Numerical modular symbols for elliptic curves

Christian Wuthrich

20th March 2017

Abstract

We present a detailed analysis of how to implement the computation of modular symbols for a given elliptic curve by using numerical approximations. This method turns out to be more efficient than current implementations as the conductor of the curve increases.

1 Introduction

The aim of the article is to describe an alternative algorithm for computing modular symbols for a given fixed elliptic curve $E/\mathbb{Q}$ of conductor N. The current implementations use linear algebra with rational coefficients to determine the space of modular symbols attached to E within the space of all symbols of level N. Instead we wish to compute efficiently the value of a modular symbols for a fixed E avoiding to work with the full space whose dimension grows linearly in N. We build on the work of Goldfeld [12] using numerical approximation to path integrals in the upper half plane. He already noted that the Atkin-Lehner involutions can be used to avoid integrating close to the real line where the convergence is bad. Goldfeld obtained that a single modular symbol for a semistable elliptic curve can be computed roughly in $N^{1/2}$ steps. See Theorem 10 where we recall the precise statement.

We improve on his work in several directions. First we prove all the rigorous bounds and we present the finer details of an implementation that returns provably correct rational numbers. This uses some theoretical knowledge about the possible denominators. Moreover, we explain what methods can be used for elliptic curves that are not semistable. Furthermore, we explain an idea that allows us to compute the modular symbols at all rational numbers with a fixed denominator. This is very useful for the practical applications we have in mind, for instance computing the p-adic L-functions of E. Finally, we analyse where possible the complexity of our algorithms. For instance, we prove that the set of all Manin symbols can be evaluated in roughly $N^{7/4}$ steps. See Theorem 13 for the precise statement.

We implemented the algorithms in `SageMath` [9]. It turns out to be faster when computing a single modular symbol and allows for computations with larger conductors than all previous implementations.

In order to describe the methods and results in more detail, we start by defining modular symbols (sometimes called modular elements). Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. Let $f = \sum a_n q^n$ be the newform of weight 2 and level $\Gamma_0(N)$ associated to the isogeny class of E. We know that f exists by modularity [5]. Given a rational number $r \in \mathbb{Q}$, we consider the integrals

$$\lambda(r) = 2\pi i \int_{i\infty}^{r} f(z)dz = 2\pi \int_0^\infty f(r+yi)dy \ \in \ \mathbb{C}. \tag{1}$$

Let γ^+ be a $\mathbb{Z}$-basis of the subgroup of $H_1\big(E(\mathbb{C}),\mathbb{Z}\big)$ fixed by complex conjugation and, similarly, let γ^- be a generator for the subgroup on which complex conjugation acts by multiplication with -1. Let ω_E be a Néron differential on E. Let Ω^+ be the smallest positive period of E, i.e., $\Omega^+ = \big|\int_{\gamma^+}\omega_E\big|$. Similarly, we set Ω^- to be $\big|\int_{\gamma^-}\omega_E\big|$ in $\mathbb{R}_{>0}$. The period lattice Λ_E of E is either $\mathbb{Z}\Omega^+ \oplus \mathbb{Z}\Omega^- i$ or $\mathbb{Z}\Omega^+ \oplus \mathbb{Z}\frac{1}{2}(\Omega^+ + \Omega^- i)$ depending on whether the discriminant of E is negative or positive.

Manin [16] and Drinfeld [10] showed that there exists an integer t such that $t\cdot\lambda(r) \in \mathbb{Z}\Omega^+ \oplus \mathbb{Z}\Omega^- i$ for all $r \in \mathbb{Q}$. In Section 2, we will look for a good bound on t in practice. Now we define the rational numbers

$$[r]^+ = \frac{\mathrm{Re}\big(\lambda(r)\big)}{\Omega^+} \in \mathbb{Q} \qquad \text{and} \qquad [r]^- = \frac{\mathrm{Im}\big(\lambda(r)\big)}{\Omega^-} \in \mathbb{Q}.$$

In this article, the map $r \mapsto [r]^{\pm}$ will be called a modular symbol rather than the homological version where the paths are called modular symbols. Our main goal is to find a fast algorithm for computing the values $[r]^{\pm}$ for a given curve E and $r \in \mathbb{Q}$.

Current implementations of modular symbols of E compute the values $[r]^{\pm}$ as follows. First they determine the vector space over $\mathbb{Q}$ of all modular symbols as the modular form varies through all rational cuspidal forms of weight 2 and level N. Then the matrices for the first few Hecke operators are computed and they are used, together with the known eigenvalues a_p for our curve E, to find the subspace corresponding to our fixed cuspform f. (Or rather quotient as they work with the dual space.) Once this initial step of finding a basis for this subspace is done, the value of $[r]^{\pm}$ for a given r is computed efficiently using the continued fractions expansion of the rational number r.

A thorough explanation of this method is given in Stein's book [21] and in Cremona's book [7]. It is implemented in Cremona's library `eclib` [8], `Magma` [4], `PARI/GP` [19] and [26], and `SageMath` [9]. Originally these implementations were written to find the elliptic curves of a given conductor as explained in [7]. In particular, the modularity of E was proven with this method, too.

Instead, we use here that the modularity of the elliptic curve is known. We wish to avoid to work with the space of all modular symbols of level N because this involves manipulations with sparse matrices of size $N/3 \times N/4$ as explained in § 8.9 of [21]. As N increases the initial step takes up a very long time and it currently makes it difficult to work with elliptic curves of conductor larger than 10^5.

The approach in this paper is to compute the values of $\lambda(r) \in \mathbb{C}$ by finding a numerical approximation to the integral in (1). We assume that we are given the values of the Fourier coefficients a_n of f; for instance `PARI` [19] yields these very fast by point counting on the reductions of E. We also know how to compute good approximations to the values of the periods $\Omega^{\pm}$. We make one assumption: We suppose that the Manin constant of the strong Weil curve in the isogeny class of E is 1. See Section 2.2 for the concrete implication of this assumption.

Here is how the modular symbol $[r]^{+}$ is computed in practice. First we use Manin's trick [16] with continued fractions to split the path from $i\infty$ to r into pieces (Section 6). This reduces our problem to evaluating so called Manin symbols (Section 6.2). These are integrals between two cusps r and r'. The main advantage is that the denominators of r and r' are now small compared to N. The path from r to r' is split up at the best place into two pieces (Section 4). We use an Atkin-Lehner involution as in [12] to move the path close to r to a path close to $i\infty$ where the Fourier expansion of f allows for fast integration. This integration is done by a summation where the number of terms and the precision of the floating point numbers is determined rigorously to guarantee the result within a given error (Section 3).

However, this is not possible for all cusps r. A cusp is called "unitary" if it is in the orbit of $i\infty$ under the group of Atkin-Lehner involutions. If we encounter a non-unitary cusp, we have to fall back to a much slower method using so-called transportable paths (Section 5), which we would like to avoid, if at all possible. The most important idea for this is to replace the curve by its quadratic twist of minimal conductor (Section 7.2). Furthermore there is also some flexibility in the continued fraction method.

The main application we have in mind is to compute algebraic L-values $L(E, \chi, 1)$ for Dirichlet characters χ and to compute p-adic L-functions. In both cases one only needs to find all values $[\frac{a}{m}]$ for a fixed m. Typically they are all unitary symbols. In Section 7.1, we explain an idea using partial sums that allows us to evaluate all of these symbols almost as fast as a single evaluation. This has also theoretical implications for the complexity estimates proven in Theorem 13.

The structure of the paper goes through the above explanation of the computation in reversed order. It is important first to understand the bounds for the possible denominators of $[r]^{\pm}$ in Section 2. Then we deal with the numerical approximation in Section 3 followed by how to split up and move the integration paths in Section 4 and 5. How to use and compute Manin symbols is explained in Section 6. Then, Section 7 describes how to take advantage of quadratic twists and partial sums and Section 8 looks at the complexity of all steps for unitary symbols.

We end the paper with examples and numerical comparisons with current implementations. We will illustrate that our method proves to be much faster when we need to evaluate a single, or a small number of values of $[r]^{\pm}$. It is even comparable when the task is to evaluate all Manin symbols as long as we assume that the curve is semistable. When N is really large, say 10^{10}, our method still determines single values of modular symbols quite fast, while the current implementations cannot perform the initial step any more. We refer to Section 9 for precise timings.

The methods in this paper could be extended to modular forms that do not come from elliptic curves; for instance forms associated to $\mathbb{Q}$-curves. We have not explored this or any potential gener-

alisations to other groups or situations.

Acknowledgements

It is a pleasure to thank John Cremona, Christophe Delaunay, Marc Masdeu, Dave Parkin and Fredrik Strömberg for help with the research and the implementation.

2 Denominator of modular symbols

We will compute a numerical approximation to the rational numbers $[r]^{\pm}$ defined in the introduction. In order to know to what precision we need to compute the approximation, we have to find a good bound on the denominator of the rational numbers $[r]^{+}$ and $[r]^{-}$. This will also lead us to the issue concerning the Manin constant. See [27] for further investigations on these denominators.

First, we need a few further definitions. Throughout this text E will be an elliptic curve defined over $\mathbb{Q}$ of conductor N. We know that E is modular and so let

$$\varphi\colon X_0(N) \longrightarrow\!\!\!\!\!\rightarrow E$$

be a modular parametrisation of minimal degree sending $i\infty$ to O. This is defined up to an automorphism of E defined over $\mathbb{Q}$, so up to multiplication by $[-1]$. The Manin constant is defined to be the rational number $c_E > 0$ such that

$$\varphi^*(\omega_E) = \pm c_E \cdot 2\pi i\, f(z) dz.$$

We choose φ uniquely such that it is the $+$ sign that appears in the above equality.

In the isogeny class of E there is a unique strong Weil curve E_0 (also called the X_0-optimal curve in [25]).

Assumption. The Manin constant $c_0 = c_{E_0}$ of the strong Weil curve E_0 is 1.

It is known that c_0 is an integer [11], and it is believed to be equal to 1 in all cases. See [1] for a discussion of known results about c_0. In particular, it is known that c_0 is either 1 or 2 when E is semistable. See Section 2.2 for an explanation of how harmful the above assumption is. A consequence of this assumption is that the Néron lattice Λ_0 of E_0 is equal to the lattice generated by all values $\lambda(r) - \lambda(s)$ as r and s run through all pairs of $\Gamma_0(N)$-equivalent cusps.

When comparing the period lattices of E and E_0, the quotient of the Néron periods will become important. Define the rational numbers

$$q^+ = \frac{\Omega_E^+}{\Omega_{E_0}^+} \qquad \text{and} \qquad q^- = \frac{\Omega_E^-}{\Omega_{E_0}^-}.$$

Let r be a rational number.

Definition. We write $r = \frac{a}{m}$ as a reduced fraction of integers. Let M be the greatest common divisor of m and the conductor N. Hence we can write $N = Q \cdot M$ and $m = d \cdot M$. Following Section 3.1 in [17], we call a cusp r *unitary* if Q and M are coprime. The integer $Q/\gcd(M, Q)$ is called the *width* of the cusp r; for unitary cusps it is simply Q.

It is known that r is unitary if and only if the cusps r and $i\infty$ on $X_0(N)$ are in the same orbit under the action by the group of Atkin-Lehner involutions. In the application we have in mind, no prime of additive reduction for E divides the denominator m. Then the cusp $\frac{a}{m}$ is unitary. For a semistable curve all cusps are unitary.

Further, we set δ^2 to be the largest square dividing N. Thus $\delta = 1$ if and only if E is semistable.

Proposition 1. *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N. Choose a few primes $\ell > 2$ coprime to N, with $\ell \equiv 1 \pmod{\delta}$ and set t_0 to be the greatest common divisor of the number N_ℓ of points on the reduction of E modulo ℓ. Let $t^{\pm}$ be the numerator of $t_0 q^{\pm}$. Assume $c_0 = 1$. We have*

$$[r]^{\pm} \in \frac{c_\infty(E_0)}{2t^{\pm}}\,\mathbb{Z},$$

where $c_\infty(E_0)$ is the number of connected components of $E_0(\mathbb{R})$. If r is unitary, we also get

$$[r]^{+} \in \frac{c_\infty(E)}{2 \cdot \#E(\mathbb{Q})_{\mathrm{tors}}} \qquad \text{and} \qquad [r]^{-} \in \frac{1}{2}\mathbb{Z}.$$

where $c_\infty(E)$ is the number of connected components of $E(\mathbb{R})$.

For a semistable curve, even without assuming $c_0 = 1$, we get the bounds $4 \cdot \#E(\mathbb{Q})_{\mathrm{tors}}/c_\infty(E) \leqslant 24$ and 4 for the denominator of $[r]^+$ and $[r]^-$ respectively. If t is a bound for the denominator of a modular symbols $[r]^+$ as above then, in our implementation we now round $t \cdot \mathrm{Re}\big(\lambda(r)\big)/\Omega^+$ to the closest integer and find $[r]^+$ by dividing again by t. Hence we must compute $\mathrm{Re}\big(\lambda(r)\big)$ with a proven error smaller than $\Omega^+/(2t)$.

Proof. Consider the modular parametrisation $\varphi_0 \colon X_0(N) \to E_0$. After identifying $E_0(\mathbb{C})$ with $\mathbb{C}/\Lambda_{E_0}$ via the integration of ω_{E_0}, we get an induced map $\tilde{\varphi}_0$ from the upper half plane to $\mathbb{C}/\Lambda_{E_0}$. We find

$$\tilde{\varphi}_0(r) \equiv \int_O^{\varphi_0(\Gamma_0(N)r)} \omega_{E_0} \equiv \int_{\Gamma_0(N)i\infty}^{\Gamma_0(N)r} \varphi_0^*(\omega_{E_0}) \equiv c_0 \int_{i\infty}^{r} 2\pi i f(z) dz \equiv c_0 \lambda(r) \pmod{\Lambda_{E_0}}$$

By the theorem of Manin and Drinfeld, the modular parametrisation $\varphi \colon X_0(N) \to E$ maps the cusp r to the torsion point $\varphi(r) \in E(\bar{\mathbb{Q}})$. The action of the Galois group on the cusps on $X_0(N)$ is given in Theorem 1.3.1 in [24]. The cusps on $X_0(N)$, and hence all points $\varphi(r)$ for $r \in \mathbb{Q}$, are defined over the cyclotomic field $K = \mathbb{Q}(\zeta_\delta)$. The image of the unitary cusps is in the torsion subgroup of $E(\mathbb{Q})$ instead.

If $\ell \equiv 1 \pmod{\delta}$, then there is a place v in K above ℓ with residue field $\mathbb{F}_\ell$. If $\ell \nmid N$, then we get a reduction map $E_0(K) \to \tilde{E}_0(\mathbb{F}_\ell)$ of elliptic curves. Since $v \mid \ell$ is unramified and $\ell > 2$, we conclude from Theorem VII.3.4 in [20] that the reduction map is injective on torsion points in $E(K)$. Hence N_ℓ is a multiple of the order of the torsion subgroup of $E(K)$ for all these ℓ. We conclude that $t_0\, \varphi_0(r) = O$ in $E_0(K)$. Therefore $t_0\, \tilde{\varphi}_0(r)$ and hence $t_0\, c_0\, \lambda(r)$ belong to Λ_{E_0}. Recall that $\lambda(r) = [r]^+ q^+ \, \Omega^+_{E_0} + [r]^- q^- \, \Omega^-_{E_0}\, i$. Now if Λ_{E_0} is rectangular, then $c_\infty(E_0) = 2$ and $c_0 t_0 q^\pm \, [r]^\pm \in \mathbb{Z}$. If Λ_{E_0} is not rectangular, then $c_\infty(E_0) = 1$ and $c_0 t_0 q^\pm \, [r]^\pm \in \frac{1}{2}\mathbb{Z}$. Thus combined we find that $c_0 t_0 q^\pm \, [r]^\pm$ belong to $c_\infty(E_0)/2\,\mathbb{Z}$.

Finally if r is unitary, then $\varphi(r)$ belongs to $E(\mathbb{Q})$ and hence to $E(\mathbb{R})$. This implies in both the rectangular and the non-rectangular case that $[r]^- \in \frac{1}{2c_0}\mathbb{Z}$. We find that $2 \cdot \#E(\mathbb{Q})_{\mathrm{tors}}\, [r]^+/(c_0\, c_\infty(E))$ belongs to $\mathbb{Z}$. □

By the way, the original proof of Manin [16] and Drinfeld [10] used the Hecke operators and found that N_ℓ for $\ell \equiv 1 \pmod{N}$ is a bound for the order of $\varphi_0(r) \in E(K)$; our bound involving δ rather than N is better.

We add the example of the strong Weil curve 121d1. Here $E(\mathbb{Q})$ is trivial and $E(\mathbb{R})$ is connected. So we expect a denominator 1 or 2 for all unitary cusps. We have $\delta = 11$ for this curve and $N_{23} = 25$ which is also the greatest common divisor of the first few $\ell \equiv 1 \pmod{11}$. In fact the torsion subgroup of $E(\mathbb{Q}(\mu_{11}))$ is isomorphic to $\mathbb{Z}/25\mathbb{Z}$. Hence we can bound the denominator of $[r]^\pm$ by 50. One can show that $[r]^+ \in \frac{1}{2}\mathbb{Z}$ and $[r]^- \in \frac{1}{50}\mathbb{Z}$. For instance, $\lambda\big(\frac{3}{11}\big) = -\frac{1}{2}\Omega^+ + \frac{27}{50}\Omega^- \, i$.

2.1 Implementation of the Manin constant

We add here an explanation of how to implement the Manin constant under the above assumption (as it is now done in SageMath).

Proposition 2. *Let E be an elliptic curve defined over $\mathbb{Q}$. Let $n^\pm$ be the numerator of $q^\pm$ as defined above. Then the Manin constant c_E is equal*

$$c_E = \begin{cases} c_0 \cdot \frac{1}{2} \cdot n^+ \cdot n^- & \text{if } n^+ \text{ and } n^- \text{ are both even and} \\ & E_0(\mathbb{R}) \text{ has more components than } E(\mathbb{R}), \\ c_0 \cdot 2 \cdot n^+ \cdot n^- & \text{if } E(\mathbb{R}) \text{ has more components than } E_0(\mathbb{R}) \text{ and} \\ & \text{at least one of the denominators of } q^+ \text{ and } q^- \text{ is odd}, \\ c_0 \cdot n^+ \cdot n^- & \text{otherwise.} \end{cases}$$

If $\alpha \colon E \to E'$ is an isogeny defined over $\mathbb{Q}$, then set $n^\pm_\alpha$ to be the numerator and $d^\pm_\alpha$ the denominator of $\Omega^\pm_{E'}/\Omega^\pm_E$. Further we define c_α by $\alpha^*(\omega_{E'}) = c_\alpha\, \omega_E$. We may choose the differentials such that c_α is a positive integer.

Let $\psi \colon E_0 \to E$ be the isogeny of smallest degree. By the definition of the strong Weil curve, the modular parametrisation of E factors through φ_0 and ψ. Thus the Manin constant c_E is equal to $c_0 \cdot c_\psi$. It is convenient to prove a lemma first.

Lemma 3. *Let $\alpha\colon E \to E'$ be a cyclic isogeny defined over $\mathbb{Q}$ of degree p^k for some prime p. If $p = 2$ assume that $E(\mathbb{R})$ and $E'(\mathbb{R})$ have the same number of connected components. Then $\gcd(n_\alpha^+, n_\alpha^-) = \gcd(d_\alpha^+, d_\alpha^-) = 1$ and $c_\alpha = n_\alpha^+ \cdot n_\alpha^-$.*

Proof. Integrating against the fixed Néron differentials ω and ω' on E and E' respectively, we identify $E(\mathbb{C})$ with $\mathbb{C}/\Lambda_E$ and $E'(\mathbb{C})$ with $\mathbb{C}/\Lambda_{E'}$. The isogeny $\mathbb{C}/\Lambda_E \to \mathbb{C}/\Lambda_{E'}$ is then induced by the multiplication by c_α on $\mathbb{C}$. Recall also that c_α divides p^k as $c_\alpha \cdot c_{\hat\alpha} = c_{[p^k]} = p^k$.

If n_α^+ and n_α^- have a common divisor n, then the isogeny α would factor through $[n]\colon \mathbb{C}/n\Lambda_{E'} \to \mathbb{C}/\Lambda_E$, but that is not possible as α is cyclic. Similarly d_α^+ and d_α^- are coprime.

Since $\ker(\alpha)$ is not a direct sum, it is either contained in $E(\mathbb{R})$ or in $E(\mathbb{C})^-$, the set of points Q in $E(\mathbb{C})$ whose complex conjugate is $\bar Q = -Q$. Let $z + \Lambda_E$ be a generator of $\ker(\alpha)$. Now if $z = x + iy$, then $\bar z \equiv \pm z \pmod{\Lambda_E}$ implies that either $2x \in \Lambda_E$ or $2iy \in \Lambda_E$.

Assume now that p is odd. Then the above implies that either $\ker(\alpha)$ is generated by $\Omega_E^+/p^k + \Lambda_E$ or it is generated by $i\Omega_E^-/p^k + \Lambda_E$. In the first case, we have $c_\alpha\Omega_E^+/p^k = \Omega_{E'}^+$ and $c_\alpha\Omega_E^- = \Omega_{E'}^-$. Together with $c_\alpha \mid p^k$, this implies that $n_\alpha^+ = d_\alpha^- = 1$ and $c_\alpha = n_\alpha^-$. The lemma is then proved in this case. The second case, when $\Omega_E^- i/p^k$ is in $\ker(\alpha)$, is similar but with signs swapped.

Finally, we assume $p = 2$. Consider $w = 2^{k-1}z$. Then $w + \Lambda$ is a 2-torsion point on E and, since α is defined over $\mathbb{Q}$, it lies in $E(\mathbb{R})[2]$. First, if $E(\mathbb{R})$ is connected, then $w \in \Omega_E^+/2 + \Lambda_E$ and $z \in \Omega_E^+/2^k + \Lambda_E$. We find ourselves in a case in which the explanation for general p treated above extends to $p = 2$. Also when $E(\mathbb{R})$ has two connected components, we fall back onto the two cases treated above, except when $w \in (\Omega_E^+ + i\Omega_E^-)/2 + \Lambda_E$. However in this last case, $E'(\mathbb{R})$ is connected, which is excluded by assumption. □

Proof of Proposition 2. We factor $\psi = \beta \circ \alpha$ with $\alpha\colon E_0 \to E'$ and $\beta\colon E' \to E$. We can impose that α is cyclic and $E_0(\mathbb{R})$ and $E'(\mathbb{R})$ have the same number of connected components and that β has degree 2 when $E_0(\mathbb{R})$ and $E(\mathbb{R})$ do not have the same number of connected components otherwise β is trivial.

Decomposing α into isogenies of prime power degrees, we can apply the previous lemma repeatedly. It follows that $\gcd(n_\alpha^+, n_\alpha^-) = \gcd(d_\alpha^+, d_\alpha^-) = 1$ and $c_\alpha = n_\alpha^+ \cdot n_\alpha^-$. This concludes the case when $E_0(\mathbb{R})$ and $E(\mathbb{R})$ have the same number of connected components.

Assume now that $E_0(\mathbb{R})$ has two and $E(\mathbb{R})$ has one connected component. As seen above in the case $p = 2$, it follows that the kernel of β is generated by $(\Omega_{E'}^+ + i\Omega_{E'}^-)/2 + \Lambda_{E'}$. Either $c_\beta = 1$ or 2. In the first case, we have $\Omega_{E'}^\pm = \Omega_E^\pm$ and hence $c_\psi = c_\alpha \cdot c_\beta = n_\alpha^+ \cdot n_\alpha^- = n_\psi^+ \cdot n_\psi^-$ proves this case. In the second case, we have $c_\beta = 2 = \Omega_E^\pm/\Omega_{E'}^\pm$. Thus $n_\psi^\pm/d_\psi^\pm = 2n_\alpha^\pm/d_\alpha^\pm$. We have to split up into two cases according to the parity of $d_\alpha^+ d_\alpha^-$. If it is even, then exactly one of d_α^+ and d_α^- is even and we find that $c_\psi = 2n_\alpha^+ n_\alpha^- = n_\psi^+ n_\psi^-$. Otherwise, if it is odd, $c_\psi = 2n_\alpha^+ n_\alpha^- = \frac{1}{2} n_\psi^+ n_\psi^-$. It remains to note that $d_\alpha^+ d_\alpha^-$ is odd if and only if n_ψ^+ and n_ψ^- are both even.

Finally, we can treat the case when $E_0(\mathbb{R})$ has one and $E(\mathbb{R})$ has two connected components by a similar case-by-case treatment. Alternatively one can just apply the above to the dual of β. □

There are other ways to find c_α for an isogeny $\alpha\colon E \to E'$. For instance, the expansion of α using the formal groups for E and E' will have c_α as the leading coefficient. Also there is the useful formula $c_\alpha^2 = \deg(\alpha)\, c_\infty(E)\, \Omega_E^+\, \Omega_E^- / (c_\infty(E')\, \Omega_{E'}^+\, \Omega_E^-)$. The advantage of the formula in Proposition 2 is that all terms can be read off E and E' without reference to ψ any more.

As an example we add here the case of the isogeny class 27a. There are four curves in this class and they are linked by the following 3-isogenies

where the direction of the arrow indicates the isogeny α for which $c_\alpha = 1$. In other words, the inclusion of the Néron lattices is in the opposite direction. The curve 27a1 is the strong Weil curve, while 27a3 is the minimal curve in the sense of [25]. The three curves on the right have each exactly 3 points in $E(\mathbb{Q})$ and they lie in the kernel of the isogeny to the curve on their left. The Manin constants are equal to 1 for 27a2 and 27a1 and they are equal to 3 for the two curves 27a3 and 27a4.

2.2 Outstanding issues

There are two outstanding issues. First, what happens if $c_0 \neq 1$ and secondly how do we find the strong Weil curve in the isogeny class.

Suppose that the Manin constant c_0 were larger than 1. If we knew the value of c_0 we could simply multiply the bounds t_0 and $t_0^{\pm}$ by c_0, too. However, it is then likely that we would at some point find a modular symbol where c_0 appears as a factor of the denominator. When rounding our numerical approximation, we would find a large error. If this happens, we could verify that $c_0 \neq 1$ and announce the exceptional news to the world. Therefore we do not really have to worry about this assumption in practice.

For all isogeny classes in Cremona's tables [7] it has been verified that $c_0 = 1$ when the table was created. For a few curves this is slightly more complicated and the issue is well explained in the appendix of [1].

The second issue is related to the first. Even for the curves in the tables, it is not always possible to say with certainty which curve in the isogeny class is the strong Weil curve. This arises because the computation in creating the table is done mostly with +-modular symbols only. At worst, we are off by a lattice of index 2.

Finally, suppose the curve lies outside the range of the table. We can still determine the isogeny class of the curve fairly quickly. However we have no means of knowing which curve is the strong Weil curve. To be on the safe side, we have to assume that it is one of the curves with maximal lattice. In practice it is very often on the contrary the minimal curve that is the strong Weil curve, but we have no way of showing this for our curve. If we are really unlucky, we even picked the wrong curve among the maximal curves; hence we should really work with the lattice generated by all Néron lattices in the isogeny class.

3 Numerical integration

Let f be the newform associated to the isogeny class of the elliptic curve E. Let $\varepsilon > 0$. In this section, we consider the finite sum that approximates the integral of $2\pi i f(z)dz$ from $i\infty$ to a point τ in the upper half plane. We prove bounds on the number of terms and the bit precision to work with in order to determine the integral with an error of at most ε.

Generalising the definition of $\lambda(r)$, we will consider

$$\lambda(\tau) = 2\pi i \int_{i\infty}^{\tau} f(z)dz$$

for any point $\tau = x + yi$ in the upper half plane. As $y > 0$, we can express it as the evaluation of a power series in $q = e^{2\pi i \tau}$, namely

$$\lambda(\tau) = \sum_{n=1}^{\infty} \frac{a_n}{n} q^n = \sum_{n=1}^{\infty} \frac{a_n}{n} \exp(-2\pi n y + 2\pi n x i). \tag{2}$$

We will approximate this sum by its finite partial sum for $n \leqslant T$ for a bound T. It is the value of y that determines how quickly the sum will converge and so how large T should be. In Section 7.1, we will be interested in the following partial sums: for any $m > 1$ and $0 \leqslant j < m$ and $y > 0$, we define

$$\kappa_{j,m}(y) = \sum_{\substack{n \geqslant 1 \\ n \equiv j \bmod m}} \frac{a_n}{n} \exp(-2\pi n y) \in \mathbb{R}. \tag{3}$$

3.1 Truncation

We now proceed to determine how many terms in the sums in (2) and (3) we have to add to be guaranteed a value that differs from the infinite sum by less than a given error ε. Afterwards we will decide with what level of precision we have to do the numerical computations so that the error due to precision loss will be smaller than a given bound ε'. Recall that we have determined the value of $\varepsilon + \varepsilon'$ in Section 2.

Define the following function for $y > 0$ and $\varepsilon > 0$.

$$T(y, \varepsilon) = \frac{-\log(2\pi y \varepsilon)}{2\pi y} \tag{4}$$

which is, for a fixed ε, a function that grows like a constant multiple of $\frac{1}{y}\log(\frac{1}{y})$ as $y \to 0$.

Lemma 4. *Let τ be an element of the upper half plane with $y = \mathrm{Im}(\tau)$ and let $\varepsilon > 0$. If $T > T(y, \varepsilon)$ then we have*

$$\left| \lambda(\tau) - \sum_{n=1}^{T} \frac{a_n}{n} \exp(2\pi i n \tau) \right| < \varepsilon \,.$$

Proof. Write $\tau = x + yi$. Now we use that $|a_n| \leqslant n$ as proven in Lemma 2.9 in [13]. The difference to bound is

$$\left| \sum_{n>T} \frac{a_n}{n} \exp\Big(2\pi i n(x + i\,y)\Big) \right| \leqslant \sum_{n>T} \frac{|a_n|}{n} \exp(-2\pi n y) \leqslant \sum_{n>T} \exp(-2\pi n y) = \frac{e^{-2\pi(T+1)y}}{1 - e^{-2\pi y}} = \frac{e^{-2\pi T y}}{e^{2\pi y} - 1}.$$

Now the condition on T implies that

$$\frac{e^{-2\pi T y}}{e^{2\pi y} - 1} < \frac{2\pi y \varepsilon}{e^{2\pi y} - 1} < \varepsilon. \qquad \square$$

In this proof, we have used the inequality $|a_n| \leqslant n$. In fact, we even know that $|a_n| \leqslant \sigma_0(n)\sqrt{n}$ where $\sigma_0(n)$ is the number of positive divisors of n. However even this asymptotically sharper inequality will not lead to a substantially better theoretical bound on the number of terms.

Nonetheless, in practice we use the following estimates. First we have the trivial bound $\sigma_0(n) \leqslant 2\sqrt{n}$. Moreover for every $2 > \varsigma > 0$ the equality $\sigma_0(n) < \varsigma \cdot \sqrt{n}$ holds for all $n > B(\varsigma)$ for some $B(\varsigma)$. Here are a few values of this bound used in the implementation:

ς	1	2/3	1/2	1/3	1/4	1/5	1/6
$B(\varsigma)$	1260	10080	55440	277200	831600	2162160	4324320

(5)

With the same method as in Lemma 4 one proves the bound on the approximation for the partial sum $\kappa_{j,m}(y)$. When we will compare the methods it will be clear that the corresponding error bound that we ask for is ε/m.

Lemma 5. *Let $y > 0$, $m > 1$, $0 \leqslant j < m$ and $\varepsilon > 0$. If $T > T(y, \varepsilon) + m$, then*

$$\left| \kappa_{j,m}(y) - \sum_{\substack{n \equiv j \bmod m \\ 1 \leqslant n \leqslant T}} \frac{a_n}{n} \exp(-2\pi n y) \right| < \frac{\varepsilon}{m}.$$

We have seen that the value of $\frac{1}{y}$ is an important measure of how difficult it will be to approximate the integral. This motivates the following definition.

Definition. We call the value of y the *speed* of the evaluation of $\lambda(x + yi)$.

The larger the speed the faster we can compute $\lambda(\tau)$.

Of course, since the sums are alternating in average (because the a_p for primes p follow the Sato-Tate distribution), they actually converge much faster. In [12], Goldfeld suggests that it is probable that the computation complexity is polynomial in N; in other words that the bound for T could behave like a power of $\log(\frac{1}{y})$. However this is still far beyond current knowledge. Even an unproven effective version of the Sato-Tate distribution does not seem to help here.

3.2 Implementation

For implementing these finite sums we use Horner's rule. Here is the algorithm to evaluate an approximation to $\lambda(\tau)$. We are given τ in the upper half plane and an bound ε on the allowed error.

Algorithm: Numerical approximation to $\lambda(\tau)$.

[Initialisation]: Set $s \leftarrow 0$ and $n \leftarrow \lceil T(y, \varepsilon) \rceil$ and compute $q \leftarrow \exp(2\pi i \tau)$.

[Loop]: While n is positive, replace $s \leftarrow s \cdot q + \frac{a_n}{n}$ and decrease n by one.

[End]: Return $s \cdot q$ as a good approximation to $\lambda(\tau)$.

The same idea can be used to compute an approximation to the partial sum $\kappa_{j,m}(y)$ for all j simultaneous. We are given m and y and the allowed error ε/m.

Algorithm: Simultaneous numerical approximation to $\kappa_{j,m}(y)$.

[Initialise]: Set $v_j \leftarrow 0$ for all $0 \leqslant j < m$. Compute $q \leftarrow \exp(-2\pi y)$ and $q' \leftarrow \exp(-2\pi m y)$. Set to start $n \leftarrow \lceil T(y,\varepsilon) \rceil$.

[Loop]: As long as n is positive, replace $v_j \leftarrow v_j \cdot q' + \frac{a_n}{n}$, where $j \equiv n \pmod m$ and then decrease n by 1.

[End]: At the end the value $v_j \cdot q^j$ for $1 \leqslant j < m$ and $v_0 \cdot q'$ are good approximations to $\kappa_{j,m}(y)$.

3.3 Precision

We wish to determine with how many bits b of precision we have to work with to make sure that the error in the above algorithm is smaller than a given error ε'. In practice this error will be chosen to be a tiny fraction of the error ε that we allowed for finding the above bound T.

Lemma 6. *Let $1 > \varepsilon' > 0$ and τ in the upper half plane. Let T be the number of terms evaluated in the sum to approximate $\lambda(\tau)$. If*

$$2^{-b} < \frac{\varepsilon'}{2T(T+\varepsilon')}$$

then the numerical value computed differs from the actual sum $\sum_{n=1}^{T} \frac{a_n}{n} q^n$ by less than ε' in absolute value.

Proof. We may suppose that the value of q can be pre-computed to b bits of precision. We use the absolute error estimate on the Horner's rule given on page 105 of [14]. If we write $\delta = 2^{-b}$, then the absolute error is smaller than

$$\frac{2T\delta}{1-2T\delta} \cdot \sum_{n=1}^{T} \left|\frac{a_n}{n}\right| \cdot |e^{2\pi i \tau}|^n \leqslant \frac{2T\delta}{1-2T\delta} \cdot \sum_{n=1}^{T} e^{-2\pi n y} \leqslant \frac{2T\delta}{1-2T\delta} \cdot T,$$

where we used again that $|a_n| \leqslant n$. It is now easy to see that the given inequality on δ in the lemma implies that the above right hand side is smaller than ε' □

For the approximation of $\kappa_{j,m}(y)$ to have an error smaller than ε'/m, we have to impose the bound

$$2^{-b} < \frac{\varepsilon' m}{2T'(T'+\varepsilon')}$$

where $T' = T(y,\varepsilon) + m$ is the upper limit of the finite sums in Lemma 5.

Later, it will be clear later that, in view of Lemma 6, we may neglect the issue of memory usage because the floating point numbers will take up approximatively as many bits as the conductor or the coefficients of E take up.

Within the range of interesting examples, the standard double precision of 53 bits is often sufficient. For example, the period Ω^+ of the curve 100002a1 is approximatively equal to 1.125. If we set $\varepsilon = 0.278427$ and $\varepsilon' = 0.002812$, then we are allowed to sum up $T = 3558923$ terms using 53 bits precision, which would allow for $\frac{1}{y}$ to be as large as 1627105. From the results in the following sections one can deduce that this allows to evaluate all Manin symbols using standard double precision.

Instead, for a curve like $E\colon y^2 = x^3 + 101x + 103$ of conductor 35261176, the evaluation of $\lambda(\frac{1}{107})$ will require precision above 53 bits to obtain provable results.

4 Computation of unitary symbols

In this section we assume that r is a unitary cusp. It is equivalent to the definition given at the start of Section 2 to ask that the cusp r on $X_0(N)$ is in the orbit of $i\infty$ under the group of Atkin-Lehner involutions. This section explains how to compute $\lambda(r)$ under this assumption. In Section 3, we explained how to compute integrals from a point τ within the upper half plane to the cusp $i\infty$. Now, we wish to explain how one integrates paths from τ to another cusp $r \in \mathbb{Q}$. The idea to use the Atkin-Lehner involution to bring r to ∞ is already presented in [12].

4.1 Moving unitary cusps with Atkin-Lehner involutions

By assumption, $r = \frac{a}{m}$ is unitary. Recall that we denote by M be the greatest common divisor of m and the conductor N. Further we write $N = Q \cdot M$ with Q and M coprime. Then the greatest common divisor of Qa and m is 1 and hence we find integers u and v such that $Qau + mv = 1$. We define

$$W_r = \begin{pmatrix} Qu & v \\ -Qm & Qa \end{pmatrix}$$

which is of determinant Q and sends r to $i\infty$ under the action of $\mathrm{GL}_2(\mathbb{Q})$ on the completed upper half plane. Since Qm is divisible by N, the matrix W_r induces an Atkin-Lehner involution on $X_0(N)$. Since f is a newform it is also an eigenfunction for W_r. We have $f|_{W_r} = \epsilon_Q \cdot f$ for $\epsilon_Q \in \{\pm 1\}$. In fact, ϵ_Q is easy to compute as it is just the product of the local root numbers for $\ell \mid Q$; and for a product of semistable primes, we have $\epsilon_Q = -a_Q$. We get

$$2\pi i \int_\tau^r f(z)dz = \epsilon_Q \cdot 2\pi i \int_\tau^r f|_{W_r}(z)dz = \epsilon_Q \cdot 2\pi i \int_{W_r(\tau)}^{i\infty} f(z)dz = -\epsilon_Q \cdot \lambda\big(W_r(\tau)\big) \tag{6}$$

which can be evaluated with the previously described numerical method. Note that the speed of this evaluation is equal to

$$\mathrm{Im}\big(W_r(\tau)\big) = \frac{Q \cdot \mathrm{Im}(\tau)}{|-Qm\tau + Qa|^2} = \frac{\mathrm{Im}(\tau)}{Q \cdot m^2 \cdot |r-\tau|^2}. \tag{7}$$

4.2 Splitting up the path from $i\infty$ to r

We wish to compute $\lambda(r)$ by splitting up the path of integration from r to $i\infty$ at a certain τ in the upper half plane. Using (6), we find, for any such τ and any unitary cusp r,

$$\lambda(r) = 2\pi i\Big(\int_{i\infty}^{\tau} + \int_\tau^r\Big) f(z)dz = \lambda(\tau) - \epsilon_Q \cdot \lambda\big(W_r(\tau)\big). \tag{8}$$

These two values of λ can be evaluated using the numerical method. We are now looking for the choice of τ such that the computation is fastest. The following lemma will show that this is achieved when the speed of computing $\lambda(\tau)$ is equal to the speed of computing $\lambda\big(W_r(\tau)\big)$ and they are maximal. From (7), the first condition is equivalent to the equation

$$\mathrm{Im}(\tau) = \frac{\mathrm{Im}(\tau)}{Q \cdot m^2 \cdot |r-\tau|^2}.$$

So we are looking for the $\tau = x + yi$ with maximal y such that $|r-\tau| = 1/\big(\sqrt{Q}\, m\big)$. This is obtained for

$$\tau = r + \frac{1}{m\sqrt{Q}}\, i.$$

It is not difficult to see that

$$W_r\Big(\frac{a}{m} + \frac{1}{m\sqrt{Q}}\, i\Big) = -\frac{u}{m} + \frac{1}{m\sqrt{Q}}\, i \tag{9}$$

where u is an inverse of $Q \cdot a$ modulo m. We still have to justify the claim that our choice of τ is optimal.

Lemma 7. *For a fixed curve E, a fixed unitary cusp r, and a fixed error $\varepsilon > 0$, the minimum of*

$$T\Big(y, \frac{\varepsilon}{2}\Big) + T\Big(\frac{y}{Qm^2|r-\tau|^2}, \frac{\varepsilon}{2}\Big)$$

is attained when $\tau = r + \frac{1}{m\sqrt{Q}}\, i$.

Proof. Write as before $\tau = x + y\, i$. Since $t(y) := T(y, \varepsilon/2)$ is decreasing in $y > 0$, the best choice for x must occur when $x = r$. The function to minimise simplifies then to $t(y) + t(1/(Cy))$ with $C = Q\, m^2$. Taking the derivatives with respect to y, we see that at the minimum, we must have

$$y \cdot t'(y) = \frac{1}{Cy} \cdot t'\Big(\frac{1}{Cy}\Big).$$

Now from the definition we see that t satisfies the differential equation

$$y \cdot t'(y) = -t(y) - \frac{1}{2\pi y}$$

and hence, since t is decreasing, $y \mapsto y \cdot t'(y)$ is increasing. Hence there is only one minimum, namely when $y = \frac{1}{Cy}$. □

4.3 Integrals from cusp to cusp

Let $r = \frac{a}{m}$ and $r' = \frac{a'}{m'}$ be two unitary cusps of widths Q and Q' respectively. Our aim is to compute

$$\lambda\big(\{r' \to r\}\big) = 2\pi i \int_{r'}^{r} f(z)dz = \lambda(r) - \lambda(r'),$$

where the integration follows any path from r' to r in the upper half plane. One way to do so, indicated by the last expression above, is to integrate from r' to $i\infty$ and then subtract the integration from r to $i\infty$ using the method explained above. We call this the *indirect* way.

Instead, the *direct* way splits up the integration path from r' to r into two pieces: First find a good τ in the upper half plane. Then use the Atkin-Lehner involution $W_{r'}$ to move the path from r' to τ to a path from $i\infty$ to $W_{r'}(\tau)$. Similarly use W_r to move the second piece to a path from $W_r(\tau)$ to $i\infty$. As before, on these two paths we can use the methods from the previous section. We get

$$\lambda\big(\{r' \to r\}\big) = 2\pi i \left(\int_{r'}^{\tau} + \int_{\tau}^{r} \right) f(z)dz = \epsilon_{Q'} \cdot \lambda\big(W_{r'}(\tau)\big) - \epsilon_Q \cdot \lambda\big(W_r(\tau)\big)\,.$$

We expect again the best choice for $\tau = x + i\,y$ to be such that the speeds of both integrals are equal and they are maximal. We get the equation

$$\frac{y}{m^2 Q \cdot |\tau - r|^2} = \frac{y}{m'^2 Q' \cdot |\tau - r'|^2}. \tag{10}$$

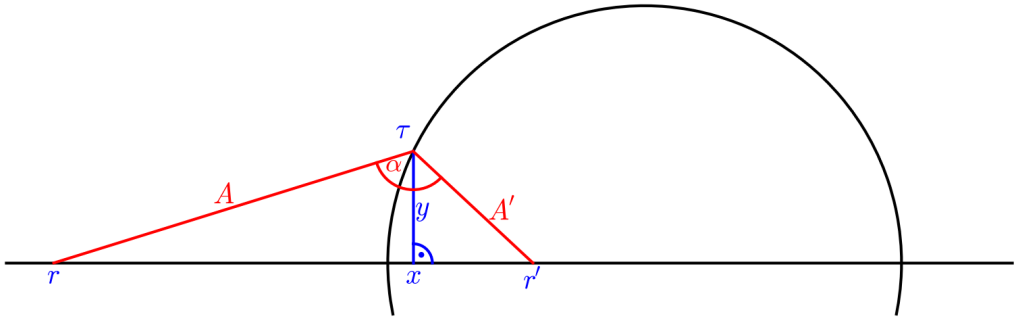
If we denote

$$c = \frac{m'}{m} \cdot \sqrt{\frac{Q'}{Q}} > 0$$

then the above equation (10) becomes

$$|\tau - r| = c \cdot |\tau - r'|.$$

The set of all complex numbers satisfying this equation forms a circle around either r or r'. More precisely, if $c > 1$ then it is a circle with r in its interior and r' in the exterior. Conversely if $c < 1$ then it is r' that lies in the interior and r in the exterior. Finally if $c = 1$, we deal with a vertical line bisecting the segment from r to r'. Write $A = |\tau - r|$ and $A' = |\tau - r'|$.



Our aim now is to maximise the function in (10), which is the same as to maximise $\frac{y}{A^2}$, on this circle $A = c \cdot A'$. We have

$$\frac{y}{A^2} = \frac{y}{A \cdot A' \cdot c} = \frac{\sin(\alpha)}{|r' - r| \cdot c}$$

where α is the acute angle between the segments from τ to r and r' respectively. This is maximal when $\alpha = \pi/2$. So τ is the intersection of the circle $A = c \cdot A'$ with the circle centred on the real axis and passing through r' and r. It is now easy to compute that

$$y = \frac{c}{c^2+1} \cdot |r' - r| = \frac{\sqrt{Q\,Q'}}{m^2\,Q + m'^2\,Q'} \cdot |a\,m' - a'\,m|$$
$$x = \frac{c^2 r' + r}{c^2+1} = \frac{a\,m\,Q + a'\,m'\,Q'}{m^2\,Q + m'^2\,Q'}$$

The maximum value for the speed in (10) is

$$\frac{1}{m\,m'\sqrt{Q\,Q'} \cdot |r - r'|} = \frac{1}{\sqrt{QQ'} \cdot |am' - a'm|}.$$

Furthermore, we find

$$W_r(\tau) = \frac{1}{Q}\frac{Qa'u + m'v}{am' - a'm} + \frac{i}{\sqrt{QQ'} \cdot |am' - a'm|} \qquad \text{and}$$
$$W_{r'}(\tau) = \frac{1}{Q'}\frac{Q'au' + mv'}{a'm - am'} + \frac{i}{\sqrt{QQ'} \cdot |am' - a'm|}.$$

We could not spot any general rule to distinguish the cases when the direct or the indirect method is faster. In practice it is easy to test before starting to sum. For the curve E=5077a1 and $\varepsilon = 0.001$, the direct method is faster for $(r, r') = (0, \frac{70}{5077})$, but slower for $(r, r') = (\frac{123}{456}, \frac{789}{5077})$.

5 Computation of non-unitary symbols

If N is not square-free then there are modular symbols that we do not know how to compute with the above methods. In the section, we analyse how to compute $\lambda(r)$ when r is non-unitary. We cannot move the cusp to $i\infty$ using an Atkin-Lehner involution. If the elliptic curve admits a quadratic twist $E^\dagger$ whose conductor is square-free, then it is best to use the formula for twisting modular symbols, see Section 7.2. But this is not always possible.

There is one special case when we can transform a non-unitary symbol to a unitary one: Suppose $4 \mid N$ and $r = \frac{a}{2m}$ with odd a and m. Then the action of the Hecke operator T_2 yields the equality

$$\lambda\Big(\frac{r'}{2}\Big) + \lambda\Big(\frac{r'-1}{2}\Big) = 0$$

because $a_2 = 0$. For $r'/2 = r$, we get $\lambda(r) = -\lambda\big(\frac{a-m}{2m}\big)$. The latter is now at a cusp with an odd denominator and has a chance of being a unitary cusp. This little trick only works for $4 \mid N$ not any other square dividing N.

In general, however, we know no better method than to rewrite $\lambda(r)$ as the sum of so-called transportable symbols via the use of a Hecke operator. We start by explaining what transportable symbols are and how they can be computed.

5.1 Transportable modular symbols

Definition. We will call $\lambda(\{r' \to r\})$ a *transportable* modular symbol if the two rational numbers r and r' are $\Gamma_0(N)$-equivalent.

This is a more restrictive definition of this term than in [22] where they allow also sums of transportable symbols in the more general setting of higher weight modular forms.

Let $\lambda(\{r' \to r\})$ be a transportable modular symbol. We may compute it by transporting the path: if $\gamma \in \Gamma_0(N)$ is such that $r' = \gamma(r)$, then

$$\lambda\big(\{r' \to r\}\big) = 2\pi i \int_{\gamma(r)}^{r} f(z)dz = 2\pi i \int_{\gamma(\tau)}^{\tau} f(z)dz = \lambda(\tau) - \lambda(\gamma(\tau)) \tag{11}$$

for any τ in the upper half plane.

Write $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Note first that if γ is not hyperbolic, i.e., if $|a + d| \leqslant 2$, then there is a point τ in the upper half plane or among the cusps with $\gamma(\tau) = \tau$ and thus $\lambda\big(\{r' \to r\}\big) = 0$. Hence we may assume that γ is hyperbolic.

Let us now find the best choice of $\tau = x + yi$ in the upper half plane. It will be such that the speeds of summing up $\lambda(\tau)$ and $\lambda(\gamma(\tau))$ are equal and as large as possible. This implies that

$$(cx + d)^2 + c^2y^2 = 1.$$

We want to maximise y under this restriction, so obviously the best choice is $y = \frac{1}{|c|}$ and $x = -\frac{d}{c}$ and the speed will be $1/|c|$. See Algorithm 10.6 in [21]. Since $c \in N\mathbb{Z}$, the speed will be smaller than $\frac{1}{N}$, which is often quite worse than the previous methods.

Given two $\Gamma_0(N)$-equivalent cusps r and r', we should try to find the matrix $\gamma \in \Gamma_0(N)$ with $\gamma(r) = r'$ in such a way as to make its lower left entry c as small as possible in absolute value. Proposition 2.2.3 in [7] gives an algorithm to construct such a matrix, but reading carefully the proof one sees that it actually gives the construction of all possible γ. We repeat it here in our notations for the convenience of the reader.

Write $r = \frac{e}{m}$ and $r' = \frac{e'}{m'}$ in reduced fractions. Using the euclidean algorithm, we can find matrices $\delta = \left(\begin{smallmatrix} e & u \\ m & v \end{smallmatrix}\right)$ and $\delta' = \left(\begin{smallmatrix} e' & u' \\ m' & v' \end{smallmatrix}\right)$ in $\mathrm{SL}_2(\mathbb{Z})$ such that $\delta(i\infty) = r$ and $\delta'(i\infty) = r'$. We have that $\gamma_0 = \delta' \cdot \delta^{-1} = \left(\begin{smallmatrix} a_0 & b_0 \\ c_0 & d_0 \end{smallmatrix}\right)$ sends r to r'. We can obtain all such matrices as

$$\gamma = \delta' \cdot \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \cdot \delta^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

for some $t \in \mathbb{Z}$, because $\delta'^{-1}\gamma\delta$ stabilises the cusp $i\infty$. (Alternatively one can view this as the indeterminacy of u and v in the Bézout equation $ev - mu = 1$ modulo e and m respectively.) The equation

$$0 \equiv c = c_0 + t\, m\, m' \pmod{N}$$

is solvable in t if and only if r and r' are $\Gamma_0(N)$-equivalent. The choice of c is unique up to multiples of $\mathrm{lcm}(mm', N)$. So we can take t such that c is the least residue modulo $\mathrm{lcm}(mm', N)$; hence we have that the speed will be at least $\frac{2}{\mathrm{lcm}(mm', N)}$.

Now so far, we have considered to transport the path close to the cusp $i\infty$. However, we could also choose another unitary cusp r_0 of width Q instead. We compute

$$\lambda\big(\{r' \to r\}\big) = 2\pi i \left(\int_{r_0}^{\tau} - \int_{r_0}^{\gamma(\tau)} \right) f(z)dz = \epsilon_Q \Big(\lambda\big(W_{r_0}(\tau)\big) - \lambda\big(W_{r_0}(\gamma(\tau))\big) \Big)$$

by (6). Renaming $W_{r_0}(\tau)$ as τ and writing $\gamma_{r_0} = W_{r_0} \cdot \gamma \cdot W_{r_0}^{-1} \in \Gamma_0(N)$, this is equal to

$$\lambda\big(\{r' \to r\}\big) = \epsilon_Q \Big(\lambda(\tau) - \lambda\big(\gamma_{r_0}(\tau)\big) \Big)$$

for all τ in the upper half plane. The best γ is obtained when the lower left entry of γ_{r_0} is minimal. Again, this entry is divisible by N and we expect a rather low speed.

For example, we can take $r_0 = 0$ of width N. Then $\gamma_{r_0} = \left(\begin{smallmatrix} d & -c/N \\ -bN & a \end{smallmatrix}\right)$ and so we are now looking for γ such that $|b|$ is minimal. As before $b = b_0 + t\, e\, e'$ and we are looking for the least residue of b_0 modulo ee'. It may be that the resulting computation is faster with r_0 than with $i\infty$. It seems difficult to find the best choice of the unitary cusp r_0 in general.

Finally, we could also transport the path in such a way as to have $\gamma(\tau)$ close to $i\infty$ and τ close to another unitary cusp r_0. For instance if $r_0 = 0$, this would give a speed of $1/\big(\sqrt{N}|d|\big)$ and we would have to minimise $|d|$. However, this time it will also involve the computation of the integral from $i\infty$ to r_0.

5.2 Hecke operators to get transportable paths

Let $r = \frac{a}{m}$ be a non-unitary cusp. Set M to be the greatest common divisor of m and N. Further put d equal to the greatest common divisor of M and $Q = \frac{N}{M}$. The previous methods explain how to compute $\lambda(r)$ only in the case that $d = 1$. In this section, we will suppose $d > 1$.

First, for any integer n coprime to N, we have the action of the Hecke operator, which gives us

$$a_n \cdot \lambda(r) = \sum_{k|n} \sum_{u=0}^{k-1} \lambda\Big(\frac{nr + uk}{k^2}\Big).$$

The cusp $\frac{nr+uk}{k^2}$ is $\Gamma_0(N)$-equivalent to r if and only if $n \cdot k^{-2}$ is congruent to 1 modulo d. This implies that $n \equiv 1 \pmod{d}$ and that $k^2 \equiv 1 \pmod{d}$ for all divisors $k \mid n$. If n is not a prime or a square of a prime, then the smallest prime divisor of n will provide a smaller choice for n.

Let ℓ be a prime congruent to 1 modulo d. If ℓ does not divide N then

$$\big(a_\ell - \ell - 1\big)\lambda(r) = \lambda\big(\{\ell r \to r\}\big) + \sum_{u=0}^{\ell-1} \lambda\Big(\Big\{\frac{r+u}{\ell} \to r\Big\}\Big).$$

The right hand side is now a sum of $\ell+1$ transportable symbols. The integer $a_\ell - \ell - 1 = -N_\ell$ is non-zero since N_ℓ is the number of points on the reduction of E to $\mathbb{F}_\ell$. If ℓ divides N, then we get

$$\big(a_\ell - \ell\big)\lambda(r) = \sum_{u=0}^{\ell-1} \lambda\Big(\Big\{\frac{r+u}{\ell} \to r\Big\}\Big)$$

instead. This time $|a_\ell| \leqslant 1$.

The other option is to take a prime ℓ such that $\ell^2 \equiv 1 \pmod{d}$. For instance, let $\ell \equiv -1 \pmod{d}$. Then we have the following formula

$$\big(a_{\ell^2} - \ell^2 - \ell - 1\big)\lambda(r) = \lambda\Big(\{\ell^2 r \to r\}\Big) + \sum_{u=0}^{\ell-1} \lambda\Big(\Big\{r + \frac{u}{\ell} \to r\Big\}\Big) + \sum_{v=0}^{\ell^2-1} \lambda\Big(\Big\{\frac{r+v}{\ell^2} \to r\Big\}\Big)$$

which expresses a non-zero multiple of $\lambda(r)$ as a sum of transportable symbols. If ℓ is the smallest prime congruent to 1 modulo d and $\ell' \not\equiv 1 \pmod{d}$ is the smallest prime such that $\ell'^2 \equiv 1 \pmod{d}$, then the above formula for ℓ' will have $\ell'^2 + \ell' + 1$ terms, which may be smaller than the $\ell+1$ terms in the corresponding sum for ℓ. Although not frequent, there are cases when this is useful. For instance if $d = 6441$, we have $\ell = 231877$ and $\ell' = 227$.

It is hard to estimate what the complexity of this method is. It is certainly significantly slower than the computation of unitary cusps, but it is still useful when the conductor is not too large. In the most frequent applications, like for the computation of p-adic L-series, this is not important, as we will be mainly interested in unitary symbols. Note however that the following section shows that even the computation of unitary symbols for large denominators may encounter the computations explained here.

6 Manin's trick using continued fractions

Manin [16] introduced the use of the continued fraction expansion of the rational r to help speeding up the computation of $[r]^\pm$ considerably when the denominator of r is large compared to N. See also [12] and Section 3.3.1 in [21] for more details. However, we need to modify it slightly here as we should avoid non-unitary cusps if at all possible.

Definition. Recall that the set of right coset representatives of $\Gamma_0(N)$ in $\mathrm{SL}_2(\mathbb{Z})$ is in bijection with $\mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$ by sending $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to $(c:d)$. For each such coset $\Gamma_0(N)\delta$ with $\delta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we define the *Manin symbol* by

$$M(c:d) = 2\pi i \int_{i\infty}^{0} f|_\delta(z)dz = 2\pi i \int_{a/c}^{b/d} f(z)dz = \lambda\Big(\frac{b}{d}\Big) - \lambda\Big(\frac{a}{c}\Big). \tag{12}$$

We start by explaining how to reduce the computation of $[r]^\pm$ for a large denominator of r to the computation of Manin symbols and then explain how to evaluate Manin symbols.

6.1 Using continued fractions

Here is the original trick by Manin. We are given a rational number $r = a/m$. Consider the sequence of convergents of the continuous fraction of r:

$$\frac{a_{-1}}{m_{-1}} = \frac{1}{0}, \qquad \frac{a_0}{m_0} = \frac{a_0}{1}, \qquad \dots, \qquad \frac{a_n}{m_n} = \frac{a}{m}.$$

We have $a_k m_{k-1} - a_{k-1} m_k = (-1)^{k-1}$. So the matrix

$$\begin{pmatrix} a_k & (-1)^{k-1} a_{k-1} \\ m_k & (-1)^{k-1} m_{k-1} \end{pmatrix}$$

belongs to $\mathrm{SL}_2(\mathbb{Z})$ and it sends any path linking 0 to $i\infty$ to a path from $\frac{a_{k-1}}{m_{k-1}}$ to $\frac{a_k}{m_k}$. We find

$$\begin{aligned}\lambda(r) &= -2\pi i \cdot \left(\int_{a_n/m_n}^{a_{n-1}/m_{n-1}} + \int_{a_{n-1}/m_{n-1}}^{a_{n-2}/m_{n-2}} + \cdots + \int_{a_1/m_1}^{a_0/m_0} + \int_{a_0}^{i\infty} \right) f(z)dz \\ &= M\big(m_n : (-1)^{n-1} m_{n-1}\big) + M\big(m_{n-1} : (-1)^{n-2} m_{n-2}\big) + \cdots + M\big(m_1 : 1\big) + M\big(1 : 0\big).\end{aligned}$$

This allows to compute $\lambda(r)$ as a sum of Manin symbols $M(c : d)$, each of which is a modular symbol between two rational numbers of denominator c and d smaller than N.

Now the problem with this way of splitting up is the following: Even if r is a unitary cusp, it may be that some intermediate convergent a_k/m_k is not unitary. Here is an adaptation, which may take a few steps more, but tries to avoid non-unitary cusps. In the end this is a great gain of speed.

Algorithm: Try to split up the path into unitary Manin symbols

[Initialisation]: Given $r = a/m$. If $m = 1$, return $\lambda(0)$.

[Find new cusp]: Compute with the extended euclidean algorithm x and y such that $a\,y + x\,m = 1$. Make sure that $-m/2 < y \leqslant m/2$.

[Unitary?]: If $-x/y$ is unitary, set $r' = -x/y$. Otherwise, set $r' = (x + \mathrm{sign}(y)\,a)/(y - \mathrm{sign}(y)\,m)$ if that is unitary. If both are non-unitary, set $r' = -x/y$.

[Recursion]: Call this function recursively with r' and add the result to the Manin symbol $M(m : y)$.

Here is an example of a case when both choices of cusps are non-unitary: For $N = 36$ and $r = \frac{2}{5}$, neither $\frac{1}{2}$ nor $\frac{1}{3}$ is unitary. This can only happen when the squares of two distinct primes divide N.

Note that if we have to go for the second choice for the cusp, then we still have $|y| < m$, but not $|y| < m/2$. So we are not certain any more if the algorithm takes only $O(\log(m))$ steps. In practice, the algorithm is quite effective in avoiding non-unitary cusps. We tested all elliptic curves of conductor at most 1000 which are not semistable and whose conductor cannot be decreased by a quadratic twist. Among all a/m with $m < N$, there were 77% such that the best choice for r' is unitary, for 22% the second best choice is unitary and only in 1.4% we have to pass to a non-unitary cusp r'.

6.2 Unitary Manin symbols

As explained above, we now have to compute the Manin symbol $M(c : d)$ as defined in (12). We assume here first that both c and d are denominators of unitary cusps. In this case, we say that the Manin symbol $M(c : d)$ is unitary. Note, that once we computed $M(c : d)$, we also know $M(-d : c) = -M(c : d)$. This is the formula (2.2.6) in [7]. Further $\overline{M(c : d)} = M(c : -d)$. Also, there is a three term relation $M(c : d) + M(c + d : -c) + M(d, -c - d) = 0$; which can be used to compute a further value if two of them are known.

There are now at least three possible ways of evaluating the Manin symbol $M(c : d)$. Either by direct or indirect integration or by using transportation. Further note that $M(c : d)$ only depends on $(c : d)$ in $\mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$ and we may improve the performance by choosing good representatives c and d.

First, in the cases when both c and d are coprime to N, we could transport them as both cusps $\frac{a}{c}$ and $\frac{b}{d}$ are $\Gamma_0(N)$-equivalent to 0. From the Section 5.1, we see that the speed will be at best equal to $\frac{1}{N}$ and hence this method will usually lose out on the others below.

Let Q be the width of $\frac{a}{c}$ and Q' be the width of $\frac{b}{d}$. Then the speed of using the direct integration from $\frac{a}{c}$ to $\frac{b}{d}$ is equal to

$$\frac{1}{\sqrt{QQ'} \cdot |ad - bc|} = \frac{1}{\sqrt{QQ'}} \geqslant \frac{1}{N}$$

as seen in Section 4.3. In the (most frequent) case when c and d are coprime to N, then the speed is indeed equal to $\frac{1}{N}$. Neglecting the contribution from ε, this means that we expect a single sum over approximately $\frac{1}{2\pi} N \log(N)$ terms.

By Section 4.2, the indirect integration via $i\infty$ instead uses two sums with speed $\big(|c|\sqrt{Q}\big)^{-1}$ and $\big(|d|\sqrt{Q'}\big)^{-1}$ each. If we neglect again the contribution from ε, we expect in the case $\gcd(cd, N) = 1$ to sum in total about

$$\tfrac{\sqrt{N}}{2\pi}\Big(|c|\,\log\big(|c|\sqrt{N}\big) + |d|\,\log\big(|d|\sqrt{N}\big)\Big).$$

In particular, if we can find c and d representing the point on the projective line with $|c|$ and $|d|$ both smaller than $\frac{1}{2}\sqrt{N}$, then the indirect method is faster. This leads to the problem of finding small c and d. The following lemma shows that we may just as well try to minimise $|c| + |d|$.

Lemma 8. *Let $C = \sqrt{N}/(2\pi)$. Let $\gamma\colon \mathbb{R}^2 \to \mathbb{R}_{\geqslant 0}$ be the continuous function such that $\gamma(x,y) = C\big(|x|\log(|x|\sqrt{N}) + |y|\log(|y|\sqrt{N})\big)$ for $xy \neq 0$. Let $L \subset \mathbb{Z}^2$ be a set not containing the origin. Let (x_0, y_0) be a point of L at which γ is minimal and let (x_1, y_1) be a point in L at which $|(x,y)| = |x|+|y|$ is minimal. Then*

$$\frac{\gamma(x_1,y_1)}{\gamma(x_0,y_0)} = 1 + \mathbf{O}\Big(\frac{1}{\log(N)}\Big).$$

Proof. Write $A = |(x_1, y_1)|$. Since γ is increasing on rays leaving from the origin, we see that

$$\gamma(x_1,y_1) \leqslant \max\Big\{\gamma(x,y) \;\Big|\; |(x,y)| = A\Big\}$$
$$\gamma(x_0,y_0) \geqslant \min\Big\{\gamma(x,y) \;\Big|\; |(x,y)| = A\Big\}$$

It is not hard to show that the maximum above is $C\cdot\log(A\sqrt{N})$ and the minimum is $C\cdot\log(A\sqrt{N}/2)$. Hence we find

$$\frac{\gamma(x_1,y_1)}{\gamma(x_0,y_0)} \leqslant \frac{1}{1-\frac{\log(2)}{\log(A\sqrt{N})}} = 1 + \mathbf{O}\Big(\frac{1}{\log(N)}\Big). \qquad \square$$

6.3 Small coordinates of projective points

Let N be an integer and $(u : v) \in \mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$. In the above computation of Manin symbols, we came across the problem of finding the integers c and d such that $(u : v) = (c : d)$ and $|c| + |d|$ is as small as possible. Write $|(c,d)| = |c| + |d|$ and $\|(c,d)\| = \sqrt{c^2+d^2}$.

We are looking for the smallest non-zero vector in the lattice

$$\Lambda_{(u:v)} = \Big\{(c,d) \in \mathbb{Z}^2 \;\Big|\; c\cdot v \equiv d\cdot u \pmod N\Big\}$$

such that $\gcd(c,d) = 1$. Here is an algorithm based on Algorithm 1.3.14 in [6].

Algorithm: Find good representatives for projective points

[Initialise]: Set $\vec{x}, \vec{y}$ to be a $\mathbb{Z}$-basis of $\Lambda_{(u:v)}$. If one of the coordinates u or v is invertible modulo N, say v, then we can do this as follows: Set w to be the product of u and the inverse of v modulo N. Let $\vec{x} = (1, w)$ and $\vec{y} = (0, N)$. In the general case, we set $p = \gcd(u, N)$ and $q = \gcd(v, N)$; note that they must be coprime. Set w to be the product of $\frac{u}{p}$ and the inverse of $\frac{v}{q}$ modulo $\frac{N}{pq}$. Then $\vec{x} = (\frac{N}{q}, 0)$ and $\vec{y} = (w\cdot p, q)$ is a basis.

[Euclidean step]: If the signs of x_0 and x_1 agree, then set r to be the greatest integer smaller than $\frac{y_0+y_1}{x_0+x_1}$. Otherwise set r to be the greatest integer smaller than $\frac{y_0-y_1}{x_0-x_1}$. Set $\vec{z} = \vec{y} - r\cdot\vec{x}$. If $|\vec{z} - \vec{x}| < |\vec{z}|$, then replace $\vec{z}$ by $\vec{z} - \vec{x}$.

[Finished ?]: If $|\vec{z}| < |\vec{x}|$, then set $\vec{y}$ to $\vec{x}$ and $\vec{x}$ to $\vec{z}$ and go back to the second step. Otherwise we can terminate the algorithm. If the coordinates of $\vec{x}$ are coprime, we return $\vec{x}$. If not, we run through small linear combinations of $\vec{x}$ and $\vec{z}$, starting with $\vec{z}$, until we hit one with coprime coordinates.

The proof is very analogous to the one in [6]. As long as we do the second step, we know that $\vec{x}$ and $\vec{y}$ are a $\mathbb{Z}$-basis of the lattice $\Lambda_{(u:v)}$. The integer r is chosen such that $|\vec{z}|$ is minimal. At the stage when we terminate, we are certain that $\vec{x}$ is the shortest non-zero vector of the lattice and $\vec{z}$ is the shortest, which is not a multiple of $\vec{x}$. The convex body theorem of Minkowski applied to the set of vectors of $|\cdot|$-norm at most $\sqrt{2N}$ guarantees that $|\vec{x}| \leqslant \sqrt{2N}$.

Unfortunately, we cannot be certain that the algorithm will return the best of all choices. For instance with $N = 30$ and $(u : v) = (11 : 1)$, we find that the shortest non-zero vector is $\vec{x} = (3, 3)$ and the second minimum is $\vec{z} = (5, -5)$. None of them is allowed to represent $(11 : 1)$ in $\mathbb{P}^1\big(\mathbb{Z}/30\mathbb{Z}\big)$. Even $\vec{x} + \vec{z} = (8, -2)$ and $\vec{z} - \vec{x} = (2, 8)$ are not permitted. Only when we compute $2\vec{x} - \vec{z} = (11, 1)$ and $\vec{x} + 2\vec{z} = (13, -7)$ will we find coprime coordinates. It is now not certain that the algorithm will find the shorter one first. Note that in this example $|(11, 1)| = 12$ is much larger than $\sqrt{60}$.

The following is a theoretical result about small coordinates for projective points that will be used later in Section 8.5.

Lemma 9. *There exists an absolute constant C with the following property. Let N be a square-free integer and let $P = (u : v) \in \mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$. Let $\vec{v}_P$ be the shortest non-zero vector in Λ_P and let $\vec{w}_P$ be the shortest vector in Λ_P which is not collinear to $\vec{v}_P$. Then there exists $\lambda \in \mathbb{Z}$ with $|\lambda| \leqslant C \cdot \log(N)^2$ such that the two coordinates of $\vec{w}_P + \lambda\vec{v}_P$ are coprime.*

In particular, there exists (c, d) such that $(c : d) = (u : v)$ and

$$\max\big(|c|, |d|\big) \leqslant \frac{N}{\|\vec{v}_P\|} + C' \log(N)^2 \|\vec{v}_P\|$$

for some absolute constant C'.

Note that it is vain to hope for a better bound, for instance independent of the size of $\|\vec{v}_P\|$. Suppose $N = 2n$ is even. Then the size of the coordinates of the point $P = (1 : n)$ cannot be decreased. For this example $\vec{v}_P = (2, 0)$ is very small.

Proof. We will call the content, written $\mathrm{co}(x, y)$, of a point (x, y) in $\mathbb{Z}^2$ the greatest common divisor of the two coordinates x and y. Since $(u, v) \in \Lambda_P$ and u and v are coprime, there exists at least one point with content 1 in Λ_P. It follows that the contents of two basis vectors of Λ_P must be coprime integers. In particular $\mathrm{co}(\vec{v}_P)$ and $\mathrm{co}(\vec{w}_P)$ are coprime.

Let $\vec{z} = (x, y)$ be a vector in Λ_P. Then there exists an integer k such that $cx - dy = kN$. If $b = \gcd\big(k, \mathrm{co}(\vec{z})\big)$, then $\big(x/b, y/b\big)$ also belongs to Λ_P. Hence if we assume now that $\vec{z}$ is not divisible by any integer greater than 1, then $b = 1$. Thus $\mathrm{co}(\vec{z})$ divides N. In particular all points on the line $\mathcal{L} = \big\{\vec{w}_P + \lambda\vec{v}_P \;\big|\; \lambda \in \mathbb{Z}\big\}$ have contents equal to a divisor of N.

Consider two points $\vec{z} = \vec{w}_P + \lambda\vec{v}_P$ and $\vec{z}' = \vec{w}_P + \lambda'\vec{v}_P$ on the line $\mathcal{L}$. We claim that the greatest common divisor of $\mathrm{co}(\vec{z})$ and $\mathrm{co}(\vec{z}')$ divides $\lambda - \lambda'$: It is not hard to show that this greatest common divisor divides $(\lambda' - \lambda) \cdot \gcd\big(\mathrm{co}(\vec{v}_P), \mathrm{co}(\vec{w}_P)\big)$ and so the above justifies the claim.

For each prime divisor $\ell \mid N$, either ℓ does not divide the content of any point on $\mathcal{L}$ or the content of every ℓ-th point is divisible by ℓ. Let $\tilde{N}$ be the product of the prime divisors of N dividing the content of one of the points on $\mathcal{L}$. The sequence $\mathrm{co}(\vec{w}_P + \lambda\vec{v}_P)$ as λ varies in $\mathbb{Z}$ is periodic with period $\tilde{N}$. There is λ_0 such that $\vec{z}_0 = \vec{w}_P + \lambda_0\vec{v}_P \in \mathcal{L}$ has content $\tilde{N}$. Now the content of $\vec{w}_P + \lambda\vec{v}_P$ is $\gcd(\lambda - \lambda_0, \tilde{N})$.

By a theorem of Iwaniec [15] on the Jacobsthal function, there is a constant C such that any set of $C(\log(N))^2$ consecutive integers contain at least a unit modulo N. It follows that in the set $X \subset \mathcal{L}$ of $\vec{w}_P + \lambda\vec{v}_P$ with $|\lambda| \leqslant C/2 \log(N)^2$ there is a point whose coordinates are coprime.

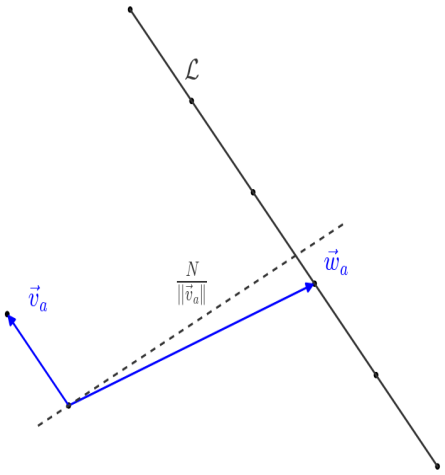


Figure 1: The two shortest vectors and the line $\mathcal{L}$

The last sentence of the lemma follows from geometric considerations (see Figure 1) measuring the length of this vector in X: The distance from $(0, 0)$ to the real line containing $\mathcal{L}$ is $N/\|\vec{v}_P\|$. The length of the point (x, y) in the set X furthest away from $(0, 0)$ satisfies

$$\|(x, y)\| \leqslant \frac{N}{\|\vec{v}_P\|} + \big(C/2 \log(N)^2 + 1\big)\|\vec{v}_P\|$$

by the triangle inequality. Finally we use $\max(|x|, |y|) \leqslant \|(x, y)\|$. □

We also remark that when $N = p$ is prime, we have the much better bound $|c| + |d| \leqslant \sqrt{2N}$: The content of $\vec{v}_P$ can only be 1 or p. But if it were p, then the representation of the form $P = (1 : d)$ with $0 \leqslant d < p$ or $(0 : 1)$ would be a smaller vector in Λ_P. Hence the shortest vector is always the best way to represent the point on $\mathbb{P}^1\big(\mathbb{Z}/p\mathbb{Z}\big)$. By Minkowski's convex body theorem $|\vec{v}_P| \leqslant \sqrt{2N}$.

6.4 Non-unitary Manin symbols

Let $(c : d)$ be such that at least one of them is not the denominator of a unitary cusp. For simplicity, we assume that d is the denominator of a unitary cusp and c is not. Given how much harder it is to work with non-unitary cusps, we should compute $M(c : d)$ as $\lambda(b/d) - \lambda(a/c)$ and we have to make c as small as possible.

Given an integer N and $(u : v) \in \mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$, we are looking for $(c : d) = (u : v)$ such that $|c|$ is minimal. Let $M = (u, N)$ and Q such that $N = MQ$. We can take $c = M$, which is minimal. The other coordinate d has now to satisfy $Mv \equiv du \pmod N$ and $(M, d) = 1$. Let x and y such that $xu + yN = M$. The congruence condition becomes $d \equiv xv \pmod Q$. Our first choice would be to take $d = xv$. However in case xv and M are not coprime, we add Q to xv until it becomes coprime to M.

7 Tweaks

In this section, we present two ideas to make certain computations faster.

7.1 Using partial sums

Let E be an elliptic curve over $\mathbb{Q}$. Let m be a small positive integer. Here is an idea that is useful for the evaluation of all symbols $[\frac{a}{m}]^{\pm}$ as a varies through all integers $1 \leqslant a < m$ coprime to m. In the application where we wish to evaluate a p-adic L-series for some small prime p, we would typically need this for $m = p^2$ or p^3. For the sake of simplicity we assume that $\frac{1}{m}$ is unitary.

In equation (3), we have defined the partial sums

$$\kappa_{j,m}(y) = \sum_{\substack{n \geqslant 1 \\ n \equiv j \bmod m}} \frac{a_n}{n} \exp(-2\pi n y).$$

We have seen that we only need m more terms in the sum to evaluate to a given precision all these partial sums for $j = 0, \dots, m-1$.

These can be used to evaluate $\lambda(\tau)$ whenever the real part x of τ is a rational number with denominator m, say $x = \frac{a}{m}$:

$$\lambda\Big(\frac{a}{m} + yi\Big) = \sum_{j=0}^{m-1} \kappa_{j,m}(y) \cdot \zeta^{ja} \tag{13}$$

where $\zeta = \exp(2\pi i/m)$. In case we are only interested in the plus modular symbols $[\cdot]^+$, we can do the computations with real numbers only.

$$\mathrm{Re}\Big(\lambda\big(\tfrac{a}{m} + yi\big)\Big) = \sum_{j=0}^{m-1} \kappa_{j,m}(y) \cdot \cos\big(2\pi a j/m\big).$$

We see here that it is possible to use fast Fourier transform if we are interested in evaluating $\lambda\big(\frac{a}{m} + yi\big)$ for all a with a fixed m and $y > 0$. Note that the radix m cannot be chosen to be a power of two, so we rely on mixed-radix algorithms. This has not yet been implemented in [28].

We can use the above formula (13) together with equations (2), (8) and (9), to give a formula for the computation of $\lambda(\frac{a}{m})$ for all a at once:

$$\lambda\Big(\frac{a}{m}\Big) = \sum_{n=1}^{\infty} \frac{a_n}{n} \cdot e^{-\frac{2\pi n}{m\sqrt{Q}}} \cdot \Big(e^{\frac{2\pi n a}{m} i} - \epsilon_Q e^{-\frac{2\pi n u}{m} i}\Big) = \sum_{j=0}^{m-1} \kappa_{j,m}\Big(\frac{1}{m\sqrt{Q}}\Big) \cdot \Big(\zeta_m^{ja} - \epsilon_Q \zeta_m^{-ju}\Big). \tag{14}$$

where u is an inverse of Qa modulo m and $\zeta_m = \exp(2\pi i/m)$.

Similarly, we can express the direct integration from $r' = \frac{a'}{m'}$ to $r = \frac{a}{m}$ as a finite sum of partial sums: Let Q and Q' be the widths and set $d = \mathrm{lcm}(Q, Q') \cdot |am' - a'm|$ and $y = \sqrt{QQ'} \cdot |am' - a'm|$

and let τ be the optimal place in the upper half plan to cut the path in two, which we found in Section 4.3. Then

$$W_r(\tau) = \frac{\xi}{d} + \frac{i}{y}$$

where $\xi = (Qa'u + vm')Q'/\gcd(Q, Q')$ and $Qau + vm = 1$. Hence we obtain

$$\lambda\big(\{r' \to r\}\big) = \sum_{j=0}^{d-1} \kappa_{j,d}\Big(\frac{1}{y}\Big)\Big(\epsilon_{Q'}\zeta_d^{j\xi'} - \epsilon_Q\zeta_d^{j\xi}\Big).$$

with $\xi' = (Q'au' + v'm)Q/\gcd(Q, Q')$ and $Q'a'u' + v'm' = 1$. Note however, that to use this formula only makes sense when d is much smaller than N.

Finally, we could also compute the transportable symbols using partial sums. For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we find

$$\lambda\big(\{r' \to r\}\big) = \sum_{j=0}^{|c|-1} \kappa_{j,|c|}\Big(\frac{1}{|c|}\Big)\Big(\zeta_c^{-dj} - \zeta_c^{aj}\Big).$$

We explain why it can be beneficial to use these partial sums: Even when computing a single one of these expressions, say $\lambda(\frac{a}{m})$ for some value of m, it may be worth wasting a bit of time and using the above formulae. We first compute all $\kappa_{j,m}(y)$ in one sum with $T(y, \varepsilon) + m$ terms. Then we do one sum involving m terms again. Hence if m is small, say $m \ll \sqrt{N}$, we lose only very little time. Since $1/y^2$ is an integer in all cases above, it is easy to cache the values $\kappa_{j,m}(y)$ for later use. If we then encounter later another symbol with the same denominator m, we have to sum up only m precomputed terms.

However note that this is not practical for transportable symbols or for the computation of all Manin symbols as m will be in the order of N rather frequently.

7.2 Quadratic twists

If N is not square-free then one can often find a quadratic twist of the elliptic curve with smaller conductor. Since all the previous computations depend heavily on the conductor, it may be an advantage to do the computation on the twisted curve instead.

Let D be a fundamental discriminant such that the quadratic twist $E^\dagger$ by D has minimal conductor among all quadratic twists of E. This needs not be unique, but for our considerations it does not seem to matter much which among them we choose. In practice we take the one with the largest period is best.

Write $\sqrt{D}$ for the square root of D in $\mathbb{R}_{>0}$, if D is positive, and in $i\mathbb{R}_{>0}$, if D is negative. We will use formula (I.8.5) in [18]

$$\lambda(r) = \frac{1}{\sqrt{D}} \sum_{u=1}^{|D|-1} \Big(\frac{D}{u}\Big)\lambda^\dagger\Big(r + \frac{u}{|D|}\Big)$$

where $\lambda^\dagger$ designates the modular symbol for the twisted elliptic curve. Since the rational numbers $r \pm \frac{u}{D}$ all have the same denominator, we can use the idea from the previous section to compute this sum with a single summation. Similar, if we wish to compute all modular symbols for E with a given denominator.

Note however that there is a small issue with this. Suppose ℓ is a prime dividing D such that the conductor $N^\dagger$ of the twisted curve $E^\dagger$ is still divisible by ℓ^2. This can happen for instance with $N = 80$, $D = -4$, and $N^\dagger = 40$. Now in this situation, we will evaluate modular symbols with denominator divisible by ℓ. If ℓ did not divide the denominator of r, then the resulting cusp $r + \frac{u}{D}$ will not be unitary. Because our method is very much slower for non-unitary cusps, it is much better to avoid this. Hence we will remove all factors of ℓ in the fundamental discriminant if the twisted curve will still have additive reduction at ℓ. Of course this affects only $\ell = 2$ or 3.

How much do we expect this to speed up our computations? We will use the notation $\mathbf{O}(f(N))$ to mean that the number of steps needed in the computations is, for sufficiently big N, bounded by $C \cdot f(N)$ for some constant $C > 0$. Suppose we wish to evaluate $\lambda(r)$ for a rational r with denominator m, which we suppose for simplicity to be coprime to N. We will compute about $\log(m)$ Manin symbols each with at worst a speed of $\frac{1}{N}$. So we will be summing about $\mathbf{O}\big(\log(m)\log(N)N\big)$ terms in total.

Instead, using the twist by D, we will have D times as many terms with a denominator of $m \cdot D$, but the conductor will be divided by $D \cdot D'$ where D' is a factor of D. Hence we get about

$\mathbf{O}\big(\log(mD)\log(N/DD')N/D'\big)$ terms to sum. If $D' > 1$, this is obviously a very good improvement. Otherwise it is negligible.

The other major advantage of twisting is that there will be less non-unitary cusps on the twist. In particular when $E^\dagger$ is semistable, then all cusps are unitary for $E^\dagger$. This way, we can compute even the non-unitary symbols for E very quickly.

8 Complexity

In [12] Goldfeld finds the complexity of evaluating one modular symbol on a semistable curve. We will refine this here. We will continue to use the notation $\mathbf{O}(f(N))$ introduced above to find an upper bound on the number of steps in an algorithm. Further the notation $\tilde{\mathbf{O}}(f(N))$ suppresses the possible further factors which are logarithmic in $f(N)$. As mentioned before, we neglect the issues with precision and simply find asymptotics for the number of terms that need to be summed up.

We will assume throughout this section that N is square-free; except for Section 8.4. Recall that this implies that E is semistable and hence the Manin constant c_0 is either 1 or 2 and hence the assumption made in Section 2 can be neglected in this section.

8.1 Periods

Although we have often neglected the size of ε in the previous consideration, we should find a proven lower bound for the size of the periods Ω^+ and Ω^-. This seems however difficult and the issue is already discussed in [12].

Conjecture (Goldfeld's period conjecture)**.** There is a constant $\kappa > 0$ such that Ω^+ and Ω^- are larger than $\mathbf{O}\big(N^{-\kappa}\big)$ as $N \to \infty$.

The graph in Figure 2 presents numerical evidence in favour of this conjecture. In fact it looks like $\kappa = 1$ is a very reasonable guess, while $\kappa < 1/2$ is not likely. In Section 2, we have shown that for

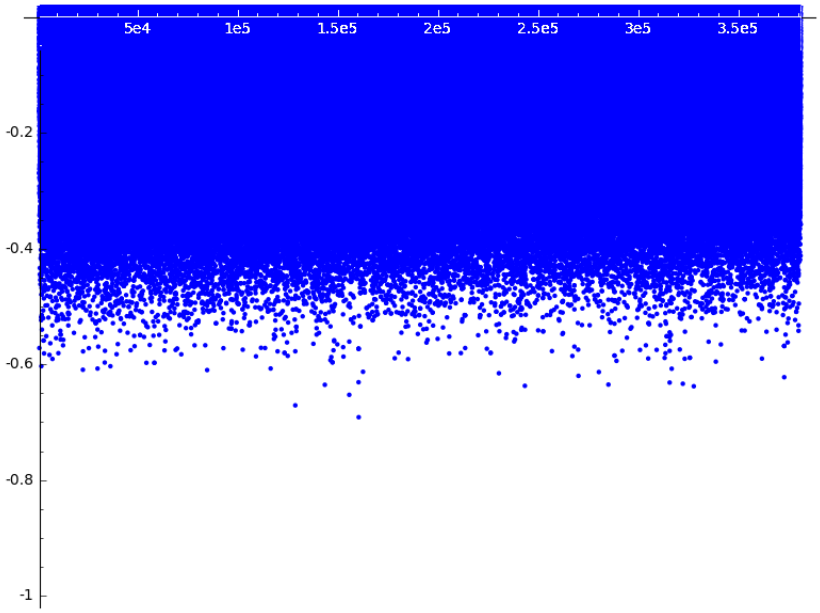


Figure 2: For each elliptic curve in the Cremona tables, the value of $\log\Omega^+/\log N$ on the vertical axis is compared with N on the horizontal axis. Only negative values are plotted.

semistable curves the bound on the denominator of $[r]^\pm$ is at most 24 for the strong Weil curve. Since the number of isogenous curves is also bounded, the denominator won't contribute to the asymptotic size of the error ε. Under the conjecture above, we find that $-\log(\varepsilon) = \mathbf{O}(\log(N))$.

Without assuming the conjecture, it seems that one only knows (see [12]) that the periods are bounded by $\mathbf{O}\big(N^{-N}\big)$. This then gives a proven bound $-\log\varepsilon = \mathbf{O}(N\log(N))$.

8.2 Fourier coefficients

We have to compute the coefficients a_n for n up to a bound T. In practice this is done by the command `ellan` in `PARI`. This function first computes the values a_p for all primes up to T. When p gets large, the preferred choice of algorithm for the Frobenius trace a_p is the Schoof-Elkies-Atkin algorithm, which is known to run in polynomial time, with a heuristic expectation of $\tilde{\mathbf{O}}(\log^4 p)$. Hence to find all a_p for $p < T$, we expect $\tilde{\mathbf{O}}(T)$ operations. The algorithm then uses the recursive formulae and the multiplicativity of a_n. This is done also in about T steps. Therefore in total we expect $\tilde{\mathbf{O}}(T)$ operations.

It is to be noted that in our implementation, this step does indeed take up a certain non-negligible portion of the total computation time. Initially, we precompute the first thousand coefficients a_n. If we later need more terms, we add them. However the way we interact with `PARI` currently it is faster to recompute all values from scratch unless we only have to add a small percentage of new values. Hence in practice, we may have to perform these computations more than once. For the theoretical considerations below, we may assume that we can determine beforehand the highest value of n ever needed and compute all values a_n only once.

8.3 Computing one modular symbol

Suppose $r = \frac{a}{m} \in \mathbb{Q}$ with $0 < a < m$ and we wish to evaluate $[r]^{\pm}$. As we supposed that N is square-free, the cusp r is unitary. We have seen in equations (8) and (9) that we can compute them by integrating to τ with imaginary part equal to $1/\left(m\sqrt{Q}\right)$ where Q is the width of r. Lemma 4 then gives us that we have to sum $T = \mathbf{O}\left(m\sqrt{Q}\log(m\sqrt{Q})\right) + \mathbf{O}(-\log(\varepsilon)\, m\sqrt{Q})$ terms. For this we need to evaluate that many Fourier coefficients, but that is done in $\tilde{\mathbf{O}}(T)$ steps. As $Q \leqslant N$, we find that the total number of steps in the computation is $\tilde{\mathbf{O}}(m\sqrt{N})$ assuming Goldfeld's period conjecture.

Of course, when m is large, one should use Manin's trick in Section 6 instead. Since N is square-free, all cusps are unitary and hence we can split up the computation of $[r]^{\pm}$ into $\mathbf{O}(\log(m))$ Manin symbols. Now using the direct integration from cusp to cusp, any unitary Manin symbol can be computed in $\tilde{\mathbf{O}}(N\log(N)) = \tilde{\mathbf{O}}(N)$ steps. We have now recovered

Theorem 10 (Goldfeld, Theorem 2 in [12])**.** *Assume Goldfeld's period conjecture holds. Then the modular symbol $\left[\frac{a}{m}\right]^{\pm}$ on a semistable curve E defined over $\mathbb{Q}$ of conductor N can be computed in less than $\tilde{\mathbf{O}}(N\log(m))$ steps.*

However, we can often do much better. For instance, when N is prime, then each Manin symbol can be computed in $\tilde{\mathbf{O}}(\sqrt{N}\log(N)) = \tilde{\mathbf{O}}(\sqrt{N})$ steps due to the fact that projective coordinates can always be chosen of size $\mathbf{O}(\sqrt{N})$, see the remark after Lemma 9. In fact a large proportion of Manin symbols are computable at that complexity:

Proposition 11. *Assume Goldfeld's period conjecture holds. For each N, there is a subset $\mathcal{P}$ containing at least 95% of all points on $\mathbb{P}^1\left(\mathbb{Z}/N\mathbb{Z}\right)$ such that each Manin symbol $M(x)$ for $x \in \mathcal{P}$ can be computed in less than $\tilde{\mathbf{O}}(N^{1/2})$ steps.*

Proof. Let $\vec{v}$ be a vector with $\|\vec{v}\| < \sqrt{N}$ and whose coordinates are coprime. Then $\vec{v}$ is the shortest vector in a lattice Λ_P for some $P \in \mathbb{P}^1\left(\mathbb{Z}/N\mathbb{Z}\right)$. Take $\mathcal{P}$ to be the set of all these points. For each P, there is only one other non-zero element of Λ_P, namely $-\vec{v}$, in the ball of radius $\sqrt{N}$. There are approximatively $\frac{1}{2}\frac{6}{\pi^2}\pi\sqrt{N}^2 = \frac{3}{\pi}N$ pairs of opposite points with coprime integers in this ball. This is asymptotically more than 95% of all elements in $\mathbb{P}^1\left(\mathbb{Z}/N\mathbb{Z}\right)$. □

8.4 Computing all modular symbol with a given small denominator

Recall that δ^2 is the largest square dividing N.

Theorem 12. *Assume Goldfeld's period conjecture and assume Manin's conjecture that $c_0 = 1$. Let $m > 1$ be an integer. Then there is a method to evaluate all modular symbols*

$$\left\{ \left[\tfrac{a}{m}\right]^{\pm} \;\middle|\; 0 \leqslant a < m \text{ and } \gcd(a,m) = 1 \right\}$$

for any elliptic curve over $\mathbb{Q}$ of conductor N with $\gcd(m,\delta) = 1$ in less than $\tilde{\mathbf{O}}(N^{1/2})$ steps.

If we restrict to semistable curves, the condition on c_0 can be dropped and m is always coprime to $\delta = 1$.

Proof. By assumption all cusps a/m are unitary. Recall from the explanations in Section 7.1 the formula (14). Hence we start by evaluating all $\{\kappa_{j,m}(y)\}_j$ with $y = m\sqrt{Q}$ using the approximation in Lemma 5. This can be done with m sums of $\mathbf{O}(m\sqrt{Q}/m)$ terms. Thus this first part takes $\mathbf{O}(m\sqrt{Q})$ steps.

Given the vector $\{\kappa_{j,m}(1/y)\}_j$, we need to obtain the vector

$$\Big\{\sum_{j=0}^{m-1} \kappa_{j,m}\big(\frac{1}{m\sqrt{Q}}\big)\zeta^{ja} \;\Big|\; a = 1, \ldots, m-1\Big\}$$

where $\zeta = \exp(2\pi i/m)$. For this we can use fast Fourier transform; in particular with Bluestein's multi-radix algorithm [3] this is done in $\mathbf{O}(m\log(m))$ steps even when m is not a prime power. Hence we get a complexity of $\tilde{\mathbf{O}}(m\sqrt{N})$ as $Q \leqslant N$, which yields the result as m is fixed. □

In practice, we may be interested in computing approximations to the p-adic L-function for varying elliptic curves. Let p^r be a fixed prime power. In order to determine the r-th approximation to the p-adic L-function as explained in [23], we will only need to compute all modular symbols with denominator p^r. By the above this can be done with a complexity $\tilde{\mathbf{O}}(\sqrt{N})$.

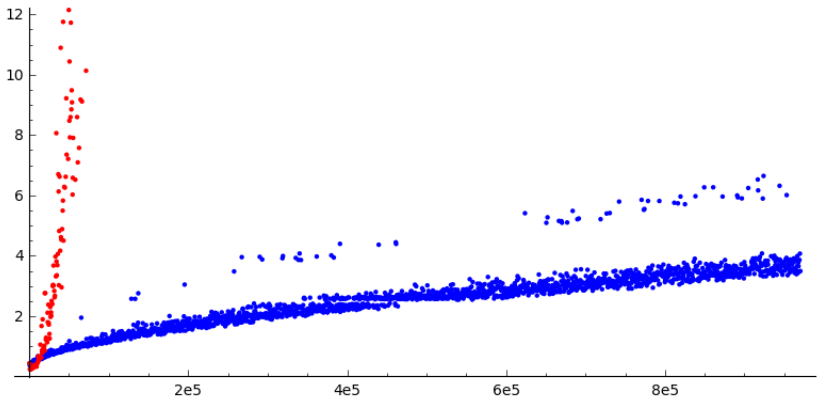


Figure 3: Comparison of approximations computing 5-adic L-functions for semistable elliptic curves

In Figure 3, we plot the time to compute the fourth approximation P_4 in the notation of [23]. We tested random semistable curves with good ordinary reduction at 5 of conductor up to 10^6, either from Cremona's table or from table of Stein and Watkins. The steeply increasing set of values uses `eclib`, the other timings are obtained with our implementation. The graph shows two anomalies: First there are a small number of values significantly higher than others. It turns out these are those examples for which the standard double precision of 53 bits is not sufficient and the implementation has to use the much slower library of arbitrary precision floating point numbers. Secondly, there is a strange vertical strip empty. This is due to the choices of the values of $B(\varsigma)$ in (5); these particular computations involve about 277200 terms in the sum.

8.5 Computing all Manin symbols

We wish to compare the numerical modular symbols to current implementations. Traditional methods start by finding a basis for the space of modular symbols attached to E in the space of all modular symbols for $\Gamma_0(N)$. This is equivalent to computing all Manin symbols $M(c : d)$ for $(c : d) \in \mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$. We will estimate therefore the complexity to compute all Manin symbols via numerical approximations. Note however that in practice, we never do this. Instead we fill up the cached values for Manin symbols as we go along.

Theorem 13. *Assume Goldfeld's period conjecture is true. Then there is a method to evaluate all Manin symbols for any semistable elliptic curve over $\mathbb{Q}$ of conductor N in less than $\tilde{\mathbf{O}}(N^{7/4})$ steps.*

Proof. As in Section 6.3 we denote for each $P \in \mathbb{P}^1\big(\mathbb{Z}/N\mathbb{Z}\big)$ the lattice Λ_P whose points with coprime coordinates are the possible representations of P. Let $\vec{v}_P$ be the shortest non-zero vector in Λ_P.

We start by evaluating all $M(P)$ for those $P \in \mathbb{P}^1\left(\mathbb{Z}/N\mathbb{Z}\right)$ with $\|\vec{v}_P\| \leqslant 2N^{3/8}$. There are at most $4\pi N^{3/4}$ of them and each such Manin symbols can be evaluated in $\tilde{\mathbf{O}}(N)$ steps using the direct method. Hence all of them are done in $\tilde{\mathbf{O}}(N^{7/4})$ steps.

Now, we may assume that $\|\vec{v}_P\| > 2N^{3/8}$. By Minkowski's convex body theorem, we also know that $\|\vec{v}_P\| \leqslant 2/\sqrt{\pi}\, N^{1/2}$. We apply Lemma 9 and find that P can be written as $(c:d)$ with

$$\max(|c|,|d|) \leqslant \frac{N}{\|\vec{v}_P\|} + C' \log(N)^2 \, \|\vec{v}_P\| < \tfrac{1}{2}\, N^{5/8} + C' \log(N)^2 2/\sqrt{\pi}\, N^{1/2} = \mathbf{O}(N^{5/8})$$

if N is sufficiently large. Therefore, we can evaluate all the remaining Manin symbols using the indirect method if we can compute all $\lambda(\frac{a}{m})$ with $m < N^{5/8}$ and $0 < a < m$. To do this, we use the idea in the previous section and we can get all $\lambda(\frac{a}{m})$ for a fixed m in $\tilde{\mathbf{O}}(m\sqrt{N})$ steps. Hence to find all $\lambda(\frac{a}{m})$ for $m < N^{5/8}$ we require $\tilde{\mathbf{O}}\big((N^{5/8})^2\sqrt{N}\big) = \tilde{\mathbf{O}}(N^{7/4})$ steps. □

Again, we can comment that this complexity is not always optimal. If N is prime, then all Manin symbols can be computed using all $\lambda(\frac{a}{m})$ with $m < \mathbf{O}(\sqrt{N})$. This gives a total complexity of $\tilde{\mathbf{O}}(N^{3/2})$.

To get unconditional results, i.e., independent of Goldfeld's conjecture, one may multiply all the complexities above with N.

The current implementations involve Gaussian elimination on sparse matrices of size $\mathbf{O}(N) \times \mathbf{O}(N)$. More precisely, as explained in Algorithm 8.38 in [21], each matrix has about $N/3$ rows each containing at most three non-zero values. It is not hard to see that Gaussian elimination needs at least $\mathbf{O}(N^{3/2})$ steps for each such matrix as we expect to reach a dense matrix by the time we are dealing with the last $\sqrt{N}$ rows. However it would be rather hard to prove a precise complexity for the full algorithm.

9 Examples

The computations below are performed with our implementation [28] written in `Cython` [2]. Note that this implementation is not fully optimised. The emphasis was on getting correct results for unitary cusps and for computing all modular symbols for a given denominator. For instance, it does not include the algorithm with the complexity of Theorem 13, though for the range of considered conductors this will not matter much.

First, we present a concrete example of our methods. We choose the curve $E = 234446\text{a}1$, famous for being the first curve in Cremona's tables of rank 4. It is semistable so we do not have to worry about non-unitary cusps. We are interested in computing the p-adic L-function $\mathscr{L}_p(E,T)$ as explained in [23] at the good ordinary prime $p = 5$. There are no isogenies from E defined over $\mathbb{Q}$ and the Néron period lattice $\Lambda_E = 1.486336\ldots\mathbb{Z} \oplus 0.800625\ldots\mathbb{Z}i$ is rectangular. Therefore the modular symbols $[r]^{\pm}$ are integers. In fact $[\frac{1}{27}]^+ = [\frac{1}{7}]^- = 1$ and $[\frac{1}{7}]^+ = 0$ show that the values $\lambda(r)$ generate Λ_E. In particular, we have to approximate the real part of $\lambda(r)$ to precision 0.743168. When computing all values of $[\frac{a}{5}]^+$ using the partial sums $\kappa_{j,5}(y)$ we need $T = 2923$ terms and the precision of 53 bits is enough. The largest error in evaluating these was smaller than $0.00032\,\Omega^+$. Similar for all values $[\frac{a}{25}]^+$ we only need to sum 17716 terms, still with precision of 53 bits. Using these values one finds that the fourth coefficient of $\mathscr{L}_5(E,T)$ is congruent to 1 modulo 5. This implies that the rank of $E(\mathbb{Q})$ is at most 4. Together with the explicit basis of $E(\mathbb{Q})$ one can deduce without much further effort that the 5-primary part of the Tate-Shafarevich group $Ш(E/\mathbb{Q})$ is trivial.

Next, in comparison an example involving non-unitary cusps. Let E be the elliptic curve 1017a1, which has additive reduction at 3 of Kodaira type III. Its quadratic twist by -3 is 1017e1, which has type III* at 3. The seemingly harmless computation of $[\frac{1}{3}]^+$ now involves more than 48000 terms to sum in total. Instead $[\frac{1}{5}]^+$ only requires 217 terms to sum. Though we have to admit that it is likely that the implementation for the non-unitary cases could be improved.

Now to the asymptotic behaviour as N increases. In Figure 4, we used the numerical implementation to compute all $[\frac{a}{25}]^+$ for various random semistable curves. The time in seconds is plotted against the conductor N. The quicker ones are those with conductor divisible by 5. This and the following computations were performed on rather standard hardware, for instance on a Intel Xeon E5-2660 2.6 GHz virtual machine with two cores.

We now pass to compare the various implementations. There is our implementation [28] of numerical modular symbols written in `Cython` [2] incorporated into `SageMath` [9], the implementation of `eclib` [8], written in `C`, also accessible within `SageMath`, the pure `Python` implementation in `SageMath`,

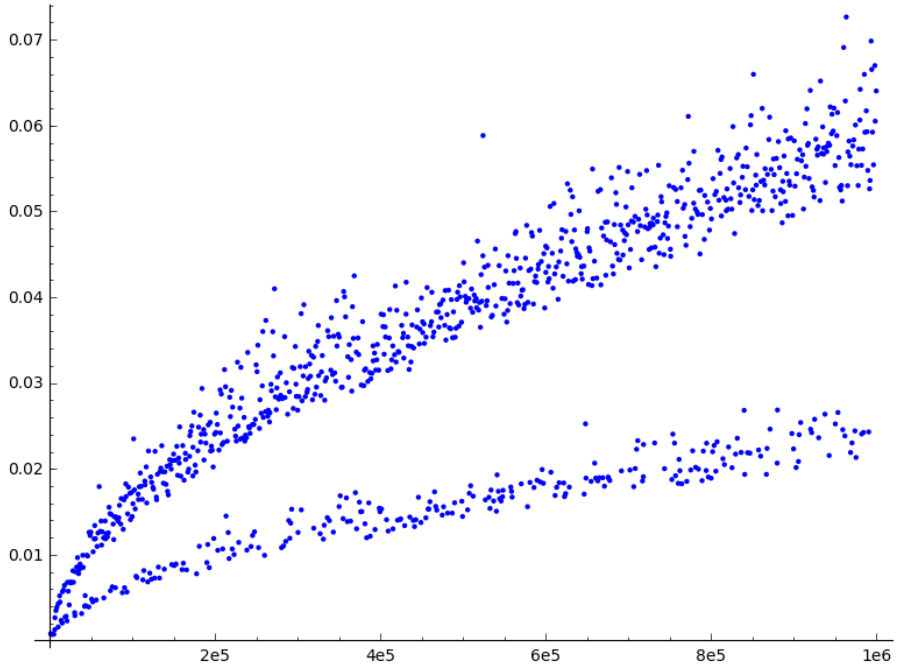


Figure 4: Time to compute all symbols $[\frac{a}{25}]^+$ for some semistable curves.

the implementation in `Magma` [4] and the implementation in `PARI` [19]. First we will exclude the pure `Python` implementation in `SageMath` and the one in `PARI`, which is still under development, as they are both significantly slower then the other three. The fact that these four implementations of the same algorithm have such different timings explains why we cannot compare them directly: they are written in different languages. Also, we call them from within `SageMath` and the time `SageMath` spends to call the underlying code varies much. Instead we want to illustrate the asymptotic behaviour of the computation.

In Figure 5 we plot the time to compute all Manin symbols $M(c : d)$ using the numerical implementation (•) against the determination of the space of modular symbols by `Magma` (■) and `eclib` (+). We do this in all three cases for random semistable curves of conductor up to 55000. The computation was stopped after 30 seconds, meaning that for some curves the plotted point would lie an unknown amount above the visible part. The computations in `Magma` became rather quickly too complicated and they were stopped after conductor 25937.

References

[1] Amod Agashe, Kenneth Ribet, and William A. Stein, *The Manin constant*, Pure Appl. Math. Q. **2** (2006), no. 2, part 2, 617–636.

[2] Stefan Behnel, Robert Bradshaw, Craig Citro, Lisandro Dalcin, Dag Sverre Seljebotn, and Kurt Smith, *Cython: The Best of Both Worlds*, Computing in Science Engineering **13** (2011), no. 2, 31–39, `http://cython.org/`.

[3] Leo I. Bluestein, *A linear filtering approach to the computation of the discrete Fourier transform*, IEEE Northeast Electronics Research and Engineering Meeting **10** (1968), 218–219.

[4] Wieb Bosma, John Cannon, Claus Fieker, and Allan Steel, *Handbook of Magma function*, 2.19-6 ed., 2013.

[5] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), no. 4, 843–939.

[6] Henri Cohen, *A course in computational algebraic number theory*, Graduate Texts in Mathematics, vol. 138, Springer-Verlag, Berlin, 1993.

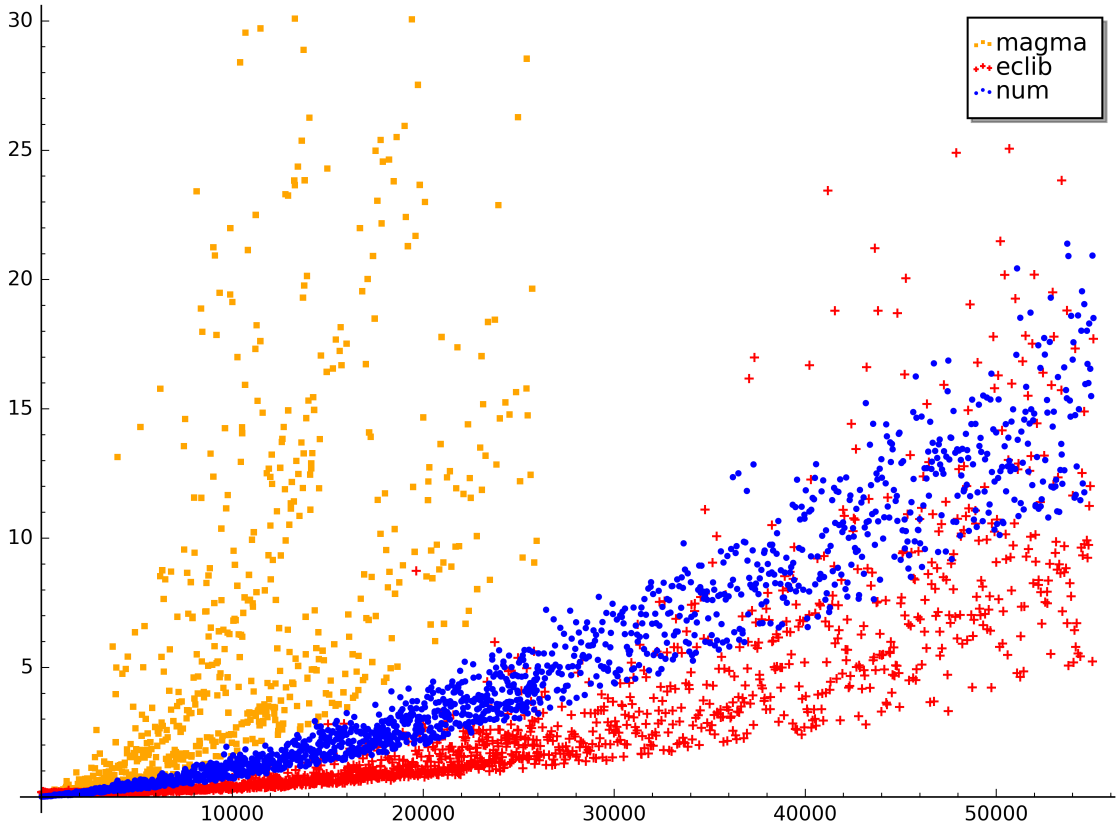


Figure 5: Time to compute all Manin symbols for some semistable curves

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

[8] ———, *The eclib package, version 20150827*, available at `https://github.com/JohnCremona/eclib`, 2015.

[9] The Sage Developers, *Sagemath, the Sage Mathematics Software System (Version 7.2)*, 2016, available from `http://www.sagemath.org`.

[10] Vladimir G. Drinfel′d, *Two theorems on modular curves*, Funkcional. Anal. i Priložen. **7** (1973), no. 2, 83–84.

[11] Bas Edixhoven, *On the Manin constants of modular elliptic curves*, Arithmetic algebraic geometry (Texel, 1989), Progr. Math., vol. 89, Birkhäuser Boston, Boston, MA, 1991, pp. 25–39.

[12] Dorian Goldfeld, *On the computational complexity of modular symbols*, Math. Comp. **58** (1992), no. 198, 807–814.

[13] Grigor Grigorov, Andrei Jorza, Stefan Patrikis, William A. Stein, and Corina Tarniţă, *Computational verification of the Birch and Swinnerton-Dyer conjecture for individual elliptic curves*, Math. Comp. **78** (2009), no. 268, 2397–2425.

[14] Nicholas J. Higham, *Accuracy and stability of numerical algorithms*, second ed., Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[15] Henryk Iwaniec, *On the error term in the linear sieve*, Acta Arith. **19** (1971), 1–30.

[16] Juri I. Manin, *Parabolic points and zeta functions of modular curves*, Izv. Akad. Nauk SSSR Ser. Mat. **36** (1972), 19–66.

[17] Barry Mazur and Peter Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1–61.

[18] Barry Mazur, John Tate, and Jeremy Teitelbaum, *On p-adic analogues of the conjectures of Birch and Swinnerton-Dyer*, Invent. Math. **84** (1986), no. 1, 1–48.

[19] The PARI Group, Bordeaux, *PARI/GP, version 2.8.0*, 2016, available from `http://pari.math.u-bordeaux.fr/`.

[20] Joseph H. Silverman, *The arithmetic of elliptic curves*, second ed., Graduate Texts in Mathematics, vol. 106, Springer, Dordrecht, 2009.

[21] William A. Stein, *Modular forms, a computational approach*, Graduate Studies in Mathematics, vol. 79, American Mathematical Society, Providence, RI, 2007, With an appendix by Paul E. Gunnells.

[22] William A. Stein and Helena A. Verrill, *Cuspidal modular symbols are transportable*, LMS J. Comput. Math. **4** (2001), 170–181.

[23] William A. Stein and Christian Wuthrich, *Algorithms for the arithmetic of elliptic curves using Iwasawa theory*, Math. Comp. **82** (2013), no. 283, 1757–1792.

[24] Glenn Stevens, *Arithmetic on modular curves*, Progress in Mathematics, vol. 20, Birkhäuser Boston Inc., Boston, MA, 1982.

[25] ———, *Stickelberger elements and modular parametrizations of elliptic curves*, Invent. Math. **98** (1989), no. 1, 75–106.

[26] Joseph L. Wetherell et al., *The pari script modsym.gp*, available at `http://pari.math.u-bordeaux.fr/Scripts/modsym.gp`., 2002.

[27] Christian Wuthrich, *On the integrality of modular symbols and Kato's Euler system for elliptic curves*, Doc. Math. **19** (2014), 381–402.

[28] ———, *Sage trac ticket # 21046: Numerical modular symbols for elliptic curves*, `https://trac.sagemath.org/ticket/21046`, 2016.