verify that the known curve of rank 3 and conductor 5077 was the first case of a modular elliptic curve of rank 3.)

For each N we found the rational newforms, and computed many Hecke eigenvalues for each; in the worst case we needed all a_p for $p < 30000$. For each form we computed a period lattice, and hence found a corresponding curve of conductor N. At present (February 1995), we have found the full period lattice and hence E_f only for $N \le 3200$; for $3200 < N < 5077$ we have only used the method of section 7 to find a suitable curve, which we only know to be isogenous to E_f. For each curve, we verified (in most cases, including all cases of rank 2) that the rank was equal to the analytic rank, and by finding the Mordell-Weil group of each curve (again, in most cases) we were able to compute the value predicted by the Birch Swinnerton-Dyer conjecture for the order of the Tate–Shafarevich group.

We summarise the results obtained in the following table, where for brevity we only give the numbers of newforms found, subdivided by rank. The first examples of each rank are: $N = 11$ for rank 0; $N = 37$ for rank 1; $N = 389$ for rank 2; and $N = 5077$ for rank 3.

Rational newforms for $\Gamma_0(N)$

$N \le 5077$

Range of N	Total	$r = 0$	$r = 1$	$r = 2$	$r = 3$
1–1000	2463	1321	1124	18	0
1001–2000	3391	1575	1737	79	0
2001–3000	3661	1663	1852	146	0
3001–4000	3836	1664	2006	166	0
4001–5000	3948	1684	2086	178	0
5001–5077	284	121	148	14	1
1–5077	17583	8028	8953	601	1

10. Further applications of the method

Using the methods desribed in these notes, several other applications are possible, which we mention here briefly:

1. Associated with each rational newform f for $\Gamma_0(N)$, we have a map $\varphi\colon X_0(N) \to E_f$, called the *modular parametrization* of the curve E_f. We can compute this map (or more precisely its lift $\mathcal{H} \to \mathbb{C}$) very explicitly numerically: if $\tau = x + yi \in \mathcal{H}$, then

$$\begin{aligned}\varphi(\tau) &= 2\pi i \int_{\infty}^{\tau} f(z)dz \quad (\tau \in \mathcal{H}) \\ &= \sum_{n=1}^{\infty} \frac{a(n,f)}{n} \exp(-2\pi n y)\exp(2\pi i n x).\end{aligned}$$

This gives a complex number z modulo Λ_f, and hence by composing with the Weierstrass functions we may obtain from each $\tau \in \mathcal{H}$ an explicit point $P(\tau) \in E_f(\mathbb{C})$. For suitable choices of τ this point $P(\tau)$ may be interesting, a Heegner point for example. Using this formula we can thus compute Heegner points on the modular elliptic curves E_f (of rank 1). We have carried this out in a few cases, but not yet systematically.

2. Given an explicit representation for the full homology $H(N)$, it is possible to compute, via an exact algebraic formula, the degree of the modular parametrization $\varphi\colon X_0(N) \to E_f$. This is described in detail in [8].

3. Variation of the methods described here can be used to find cusp forms of weight 2 for other congruence subgroups, such as $\Gamma_1(N)$. The case of $\Gamma_1(N)$ is described in more detail in [7].

4. In another direction, modular symbol methods may be used over other fields. The case of imaginary quadratic fields, which involves 3-dimensional hyperbolic geometry instead of the 2-dimensional upper half-plane $\mathcal{H}$, is the subject of active research by the author and his students past and present.

References

1. A. O. L. Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.
2. J. P. Buhler, B. H. Gross and D. B. Zagier, *On the conjecture of Birch and Swinnerton Dyer for an elliptic curve of rank 3*, Math. Comp. **44** (1985), 473–481.
3. H. Carayol, *Sur les représentations l-adiques attachées aux formes modulaires de Hilbert*, Comptes-Rendus de l'Acad. des Sci. **296,I** (1983), 629.
4. H. Cohen, *A course in Computational Number Theory*, Graduate Texts in Mathematics **138**, Springer, 1993.
5. I. Connell, *Notes on elliptic curves*, (Unpublished lecture notes, McGill University), 1991.
6. J. E. Cremona, *Algorithms for Modular Elliptic Curves*, Cambridge University Press, 1992.
7. J. E. Cremona, *Modular symbols for* $\Gamma_1(N)$ *and elliptic curves with everywhere good reduction*, Math. Proc. Camb. Phil. Soc. **111** (1992), 199–218.
8. J. E. Cremona, *Computing the degree of tha modular parametrization of a modular elliptic curve*, Math. Comp. (1995) (to appear).
9. B. Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic Algebraic Geometry, Progress in Mathematics **89**, Birkhauser, 1991, pp. 25–39.

10. A. W. Knapp, *Elliptic Curves*, Mathematical Notes **40**, Princeton University Press, 1992.
11. V. I. Kolyvagin, *Finiteness of* $E(\mathbb{Q})$ *and* $Ш_{E/\mathbb{Q}}$ *for a subclass of Weil curves*, Math. USSR Izvest. **32** (1989), 523–542.
12. A. Kraus, *Quelques remarques à propos des invariants* c_4, c_6 *et* Δ *d'une courbe elliptique*, Acta Arith. **54** (1989), 75–80.
13. S. Lang, *Elliptic functions*, Addison-Wesley, 1973.
14. S. Lang, *Introduction to modular forms*, Springer-Verlag, 1976.
15. Ju. I. Manin, *Parabolic points and zeta-functions of modular curves*, Math. USSR-Izv. **6** (1972), 19–64.
16. B. Mazur, *Courbes elliptiques et symboles modulaires*, Séminaire Bourbaki No. 414 (1972), 277–294.
17. L. Merel, *Opérateurs de Hecke pour* $\Gamma_0(N)$ *et fractions continues*, Annales de l'Institut Fourier de l'Université de Grenoble **41 no.3** (1991), 519–537.
18. M. R. Murty and V. K. Murty, *Mean values of derivatives of modular L-series*, Ann. Math. **133** (1991), 447–475.
19. G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Math. Soc. Japan No.11, 1971.
20. H. P. F. Swinnerton Dyer and B. J. Birch, *Elliptic curves and modular functions*, Modular Functions of One Variable IV, Lecture Notes in Mathematics **476**, Springer-Verlag, 1975.
21. J. Tate, *Algorithm for determining the singular fiber in an elliptic pencil*, Modular Functions of One Variable IV, Lecture Notes in Mathematics **476**, Springer-Verlag, 1975.
22. D. J. Tingley, *Elliptic curves uniformised by modular functions*, Oxford D.Phil. thesis 1975.
23. D. Zagier, *Modular Parametrizations of Elliptic Curves*, Canadian. Math. Bull. **28** (1985), 372–384.