Rational points on modular curves

Jan Vonk

Contents

Chapter **1**

Introduction

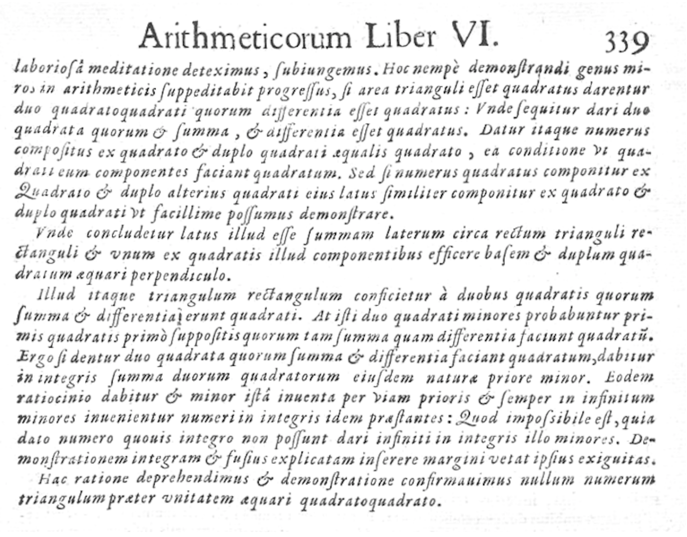
These are the notes for a series of lectures given in September 2023 at the ICTS Bangaluru. They contain an introduction to some of the ideas that have enabled important progress towards understanding Mordell–Weil groups of elliptic curves and the Birch–Swinnerton-Dyer conjecture. Particular emphasis will be on Mazur's work on torsion, and the work of Gross–Zagier on heights of Heegner points.

1.1 The historical method of descent

The group of rational points on elliptic curves is fruitfully studied using the method of *descent.* Its origins occur in the work of Fermat, who applies it to a variety of problems in his 1659 letter to Pierre de Carcavi [dF59]. Perhaps his most famous application is the only complete proof of the hand of Fermat that survives, showing that congruent numbers (areas of right angled triangles with rational side lengths) are never squares [Fer70]. Fermat reduces this problem to showing there are no non-trivial integer solutions to

$$x^4 - y^4 = z^2. \tag{1.1}$$

To solve (1.1) Fermat uses a remarkable method, which we recognise in contemporary language as a descent by 2-isogeny on an elliptic curve. Fermat sounds rather pleased with his argument, stating that "This type of demonstration will provide excellent progress in arithmetic." The proof appears in [Fer70]:



Arithmeticorum Liber VI. 339

laborioſâ meditatione deteximus, ſubiungemus. Hoc nempè demonſtrandi genus miros in arithmeticis ſuppeditabit progreſſus, ſi area trianguli eſſet quadratus darentur duo quadratoquadrati quorum differentia eſſet quadratus: Vnde ſequitur dari duo quadrata quorum & ſumma, & differentia eſſet quadratus. Datur itaque numerus compoſitus ex quadrato & duplo quadrati æqualis quadrato, ea conditione vt quadrati eum componentes faciant quadratum. Sed ſi numerus quadratus componitur ex Quadrato & duplo alterius quadrati eius latus ſimiliter componitur ex quadrato & duplo quadrati vt facillime poſſumus demonſtrare.

Vnde concludetur latus illud eſſe ſummam laterum circa rectum trianguli rectanguli & vnum ex quadratis illud componentibus efficere baſem & duplum quadratum æquari perpendiculo.

Illud itaque triangulum rectangulum conficietur à duobus quadratis quorum ſumma & differentia erunt quadrati. At iſti duo quadrati minores probabuntur primis quadratis primò ſuppoſitis quorum tam ſumma quam differentia faciunt quadratū. Ergo ſi dentur duo quadrata quorum ſumma & differentia faciant quadratum, dabitur in integris ſumma duorum quadratorum eiuſdem naturæ priore minor. Eodem ratiocinio dabitur & minor iſtâ inuenta per viam prioris & ſemper in infinitum minores inuenientur numeri in integris idem præſtantes: Quod impoſſibile eſt, quia dato numero quouis integro non poſſunt dari infiniti in integris illo minores. Demonſtrationem integram & fuſius explicatam inſerere margini vetat ipſius exiguitas.

Hac ratione deprehendimus & demonſtratione confirmauimus nullum numerum triangulum præter vnitatem æquari quadratoquadrato.

In modern language, we may describe the proof of Fermat as follows. Suppose we have a non-trivial solution (x, y, z) to the quartic equation (1.1), then we may assume x, y and z are coprime positive integers. From this coprime solution, Fermat constructs a new (smaller) solution, in two steps.

Step 1. We factorise the equation (1.1) as follows:

$$z^2 = x^4 - y^4 = (x^2 - y^2)(x^2 + y^2).$$

Note that the factors on the right hand side are coprime to each other. This implies that they must both be squares, i.e. there are positive integers s, t such that

$$\begin{cases} x^2 - y^2 &= s^2 \\ x^2 + y^2 &= t^2. \end{cases}$$

Observe that s and t must both be odd integers, and by changing the sign of s if necessary we may assume that $s - t \equiv 0 \pmod 4$. We then note that y must be even, and that we therefore have the following decomposition into integer factors:

$$\left(\frac{t+s}{2}\right)\left(\frac{t-s}{4}\right) = \left(\frac{y}{2}\right)^2.$$

The factors on the left hand side are coprime positive integers. We may once again conclude they are both perfect squares, so that we find odd coprime positive integers u, v that satisfy the equalities

$$\begin{cases} s &= u^2 - 2v^2 \\ t &= u^2 + 2v^2. \end{cases} \tag{1.2}$$

Note that we have now produced a triple (u, v, x) that satisfies $u^4 + 4v^4 = x^2$. Moreover, the triple (x, y, z) may be recovered from the triple (u, v, x) by the identities $y = 2uv$ and $z = u^4 - 4v^4$.

Step 2. Note that the relation

$$u^4 + 4v^4 = x^2$$

satisfied by the triple (u, v, x) constructed in step 1 implies in particular that $(u^2, 2v^2, x)$ is a Pythagorean triple. As such, we may find coprime positive integers m, n satisfying

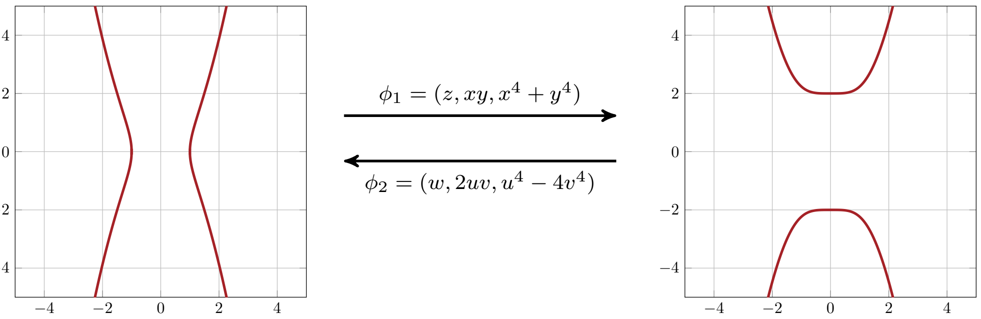
$$\begin{cases} u^2 &= m^2 - n^2 \\ 2v^2 &= 2mn \\ x &= m^2 + n^2 \end{cases} \tag{1.3}$$

Since $v^2 = mn$ we see that m and n are both squares. Writing $m = a^2, n = b^2$ with $a, b > 0$ we find that the triple (a, b, u) is a solution to (1.1), i.e. $a^4 - b^4 = u^2$. We see that $a < a^4 + b^4 = x$ so that we constructed a new solution whose first coordinate is strictly smaller than that of the original solution. This shows that if a non-trivial solution exists, we can keep descending ad infinitum, which is absurd.

If we unpack Fermat's argument a little further, we see that it considers two genus 1 curves defined by homogeneous equations in the weighted projective plane, namely

$$\begin{aligned} E_1 \quad &: \quad \left\{ (x, y, z) \in \mathbf{P}^2_{[1,1,2]} \quad : \quad x^4 - y^4 = z^2 \right\} \\ E_2 \quad &: \quad \left\{ (u, v, w) \in \mathbf{P}^2_{[1,1,2]} \quad : \quad u^4 + 4v^4 = w^2 \right\} \end{aligned} \tag{1.4}$$

They define elliptic curves after the choice of base points $(1, 0, 1) \in E_1(\mathbf{Q})$ and $(1, 0, 1) \in E_2(\mathbf{Q})$. These elliptic curves admit a pair (ϕ_1, ϕ_2) of dual rational 2-isogenies, described by:



The argument of Fermat produces for any purported non-trivial point $(x, y, z) \in E_1(\mathbf{Q})$ a preimage (a, b, u) for the multiplication by 2 map $[2] = \phi_2 \circ \phi_1$. The procedure consists of two steps, and first constructs a preimage for ϕ_2, then a preimage for ϕ_1. When taken to its natural conclusion, Fermat therefore really shows two things: First, his arguments suffice to show that

$$E_1(\mathbf{Q})/\phi_2(E_2(\mathbf{Q})) \;\simeq\; \mathbf{Z}/2\mathbf{Z} \;\simeq\; E_2(\mathbf{Q})/\phi_1(E_1(\mathbf{Q})),$$

and second, his descent argument on the ever shrinking first coordinate of a solution may be used to deduce that any solution must have one of its coordinates equal to zero. From this, one concludes that

$$\begin{aligned} E_1(\mathbf{Q}) &= \{(1,0,1),(1,0,-1),(1,1,0),(1,-1,0)\} &&\simeq \mathbf{Z}/4\mathbf{Z} \\ E_2(\mathbf{Q}) &= \{(1,0,1),(1,0,-1),(0,1,2),(0,1,-2)\} &&\simeq \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}\,. \end{aligned}$$

1.2 The Mordell–Weil theorem

Nearly three centuries later, Mordell showed in his landmark paper [Mor22] how an argument of this sort can be carried out for general elliptic curves E over $\mathbf{Q}$. Mordell shows how to establish finiteness of the group $E(\mathbf{Q})/2E(\mathbf{Q})$ and deduces by descent that the group $E(\mathbf{Q})$ is finitely generated. The work of Weil [Wei29] observes that the notion of "size" may be formalised by the theory of heights, extending the argument to general number fields, and abelian varieties. This yields the following result.

Theorem 1 (Mordell–Weil)**.** *Let A be an abelian variety defined over a number field K. The Mordell–Weil group $A(K)$ is finitely generated, i.e. there exist a finite subgroup $A(K)_{\text{tors}} \subset A(K)$ and $r \geq 0$ such that*

$$A(K) \;\simeq\; A(K)_{\text{tors}} \times \mathbf{Z}^r\,.$$

The central task is to show the *weak* Mordell–Weil theorem, which is the statement that $A(K)/nA(K)$ is finite, for some $n \geq 1$. The short exact sequence of G_K-modules defined by the multiplication by n map

$$0 \longrightarrow A[n] \longrightarrow A \xrightarrow{\cdot n} A \longrightarrow 0 \tag{1.5}$$

defines a long exact sequence in Galois cohomology, from which we extract the exact sequence

$$0 \longrightarrow A(K)/nA(K) \longrightarrow \mathrm{H}^1(K, A[n]) \longrightarrow \mathrm{H}^1(K, A)[n] \longrightarrow 0. \tag{1.6}$$

Since the cohomology group $\mathrm{H}^1(K, A[n])$ is infinite, the sequence (1.6) does not yet prove the finiteness of the weak Mordell–Weil group. To this end, we consider also the local variants of (1.6), and obtain the following commutative diagram with exact rows, where the products are taken over all places v of K, and the vertical maps are the product of the corresponding localisation maps.

$$\begin{tikzcd}
0 \arrow[r] & A(K)/nA(K) \arrow[r] \arrow[d] & \mathrm{H}^1(K, A[n]) \arrow[r] \arrow[d] \arrow[dr, dashed, "\varphi"] & \mathrm{H}^1(K, A)[n] \arrow[r] \arrow[d] & 0 \\
0 \arrow[r] & \prod_v A(K_v)/nA(K_v) \arrow[r] & \prod_v \mathrm{H}^1(K_v, A[n]) \arrow[r] & \prod_v \mathrm{H}^1(K_v, A)[n] \arrow[r] & 0
\end{tikzcd}$$

By the commutativity of this diagram and the exactness of the rows, we see that the image of the weak Mordell–Weil group must lie in the kernel of the map φ. This observation allows us to refine the sequence (1.6), by defining the *Selmer group* $\mathrm{Sel}_n(A)$ and the *Tate–Shafarevich group* $\text{Ш}(A)$ as the kernels of the products of all the natural localisation maps

$$\begin{array}{rcllcl}
\mathrm{Sel}_n(A/K) & := & \mathrm{Ker}\big(\mathrm{H}^1(K, A[n]) & \longrightarrow & \prod_v \mathrm{H}^1(K_v, A)\big) \\
\text{Ш}(A/K) & := & \mathrm{Ker}\big(\mathrm{H}^1(K, A) & \longrightarrow & \prod_v \mathrm{H}^1(K_v, A)\big)
\end{array}$$

so that we obtain a short exact sequence

$$0 \longrightarrow A(K)/nA(K) \longrightarrow \mathrm{Sel}_n(A/K) \longrightarrow \text{Ш}(A/K)[n] \longrightarrow 0. \tag{1.7}$$

The finiteness of the weak Mordell–Weil group follows from the finiteness of the Selmer group. For this latter fact, many excellent resources exist, see for instance Silverman [Sil09, Chapter VIII]. We assume the reader to be acquainted with these proofs, and sketch it using the algebro-geometric language of Milne [Mil06, Chapter IV.3], which will prepare us for the arguments in the work of Mazur in § 2.

Theorem 2. *Let A be an abelian variety defined over a number field K. For any integer $n \geq 1$, the weak Mordell–Weil group $A(K)/nA(K)$ is finite.*

Proof. Let $U = \operatorname{Spec} \mathcal{O}_K[1/S]$ where S is the set consisting of all the primes of bad reduction of A, and the primes dividing n. Then A extends to an abelian variety $\mathscr{A}$ over U, and we have an exact sequence of sheaves on the étale site of U defined by

$$0 \longrightarrow \mathscr{A}[n] \longrightarrow \mathscr{A} \xrightarrow{\cdot n} \mathscr{A} \longrightarrow 0$$

and since $\mathscr{A}(U) = A(K)$ we extract from the long sequence in cohomology an injection

$$0 \longrightarrow A(K)/nA(K) \longrightarrow \mathrm{H}^1_{\text{ét}}(U, \mathscr{A}[n])$$

and therefore it suffices to show that the target cohomology group is finite. To show this, we pick a finite étale covering $V \to U$ over which this group is more easily understood. The Hochschild–Serre spectral sequence [Mil80, III.2.20] provides the exact sequence

$$1 \longrightarrow \mathrm{H}^1(G, \mathscr{A}[n](V)) \longrightarrow \mathrm{H}^1_{\text{ét}}(U, \mathscr{A}[n]) \longrightarrow \mathrm{H}^1_{\text{ét}}(V, \mathscr{A}[n])^G$$

where $G = \mathrm{Aut}(V/U)$ is the Galois group. The left term is clearly finite, so it suffices to show that $\mathrm{H}^1_{\text{ét}}(V, \mathscr{A}[n])$ is finite. Choose $V = \operatorname{Spec} \mathcal{O}_L[1/S]$, where L/K is a finite Galois extension satisfying

- the n-torsion is constant over V, i.e. $\mathscr{A}[n] \simeq (\mathbf{Z}/n\mathbf{Z})^g$,
- the integers $\mathcal{O}_L$ contain the n-th roots of unity, so that $\mu_n \simeq \mathbf{Z}/n\mathbf{Z}$ over V.

This may always be achieved, potentially at the cost of slightly enlarging the finite set S. We are therefore reduced to showing finiteness of $\mathrm{H}^1_{\text{ét}}(V, \mathbf{Z}/n\mathbf{Z})$. Since n is invertible on V, we know the exactness of the n-th Kummer sequence of étale sheaves on V, given by

$$1 \longrightarrow \mu_n \simeq \mathbf{Z}/n\mathbf{Z} \longrightarrow \mathbf{G}_m \xrightarrow{(\cdot)^n} \mathbf{G}_m \longrightarrow 1.$$

From the associated long exact sequence in cohomology, we extract

$$1 \longrightarrow \mathcal{O}_L[1/S]^\times/(\mathcal{O}_L[1/S]^\times)^n \longrightarrow \mathrm{H}^1_{\text{ét}}(V, \mathbf{Z}/n\mathbf{Z}) \longrightarrow \mathrm{Pic}(V)[n] \longrightarrow 1.$$

Note that $\mathrm{H}^1_{\text{ét}}(V, \mathbf{Z}/n\mathbf{Z})$ is flanked by two terms that are finite: finiteness of the left follows from Dirichlet's unit theorem, whereas for the right it follows from the finiteness of class groups. □

We showed that when U is a Zariski open of $\operatorname{Spec} \mathcal{O}_K$ where A has good reduction and n is invertible, then $\mathrm{H}^1_{\text{ét}}(U, \mathscr{A}[n])$ is necessarily finite. This group consists of classes of the Galois cohomology group

$$\mathrm{H}^1(K, A[n]) \simeq \mathrm{H}^1_{\text{ét}}(\mathrm{Spec}(K), A[n])$$

that are unramified outside the set S containing the primes of bad reduction of A, and the prime divisors of n. In particular, it contains the Selmer group $\mathrm{Sel}_n(A/K)$, which satisfies additional solubility conditions at the bad places in S. The proof of the weak Mordell–Weil theorem may sometimes be turned into an algorithm for determining the weak Mordell–Weil group, or rather, the Selmer group.

1.3 The Birch–Swinnerton-Dyer conjecture

The numerical experiments of Birch and Swinnerton-Dyer [BSD65] revealed a striking connection between the *algebraic* invariants associated to an elliptic curve, and its *analytic* L-function. An extension of this conjecture to abelian varieties over number fields was formulated by Tate [Tat66], who also proved its isogeny invariance. This extended an earlier result for elliptic curves due to Cassels [Cas65].

Suppose A/K is an abelian variety of dimension g defined over a number field K. Let $\mathrm{L}(A, s)$ be the L-function associated to its ℓ-adic Tate module $V_\ell A$ for some ℓ; in other words

$$\mathrm{L}(A, s) = \prod_{\mathfrak{p}} \frac{1}{P_\mathfrak{p}(\mathrm{Nm}(\mathfrak{p})^{-s})}, \qquad \text{where } \; P_\mathfrak{p}(T) = \det\left(1 - \mathrm{Fr}_\mathfrak{p} T \mid V_l(A)^{I_\mathfrak{p}}\right).$$

Here, the product is over all primes $\mathfrak{p}$ and we used the notation $I_\mathfrak{p}$ for the inertia group of some chosen decomposition group above $\mathfrak{p}$ in G_K, and $\mathrm{Fr}_\mathfrak{p}$ its Frobenius element. The polynomial $P_\mathfrak{p}(T)$, which is the inverse characteristic polynomial of $\mathrm{Fr}_\mathfrak{p}$ on the inertia fixed part, is in $\mathbf{Z}[T]$ and independent of ℓ.

The function $\mathrm{L}(A, s)$ depends only on the isogeny class of A, and the Hasse–Weil bound on the coefficients of its Dirichlet series shows that it converges for $\mathrm{Re}(s) > 3/2$. It is conjectured to have an analytic continuation to all $s \in \mathbf{C}$. More precisely, define the completed L-function

$$\mathrm{L}^*(A, s) := \mathrm{L}(A, s) \cdot |\Delta_K|^{gs} \cdot N^{gs/2} \cdot \left(\frac{\Gamma(s)}{(2\pi)^s}\right)^{g\cdot[K:\mathbf{Q}]}$$

where Δ_K be the discriminant of K, and $N = \mathrm{Nm}_{K/\mathbf{Q}}(\mathcal{N})$ is the norm of the conductor of A. Then the completed L-function is conjectured to satisfy a functional equation

$$\mathrm{L}^*(A, 2 - s) = w(A/K)\, \mathrm{L}^*(A, s), \qquad w(A/K) = \pm 1.$$

The BSD conjecture predicts that, assuming $\mathrm{L}(A,s)$ admits analytic continuation as conjectured, the order of vanishing at $s=1$ is equal to the rank of $A(K)$, with the leading term encoding various algebraic invariants of A. These invariants are defined in terms of the Néron model $\mathscr{A}$ as follows:

- The Tamagawa numbers for every place v of K are defined by setting k_v the residue class field of K_v (when v is archimedean, we set $k_v = K_v$) and setting
$$c_v = |\,\mathscr{A}(k_v)/\,\mathscr{A}(k_v)^0|$$
where $\mathscr{A}(k_v)^0$ is the connected component of the identity.
- The real period Ω is the (suitably normalised) absolute value of the product of the integrals of $\omega_1 \wedge \ldots \wedge \omega_g$ over the connected components of the identity of $A(K_v)$, for every real place v of K. Here, the ω_i are a basis of Néron differentials for $\mathscr{A}$ over $\mathcal{O}_K$. In general, the precise definition is somewhat subtle, but for elliptic curves $E/\,\mathbf{Q}$ it is simply the integral of the Néron differential over $E(\mathbf{R})^0$.
- The regulator $\mathrm{Reg}(A/K)$ is the volume of the canonical Néron–Tate height pairing
$$\widehat{h} : A(K) \times A^\vee(K) \longrightarrow \mathbf{R}\,.$$
More precisely, note that A and its dual abelian variety $A^\vee$ are isogenous over K, so their Mordell–Weil groups have the same rank. Whenever $\{P_1,\ldots,P_r\}$ and $\{Q_1,\ldots,Q_r\}$ are bases for the free parts of the Mordell–Weil groups $A(K)$ and $A^\vee(K)$ respectively, we define
$$\mathrm{Reg}(A/K) := \left|\det\left(\widehat{h}(P_i,Q_j)\right)\right|$$

Assuming that $\text{Ш}(A/K)$ is finite, the following conjecture [BSD65, Tat66] is widely believed.

Conjecture 1 (Birch–Swinnerton-Dyer). *Let A/K be an abelian variety of rank $r = \mathrm{rk}_{\mathbf{Z}} A(K)$. Then $\mathrm{L}(A,s)$ vanishes to order r at $s=1$, and its r-th derivative $\mathrm{L}^{(r)}$ with respect to s satisfies*
$$\frac{\mathrm{L}^{(r)}(A,1)}{r!} = \Omega \cdot \frac{\prod_v c_v \cdot \mathrm{Reg}(A/K) \cdot |\text{Ш}(A/K)|}{\sqrt{\Delta_K}^{g} \cdot |A(K)_{\mathrm{tors}}| \cdot |A^\vee(K)_{\mathrm{tors}}|}$$

This conjecture was discovered through systematic numerical computations for elliptic curves. In higher dimensional examples, serious obstacles need to be overcome to compute both sides of the purported equality. The articles [FLS$^+$01, vB21, vB22] achieve this for a large collection of interesting examples.

Modular abelian varieties over Q

The statement of the BSD conjecture assumes that the L-function has an analytic continuation. This is known to be true, for instance, when $K = \mathbf{Q}$ and A is an elliptic curve [Wil95, TW95, BCDT01]. More generally, when A is a modular abelian variety over $\mathbf{Q}$ with associated newform $f \in S_2(\Gamma_0(N))$ whose coefficients generate the number field K_f, we have
$$\mathrm{L}^*(A,s) = \prod_{\sigma\in\{K_f\hookrightarrow\mathbf{C}\}} \mathrm{L}^*(f^\sigma,s).$$

We now explain how to establish the analytic continuation and functional equation for these L-functions, as well as a practical way to evaluate them numerically in concrete examples. Each individual factor in the above product may be written as

$$\mathrm{L}(f,s) = \sum_{n\geq 1} \frac{a_n}{n^s}, \qquad \text{where } \ f(q) = \sum_{n\geq 1} a_n q^n, \ a_n \in K_f.$$

It is not difficult to see that this satisfies a functional equation, with sign (known as the *root number*) equal to the opposite of the eigenvalue $w_N = \pm 1$ of the associated modular form f for the Fricke involution

$$W_N := \begin{pmatrix} 0 & -\sqrt{1/N} \\ \sqrt{N} & 0 \end{pmatrix} \ : \ f(z) \longmapsto f(-1/Nz).$$

Indeed, the completed L-function appearing as factors are each a Mellin transform

$$\begin{aligned} \mathrm{L}^*(f,s) = \int_0^\infty f\left(\frac{it}{\sqrt{N}}\right) t^s \frac{dt}{t} \ &= \ \int_0^1 f\left(\frac{it}{\sqrt{N}}\right) t^s \frac{dt}{t} + \int_1^\infty f\left(\frac{it}{\sqrt{N}}\right) t^s \frac{dt}{t} \\ &= \ \int_1^\infty f\left(\frac{it}{\sqrt{N}}\right) \left(t^{s-1} - w_N t^{1-s}\right) dt \end{aligned} \tag{1.8}$$

Note that this integral representation achieves several things at once. First, it shows that the L-function may be analytically continued to all $s \in \mathbf{C}$, and that it satisfies the functional equation

$$\mathrm{L}^*(A, 2-s) = (-w_N)^g \, \mathrm{L}^*(A,s).$$

Secondly, evaluating the integral against an explicit q-expansion gives us a numerical way to evaluate the values of any of the derivatives $\mathrm{L}^{(r)}(A,1)$ with exponentially fast convergence, via

$$\left(\frac{d^r \, \mathrm{L}^*(A,s)}{ds^r}\right)\bigg|_{s=1} = \sum_{n\geq 1} a_n G_r\left(\frac{2\pi n}{\sqrt{N}}\right), \qquad \text{with} \qquad G_r(x) := \int_1^\infty e^{-xt} g_r(t) dt \tag{1.9}$$

where the function $g_r(t)$ is the r-th derivative at $s=1$ of the second factor in (1.8). In other words

$$g_r(t) := \left(\frac{d}{ds}\right)^r \left[t^{s-1} - w_N t^{1-s}\right]\Big|_{s=1}, \qquad \text{so that} \qquad \begin{cases} g_0(t) &= \ (1-w_N) \\ g_1(t) &= \ (1+w_N)\log(t) \\ & \ \ \ldots \end{cases}$$

Example 1

Let us begin with elliptic curves over $\mathbf{Q}$ of conductor 11, of which there are precisely three. Their minimal Weierstraß models over $\mathbf{Z}$ and torsion subgroups are summarised in the following table

$E_1 \ : \ y^2 + y \ = \ x^3 - x^2$	$E_1(\mathbf{Q})_{\mathrm{tors}} = \langle (0,0) \rangle \simeq \mathbf{Z}/5\mathbf{Z}$
$E_2 \ : \ y^2 + y \ = \ x^3 - x^2 - 10x - 20$	$E_2(\mathbf{Q})_{\mathrm{tors}} = \langle (0,0) \rangle \simeq \mathbf{Z}/5\mathbf{Z}$
$E_3 \ : \ y^2 + y \ = \ x^3 - x^2 - 7820x - 263580$	$E_3(\mathbf{Q})_{\mathrm{tors}} = 1$

These curves form a single isogeny class. Indeed, from the torsion points of order 5 we discovered, we quickly find that there are two pairs of dual isogenies connecting them

$$E_1 \overset{5}{\longleftrightarrow} E_2 \overset{5}{\longleftrightarrow} E_3$$

where the kernels of the isogenies going to the right are the constant group scheme $\mathbf{Z}/5\mathbf{Z}$ (generated by the 5-torsion points listed above), and the kernels of the isogenies going to the left are isomorphic to μ_5. We remark that in fact, these statements about the kernels remain true over $\operatorname{Spec}(\mathbf{Z})$ for the induced maps on Néron models; a somewhat miraculous fact that will reappear in § 2.

- **Real periods.** All three curves have precisely one real component, and the groups $E(\mathbf{R})$ are all isomorphic to S^1 as real Lie groups. The Néron differential on all three curves is $\omega = dx/(2y+1)$, so that we may compute the complex periods numerically as

  $$\Omega = \int_{E(\mathbf{R})^0} \frac{dx}{2y+1}$$

  Switching to a short Weierstraß equation, one easily calculates this integral numerically, and obtains the following approximate values for the periods:

  $$\Omega_1 \approx 6.34604652\ldots \qquad \Omega_2 \approx 1.26920930\ldots \qquad \Omega_3 \approx 0.25384186\ldots$$

  These look like integer multiples of each other, which is no coincidence: As previously remarked (see § 2) the isogenies $E_1 \to E_2$ and $E_2 \to E_3$ induce maps on Néron models whose kernels are isomorphic to the constant group scheme $\mathbf{Z}/5\mathbf{Z}$ over $\mathbf{Z}$. This implies that Néron differentials are preserved by pullback, and hence $\Omega_1 = 5\Omega_2$ and $\Omega_2 = 5\Omega_3$, as suggested by our numerical values.

- **Tamagawa numbers.** For any place v of $\mathbf{Q}$, we let $E(\mathbf{Q}_v)^0$ be the connected component of the identity of $E(\mathbf{Q}_v)$. One uses the general algorithm of Tate [Tat72] to compute the Tamagawa numbers

  $$c_v = |E(\mathbf{Q}_v)/E(\mathbf{Q}_v)^0|.$$

  In our example, all three curves have $c_\infty = 1$, and for all finite places v of good reduction $c_v = 1$. The only nontrivial Tamagawa numbers c_v occur at the finite prime $v = 11$, and it is easily determined in this case since all curves are semi-stable over $\mathbf{Z}_{11}$, and therefore $c_{11} = \operatorname{ord}_{11}(\Delta)$. For E_1 and E_3 we have $c_{11} = 1$, and the discriminant of E_2 is $\Delta = -11^5$ so that $c_{11} = 5$ in this case.

- **The regulator and Tate–Shafaverich group.** The most involved computation is to obtain the Mordell–Weil generators. We will merely sketch the process of explicit 2-descent here, and remark that such matters are nowadays typically outsourced to computer algebra systems. We will not determine Ш but merely show that $Ш[2] = 1$ for all three curves using a 2-descent.

  Let $K/\mathbf{Q}$ be the cubic extension obtained by adjoining a 2-torsion point of any of the three curves E. Let $\{p_1, p_2, p_3\} \subset E[2]$ be the set of non-trivial points, then the triple of Weil pairings $\langle p_i, -\rangle$ defines a linear map $E[2] \to \{\pm 1\}^3$, which in fact gives an exact sequence of $G_{\mathbf{Q}}$-modules

  $$1 \longrightarrow E[2] \longrightarrow \operatorname{Ind}_K^{\mathbf{Q}}(\mu_2) \longrightarrow \mu_2 \longrightarrow 1.$$

  Passing to the long exact sequence in $G_{\mathbf{Q}}$-cohomology, we obtain a explicit identification

  $$\mathrm{H}^1(\mathbf{Q}, E[2]) = \operatorname{Ker}\left(K^\times/(K^\times)^2 \xrightarrow{\mathrm{Nm}} \mathbf{Q}^\times/(\mathbf{Q}^\times)^2\right).$$

The Selmer group contains only classes unramified outside 2 and 11, with a local solubility condition at those primes. To compute it, one is now reduced to determining the unit and class groups of K, a set of representatives modulo squares of the archimedean, 2-adic, and 11-adic places of K, and computing the norm. The reader is encouraged to work through the (laborious!) calculation, to find

$$\mathrm{Sel}_2(E/\mathbf{Q}) = 1, \qquad Ш(E/\mathbf{Q})[2] = 1.$$

In summary, we have determined quantities relevant for the Birch–Swinnerton-Dyer conjecture, summarised in the following diagram. We showed that $Ш[2] = 1$, and we might guess that in fact $Ш = 1$.

	Ω	c_∞	c_{11}	Reg	$\lvert Ш\rvert$	$\lvert E(\mathbf{Q})_{\mathrm{tors}}\rvert$
E_1	$6.34604652\ldots$	1	1	1	1?	5
E_2	$1.26920930\ldots$	1	5	1	1?	5
E_3	$0.25384186\ldots$	1	1	1	1?	1

Finally, we may numerically compute that the L-values of all three curves – which all share the same L-function, since they are isogenous – via (1.9). We obtain

$$\mathrm{L}(E, 1) \approx 0.25384186\ldots$$

which agrees to the computed precision with the prediction of the Birch–Swinnerton-Dyer conjecture, as may be verified from the above table; assuming that we indeed have $\lvert Ш(E/\mathbf{Q})\rvert = 1$.

Example 2

There are precisely four elliptic curves of conductor 37 defined over $\mathbf{Q}$, whose minimal Weierstraß models over $\mathbf{Z}$ and torsion subgroups are summarised in the following table:

$A \;:\; y^2 + y \;=\; x^3 - x$	$A(\mathbf{Q})_{\mathrm{tors}} = 1$
$E_1 \;:\; y^2 + y \;=\; x^3 + x^2 - 3x + 1$	$E_1(\mathbf{Q})_{\mathrm{tors}} = \langle (1,0) \rangle \simeq \mathbf{Z}/3\mathbf{Z}$
$E_2 \;:\; y^2 + y \;=\; x^3 + x^2 - 23x - 50$	$E_2(\mathbf{Q})_{\mathrm{tors}} = \langle (8,18) \rangle \simeq \mathbf{Z}/3\mathbf{Z}$
$E_3 \;:\; y^2 + y \;=\; x^3 + x^2 - 1873x - 31833$	$E_3(\mathbf{Q})_{\mathrm{tors}} = 1$

The curve A is its own isogeny class, and the curves $\{E_1, E_2, E_3\}$ form a single isogeny class, readily described from the determination of the torsion groups. We have two pairs of dual 3-isogenies

$$E_1 \xleftrightarrow{3} E_2 \xleftrightarrow{3} E_3$$

where the kernels of the isogenies going to the right are the constant group scheme $\mathbf{Z}/3\mathbf{Z}$ (generated by the 3-torsion points listed above), and the kernels of the isogenies going to the left are isomorphic to μ_3. We remark that also in this example, it just so happens that the same statements remain true for the induced morphisms between Néron models. This observation allows one to determine that the ranks of E_1, E_2, E_3 are all zero, using a flat 3-descent as in § 2. Alternatively, one uses a 2-descent as outlined above.

The curve A is different, and a 2-descent reveals that $A(\mathbf{Q}) \simeq \mathbf{Z}$, generated by $p = (0,0)$. The canonical height of this generator $p = (0,0)$ may be efficiently computed from the naive height function $h((x,y)) = \log\max\{\lvert\mathrm{num}(x)\rvert, \lvert\mathrm{den}(x)\rvert\}$ via the rapidly converging limit

$$\mathrm{Reg}(A/\mathbf{Q}) = \lvert\widehat{h}(p,p)\rvert = \lim_{n\to\infty} \frac{1}{2^{2n}} h([2^n]p) \approx 0.05111141\ldots$$

In conclusion, we find the following table of BSD invariants.

	Ω	c_∞	c_{37}	Reg	$\lvert \text{Ш} \rvert$	$\lvert A(\mathbf{Q})_{\mathrm{tors}} \rvert$
A	$2.99345865\ldots$	2	1	$0.05111141\ldots$	1?	1
E_1	$3.26556478\ldots$	2	1	1	1?	3
E_2	$1.08852159\ldots$	2	3	1	1?	3
E_3	$0.36284053\ldots$	2	1	1	1?	1

The L-values of the elliptic curve A may be computed numerically just as before, expressing $\mathrm{L}^*(A, s)$ as a Mellin transform (1.8) and evaluating it at a truncated q-expansion. This yields approximate values

$$\begin{array}{rcl} \mathrm{L}(A, 1) & \approx & 0.00000000\ldots \\ \mathrm{L}'(A, 1) & \approx & 0.30599977\ldots \end{array}$$

Note that our determination of the sign in the functional equation *proves* that $\mathrm{L}(A, 1) = 0$, and the value of the derivative is in perfect agreement with BSD up to the computed precision, assuming $\text{Ш} = 1$. For the three curves of rank zero on the other hand, we find the approximation

$$\mathrm{L}(E, 1) \;\approx\; 0.72568106\ldots$$

which once again agrees with the prediction of the BSD conjecture, assuming $\text{Ш} = 1$.

1.4 Outline of this course

This course aims to give an overview of some important results on rational points on modular curves that were obtained in the latter half of the 20th century, and their implications for torsion subgroups and ranks of elliptic curves over **Q**, and the Birch–Swinnerton-Dyer conjecture. We will focus mostly on the work of Mazur [Maz77b, Maz77a, Maz78] on the classification of torsion over **Q**, and on the work of Gross–Zagier [GZ85, GZ86, GKZ87] on the heights of Heegner points, and BSD in rank one. These correspond to the two main sources of rational points on modular curves, provided by cusps and CM points respectively.

- Section 2 is about cusps, and illustrates some important ideas of the work of Mazur [Maz77a, Maz78] on the explicit examples of the modular curves $X_1(11)$ and $X_1(13)$, which provide good stepping stones to the general case and the classification of torsion on elliptic curves defined over **Q**.
- Section 3 is about complex multiplication (CM) points, and illustrates the results of Gross and Zagier, starting with their work on factorisations of differences of singular moduli, and subsequent elaborations of this circle of ideas to the full determination of height pairings of Heegner points on elliptic curve, and the work of Kolyvagin on BSD in analytic rank at most one.

Chapter **2**

Cusps and the work of Mazur

We turn to the systematic study of torsion on elliptic curves $E_{\mathbf{Q}}$, and will discuss the groundbreaking work of Mazur [Maz77a, Maz78]. The goals are modest: to illustrate some of the key ideas in the work of Mazur on small examples, to try to provide the reader with some helpful intuition before reading the original papers.

Theorem 3 (Mazur)**.** *Let $E_{\mathbf{Q}}$ be an elliptic curve. The torsion subgroup $E(\mathbf{Q})_{\mathrm{tors}}$ of its Mordell–Weil group is isomorphic to one of the following groups:*

$$E(\mathbf{Q})_{\mathrm{tors}} \simeq \begin{cases} \mathbf{Z}/n\mathbf{Z} & 1 \leq n \leq 10,\ n = 12 \\ \mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2n\mathbf{Z} & 1 \leq n \leq 4 \end{cases}$$

For ℓ a large prime, Mazur studies rational points on $X_1(\ell)$ by descent on a suitably chosen[1] isogeny factor A of the Jacobian $J_0(\ell)$. In a general descent procedure, we usually discern two key steps:

- Imposing only unramifiedness conditions outside a finite set of bad primes S gives an a priori bound on the Selmer group, which can be represented by explicit classes.
- This bound is sharpened using explicit equations for twists in the Weil–Châtelet group, where classes are excluded from the image of global points using local obstructions at places in S.

An appealing feature of modular Jacobians is that their cuspidal subgroups, which we study in § 2.1, provide rational points of large order, and they have good reduction outside of ℓ. This makes the a priori bound arising in the first step quite good, though ultimately not good enough. Needless to say, methods involving the Weil–Châtelet group and explicit equations for twists are not suited to further sharpen this bound. This is already the case for modest values of ℓ, and certainly for general ℓ.

To obtain sharper bounds, one may spread out the geometry over $\mathrm{Spec}(\mathbf{Z})$, and work with respect to flat topology. Mazur shows the existence of propitious quotients A whose p-torsion is *admissible* for some $p \nmid N$, a stringent condition that assures the Jordan–Hölder factors to be $\mathbf{Z}/p\mathbf{Z}$ or μ_p. The descent formalism applied to the identity component $\mathscr{A}^0$ of the Néron model $\mathscr{A}$ over $\mathrm{Spec}(\mathbf{Z})$ then gives

$$1 \longrightarrow \mathscr{A}^0(\mathbf{Z})/p\,\mathscr{A}^0(\mathbf{Z}) \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathrm{Spec}(\mathbf{Z}), \mathscr{A}^0(\mathbf{Z})[p]) \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathrm{Spec}(\mathbf{Z}), \mathscr{A}^0(\mathbf{Z}))[p] \longrightarrow 1$$

The admissibility of the Galois module in the middle term allows Mazur to sufficiently control the corresponding pieces of the cohomology, frequently using Kummer theory through the fact that

$$1 \longrightarrow \mu_n \longrightarrow \mathbf{G}_m \xrightarrow{\cdot^n} \mathbf{G}_m \longrightarrow 1 \tag{2.1}$$

is an exact sequence of sheaves on the flat site (it is not generally right exact in the étale site).

[1] It is the so-called “Eisenstein quotient”, whose construction is a key part of the work of Mazur [Maz78].

2.1 Cusps on modular curves

The simplest systematic supply of points on modular curves is provided by the *cusps*, which correspond to degenerations of elliptic curves, with appropriate level structures. We begin with a few concrete examples as an illustration. Let E be any elliptic curve defined over a $\mathbf{Q}$-algebra, and q a rational point on E of order greater than 3. We may translate q to the origin and rescale coordinates to put E in Tate normal form

$$E : y^2 + uxy + vy = x^3 + vx^2.$$

We now explore the modular curves $X_1(p)$ for small values of p, using the group law to algebraically express the equality $[p]q = 0$, and resolving the singularities of the resulting algebraic relations between u and v.

- **The modular curve $X_1(5)$.** Suppose that q is a point of order 5. Expressing the equality $3q = -2q$ in coordinates, one finds this is equivalent to $u = v + 1$. This allows one to deduce a classification of elliptic curves with a point of order 5; they all arise from pullbacks of the universal family

  $$E_v : y^2 + (1 + v)xy + vy = x^3 + vx^2$$

  We easily compute that this curve has discriminant given by

  $$\Delta = -v^5(v^2 + 11v - 1)$$

  We conclude from these computations that the modular curve $X_1(5)$ must be isomorphic to $\mathbf{P}^1_v$ and there are precisely four cusps: the points $v = 0, \infty$ defined over $\mathbf{Q}$, and the roots of the quadratic polynomial $v^2 + 11v - 1$, which are defined over the field $\mathbf{Q}(\sqrt{5}) = \mathbf{Q}(\zeta_5)^+ = \mathbf{Q}(\zeta_5 + \zeta_5^{-1})$.

- **The modular curve $X_1(7)$.** Now suppose that q is a point of order 7. The equality $4q = -3q$ yields more complicated equations in u and v, which exhibit a singularity. We can resolve this singularity by the birational transformation $(u, v) \mapsto (s, s(t - 1))$, transforming the relations to $s + t^2 = t + 1$. Therefore, the universal elliptic curve with a point of order 7 is

  $$E_t : y^2 - (t^2 - t - 1)xy - (t^3 - t^2)y = x^3 - (t^3 - t^2)x^2.$$

  The discriminant of this curve factorises as follows:

  $$\Delta = t^7(t - 1)^7(t^3 - 8t^2 + 5t + 1).$$

  These calculations show that the modular curve $X_1(7)$ is isomorphic to the projective line $\mathbf{P}^1_t$ and there are precisely six cusps: the points $t = 0, 1, \infty$ defined over $\mathbf{Q}$, as well as the roots of the cubic polynomial $t^3 - 8t^2 + 5t + 1$, which happen to be defined over $\mathbf{Q}(\zeta_7)^+ = \mathbf{Q}(\zeta_7 + \zeta_7^{-1})$.

- **The modular curve $X_1(11)$.** Suppose q is a point of order 11, then we may algebraically express the equality $6q = -5q$, which yields (singular) equations in (u, v); I omit them here, to avoid spoiling the reader's appetite. The desingularisation is achieved by the transformation

  $$(u, v) \longmapsto \left(\frac{(r-1)(r-s-1)}{s} + 1,\ \frac{r(r-1)(r-s-1)}{s} \right)$$

  which yields an explicit equation for $X_1(11)$. It is an elliptic curve, with equation

  $$X_1(11) : r^2 - r + s^3 + s^2 = 0. \tag{2.2}$$

This is an old friend! We recognise it as the example E_1 appearing at the end of the introduction. We described there how to determine the rank by a 2-descent, using the explicit Weierstrass equation, which tells us that $X_1(11)(\mathbf{Q}) \simeq \mathbf{Z}/5\mathbf{Z}$. We will return to this example in § 2, where we will determine its Mordell–Weil group by "pure thought" without recourse to any explicit equations.

To determine the cusps, we look at the universal elliptic curve over $X_1(11)$, which has discriminant Δ given by some horrendous expression. Computing the resultants of its numerator with the relation in (2.2), we find that for the universal curve to be singular, we must have

$$s = 0 \qquad \text{or} \qquad s+1=0 \qquad \text{or} \qquad \begin{cases} r^5 - 63r^4 + 85r^3 - 4r^2 - 21r + 1 &= 0 \\ s^5 + 18s^4 + 35s^3 + 16s^2 - 2s - 1 &= 0 \end{cases}$$

The rational torsion points on $X_1(11)$ fully account for the two former conditions, whereas the latter system of equations has solutions defined over the maximal totally real subfield of $\mathbf{Q}(\zeta_{11})$. This yields precisely 5 cusps defined over $\mathbf{Q}$, and 5 cusps defined over the cyclic quintic field $\mathbf{Q}(\zeta_{11})^+$. Note that all rational points are cusps, and as a consequence we find that there are no elliptic curves defined over $\mathbf{Q}$ that possess a rational point of order 11.

- **The modular curve $X_1(p)$ for $p > 11$.** Writing down explicit equations in this way quickly becomes cumbersome, as the complexity of the singularities grows. An elegant continuation via this method can be found in Mestre [Mes81], who finds (for instance) the following equation when $p = 13$:

  $$X_1(13) : y^2 + y(x^3 - x^2 - 1) + x(1-x) = 0.$$

  Mestre finds the coordinates of the cusps, of which there are precisely 6 defined over $\mathbf{Q}$ and 6 defined over $\mathbf{Q}(\zeta_{13})^+$. If we wish to show that there are no elliptic curves over $\mathbf{Q}$ with a rational point of order 13, we would need to determine the set of rational points. It is not obvious how to do this, or how to even determine the rank of the Jacobian. We will return to this example in § 2.

After our explicit investigation of small examples, we now completely abandon our efforts to write down equations. Questions about the geometry of cusps are understood moduli-theoretically, in terms of equivalence classes of level structures on generalised elliptic curves, see [DR73, KM85]. We briefly summarise in what follows the results contained in [KM85, Chapter 10], and content ourselves with a practical discussion on how to compute the set of cusps as a Galois-set, in some concrete examples.

Calculations with cusps

The set of cusps for a subgroup $\Gamma \geq \Gamma(n)$ is the set of level structures on the Néron n-gon $\mathbf{G}_m \times \mathbf{Z}/n\mathbf{Z}$, whose n-torsion is isomorphic to $\mu_n \times \mathbf{Z}/n\mathbf{Z}$. The choices of basis with a fixed value of the Weil pairing are indexed (as row vectors) by the group $\mathrm{SL}_2(\mathbf{Z}/n\mathbf{Z})$. The symplectic automorphism group

$$\mathrm{A} := \pm \begin{pmatrix} 1 & 0 \\ * & 1 \end{pmatrix}$$

acts on the set of these choices by n-torsion points by left multiplication in $\mathrm{SL}_2(\mathbf{Z})$, whose orbit are the cusps for $\Gamma(n)$. We conclude that the set of cusps for Γ is in bijection with the double coset

$$\mathrm{A} \setminus \mathrm{SL}_2(\mathbf{Z}/n\mathbf{Z}) / \Gamma,$$

where we abuse notation and denote the image of Γ in the quotient $\mathrm{SL}_2(\mathbf{Z})/\Gamma(n) = \mathrm{SL}_2(\mathbf{Z}/n\mathbf{Z})$ simply by Γ. The Galois action on the cusps is described on the double cosets via the cyclotomic character on the entries of the first column, and the trivial action on the entries of the second column. For specific examples, this data may be computed straightforwardly by first identifying the left cosets

$$\mathrm{A} \setminus \mathrm{SL}_2(\mathbf{Z}/n\mathbf{Z}) \simeq \{\mathbf{v} = (x, y) \in (\mathbf{Z}/n\mathbf{Z})^2 \text{ of order } n\}/\pm 1$$

via the process of row reduction. This is an isomorphism of Galois sets, if we let $G_{\mathbf{Q}}$ act on the first entry of the row vectors $\mathbf{v}$ via the cyclotomic character, and trivially on the second entry. The cusps are the orbits O of Γ acting on $\mathbf{v}$ by right matrix multiplication. For every orbit O, one easily keeps track of its size, its width $p|O|/|\Gamma|$, and its field of definition i.e. the smallest subfield of $\mathbf{Q}(\zeta_n)$ that stabilises the orbit O.

We make this explicit in the case $n = p$ and $\Gamma = \mathrm{SL}_2(\mathbf{F}_p) \cap G$ where $G \leq \mathrm{GL}_2(\mathbf{F}_p)$ is maximal. The classification of maximal subgroups in $\mathrm{GL}_2(\mathbf{F}_p)$ is well-known; we discuss each option in turn.

- **Borel subgroups.** These are subgroups conjugate to the group

$$G = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \leq \mathrm{GL}_2(\mathbf{F}_p).$$

The associated modular curve is $X_0(p)$, defined over $\mathbf{Q}$. To compute the Galois set of its cusps, we compute the action of a general element of Γ, and find the identity

$$(x, y) \cdot \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} = (ax, a^{-1}y + bx).$$

From this identity it is easy to work out the orbits, their sizes, the width $2p|O|/|\Gamma|$ of the corresponding cusp, and their field of definition. This information is summarised in the following table.

O_i	$\lvert O_i \rvert$	Width	Field
$i = 0$	$(p-1)/2$	1	$\mathbf{Q}$
$i \neq 0$	$p(p-1)/2$	p	$\mathbf{Q}$

where $O_i := \{(x, y) : x = i\}/\pm$.

There are exactly two cusps, and they are defined over $\mathbf{Q}$. It is customary to denote these cusps by the symbols ∞ and 0 respectively; projective representatives of these orbits.

- **Normaliser of a split Cartan subgroup.** These are conjugate to the group

$$G := \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix} \cup \begin{pmatrix} 0 & * \\ * & 0 \end{pmatrix} \leq \mathrm{GL}_2(\mathbf{F}_p).$$

The associated modular curve $X^+_{\mathrm{spl}}(p)$ is defined over $\mathbf{Q}$. To compute the Galois set of its cusps, we take explicit elements of the group Γ and find that they act by the following identities

$$\pm(x, y) \cdot \begin{pmatrix} a & \\ & a^{-1} \end{pmatrix} = \pm(ax, a^{-1}y), \qquad \pm(x, y) \cdot \begin{pmatrix} & 1 \\ -1 & \end{pmatrix} = \pm(y, -x).$$

From these identities, we work out once again the cusps data, summarised in the following table.

O_i	Amount	$\lvert O_i \rvert$	Width	Field
$i = 0$	1	$p - 1$	p	$\mathbf{Q}$
$i \neq 0$	$(p-1)/2$	$p - 1$	p	$\mathbf{Q}(\zeta_p)^+$

where $O_i := \{(x, y) : xy = \pm i\}/\pm$.

In particular, we see that there is precisely 1 rational cusp represented by ∞, and $(p - 1)/2$ cusps defined over the cyclic extension $\mathbf{Q}(\zeta_p)^+$, all Galois conjugate to each other.

- **Normaliser of a non-split Cartan subgroup.** These are groups of the form

$$G = \mathbf{F}_{p^2}^\times \rtimes (\mathbf{Z}/2\mathbf{Z}) \leq \mathrm{GL}_2(\mathbf{F}_p), \qquad \text{where} \ \ \mathbf{F}_{p^2}^\times \leq \mathrm{Aut}_{\mathbf{F}_p}(\mathbf{F}_{p^2})$$

where the unit group of the finite field $\mathbf{F}_{p^2}$ acts by multiplication, viewed as an element of $\mathrm{GL}_2(\mathbf{F}_p)$ after choosing a basis for $\mathbf{F}_{p^2}$ as a 2-dimensional $\mathbf{F}_p$-vector space. The associated modular curve $X_{\mathrm{ns}}^+(p)$ is defined over $\mathbf{Q}$. To work out its cusps, it is more convenient to reinterpret the set of non-zero row vectors as the set of elements $\alpha \in \mathbf{F}_{p^2}^\times$. The action of

$$\Gamma = G \cap \mathrm{SL}_2(\mathbf{F}_p) = \mu_{p+1} \rtimes \mathbf{Z}/2\mathbf{Z}$$

is described by letting μ_{p+1} acts by right multiplication, and $\mathbf{Z}/2\mathbf{Z}$ by inversion. We see in this way that there are precisely $(p - 1)/2$ orbits, with their cusp data summarised in the following table:

O_i	Amount	$\lvert O_i \rvert$	Width	Field
$i \neq 0$	$(p-1)/2$	$p + 1$	p	$\mathbf{Q}(\zeta_p)^+$

where $O_i := \{\alpha : \mathrm{Nm}(\alpha) = i^{\pm 1}\}/\pm$.

In particular, we see that there are precisely $(p-1)/2$ cusps, which form a full Galois conjugacy class, and are all defined over the cyclic extension $\mathbf{Q}(\zeta_p)^+$.

- **An exceptional group.** This concerns projective lifts $G \leq \mathrm{GL}_2(\mathbf{F}_p)$ of a subgroup

$$\overline{G} = A_4, S_4, A_5 \leq \mathrm{PGL}_2(\mathbf{F}_p).$$

The arithmetic interest of these exceptional curves is more limited, and we decided to omit the cusp calculations here. The interested reader is encouraged to fill in this omission as an exercise.

Besides the maximal subgroups $G \leq \mathrm{GL}_2(\mathbf{F}_p)$ considered here, we also discuss the case

$$G = \begin{pmatrix} 1 & * \\ 0 & * \end{pmatrix} \leq \mathrm{GL}_2(\mathbf{F}_p)$$

which gives us $\Gamma = \Gamma_1(p)$, corresponding to the modular curves $X_1(p)$ which are defined over $\mathbf{Q}$. This will explain the properties of the explicit coordinates for cusps we worked out at the beginning of this section from the equations for the universal elliptic curves. In this case, we have the following identity

$$(x, y) \cdot \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = (x, y + bx).$$

from which we recover the following description of the cusps

$O_{i,j}$	Amount	$\|O_i\|$	Width	Field
$i=0$	$(p-1)/2$	1	1	$\mathbf{Q}$
$i\neq 0$	$(p-1)/2$	$(p-1)/2$	$p-1$	$\mathbf{Q}(\zeta_p)^+$

where $O_{i,j} := \{(i, j + i\cdot \mathbf{F}_p)\} / \pm$.

We conclude that there are precisely $(p-1)/2$ cusps defined over $\mathbf{Q}$, and $(p-1)/2$ Galois conjugate cusps defined over the cyclic extension $\mathbf{Q}(\zeta_p)^+$. This is no surprise, and we had long guessed this after working out the coordinates of cusps on the explicit examples of $X_1(p)$ for small p at the start of this section.

The theorem of Manin–Drinfeld

Now that we have a good grasp of the calculus of cusps on a modular curve X_Γ, and understand their fields of definition very concretely, we are naturally lead to the question:

Q: What is the structure of the subgroup of $\mathrm{Jac}(X_\Gamma)$ generated by cuspidal divisors?

Cuspidal divisors of degree zero supported on the cusps can be fruitfully understood using the transcendental description of the Jacobian over $\mathbf{C}$. We use the notation $S_2(\Gamma)$ for the $\mathbf{C}$-vector space of holomorphic cusp forms on Γ and use the description

$$\mathrm{Jac}(X_\Gamma)(\mathbf{C}) = \mathrm{Hom}_{\mathbf{C}}\left(S_2(\Gamma), \mathbf{C}\right) / \mathrm{H}_1(X_\Gamma, \mathbf{Z}).$$

Under this identification, a cuspidal divisor $(s_1) - (s_2)$ corresponds to the integration map

$$\{s_1, s_2\} : f \longmapsto \int_{s_1}^{s_2} f(z)dz$$

along any path in the upper half plane from s_1 to s_2. Changing s_1 or s_2 in their Γ-equivalence class only changes the result up to $\mathrm{H}_1(X_\Gamma, \mathbf{Z})$. Similarly, the integration pairing that sends (f, γ) to the integral over γ of the holomorphic differential $f(z)dz$ gives a canonical identification

$$\mathrm{Hom}_{\mathbf{C}}\left(S_2(\Gamma), \mathbf{C}\right) \simeq \mathrm{H}_1(X_\Gamma, \mathbf{R})$$

which allows us to think of the complex points on the Jacobian as the set $\mathrm{H}_1(X_\Gamma, \mathbf{R}) / \mathrm{H}_1(X_\Gamma, \mathbf{Z})$. The following theorem is due to Manin–Drinfeld [Man72, Dri73] who proved that cuspidal divisors $(s_1) - (s_2)$ are always torsion, i.e. the symbol $\{s_1, s_2\}$ is contained in the subgroup $\mathrm{H}_1(X_\Gamma, \mathbf{Q})$.

Theorem 4 (Manin–Drinfeld)**.** *Let Γ be a congruence subgroup of $\mathrm{SL}_2(\mathbf{Z})$. Then any divisor of degree zero supported on the set of cusps of the modular curve X_Γ has finite order in the Jacobian.*

Proof. To prove this theorem, note that it suffices to deal with the case $\Gamma = \Gamma(n)$ for some $n \geq 1$. Choose a prime $p \equiv 1 \pmod{n}$. Choose a set of representatives for the double coset

$$\Gamma(n) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} \Gamma(n) = \bigsqcup_{i=1}^{k} \Gamma(n)\alpha_i$$

The equivalence classes of cusps $s \in \mathbf{P}^1(\mathbf{Q})$ are given by the congruence classes modulo n of their projective coordinates, i.e. by the elements of $\mathbf{P}^1(\mathbf{Z}/n\mathbf{Z})$. Since any of the matrices α_i are congruent to the identity matrix modulo N, we have $\alpha_i s = s \bmod \Gamma(n)$ for all cusps s. Therefore

$$(T_p - kI)\{s_1, s_2\} \in \mathrm{H}_1(X(n), \mathbf{Z}),$$

for any pair of cusps s_1, s_2. The operator $(T_p - kI)$ is invertible, since k is larger than any eigenvalue of T_p on the space of cusp forms $S_2(\Gamma(n))$. It follows that for any pair of cusps s_1, s_2 we have

$$\{s_1, s_2\} \;\in\; \mathrm{H}_1(X(n), \mathbf{Q})$$

and therefore $(s_1) - (s_2)$ defines a torsion point on the Jacobian. □

Siegel units

Since any cuspidal divisor is torsion by Manin–Drinfeld, the Jacobian of a modular curve associated to a congruence subgroup has a finite subgroup G generated by cuspidal divisors. The structure of G may be studied with *Siegel units*. We make a few brief comments, and refer to Kubert–Lang [KL81] for details.

The key to understanding cuspidal torsion is to construct explicit generators for the unit group of the function field of the *affine* modular curve Y_Γ. Let $v = (a, b) \in (\mathbf{Q}\,/\,\mathbf{Z})^2 - \{0\}$, then the associated *Siegel unit* is a complex analytic function on the upper half plane, described explicitly by

$$g_v(\tau) = -q^{B_2(a)/2} \cdot e(b(a-1)/2) \cdot (1 - e(b)q^a) \cdot \prod_{n=1}^{\infty} (1 - e(b)q^{n+a})(1 - e(-b)q^{n-a})$$

where $e(z) = \exp(2\pi i z)$ and $B_2(x) = \{x\}^2 - \{x\} + 1/6$ is the second Bernoulli polynomial. The collection of Siegel units satisfies certain norm compatibility laws, for any $m \geq 1$, explicitly given by

$$\prod_{ma'=a} g_{(a',b)}(\tau) = g_{(a,b)}(\tau/m), \qquad\qquad \prod_{mb'=b} g_{(a,b')}(\tau) = g_{(a,b)}(m\tau).$$

These properties are very important, and lie at the basis of its forming an *Euler system*, an important property of great significance in contemporary number theory. When $v = (a, b) \in \left(\frac{1}{n}\,\mathbf{Z}\right) / \,\mathbf{Z}$ then

$$\begin{cases} g_v & \in \quad \mathcal{O}(Y(n))^\times \otimes \mathbf{Q} \\ g_v^{12n} & \in \quad \mathcal{O}(Y(n))^\times \end{cases}$$

and in general, letting $\varepsilon : \mathrm{SL}_2(\mathbf{Z}) \to \mu_{12}$ be the abelianisation suitably embedded in $\mathbf{C}$, they transform via

$$g_v(\gamma\tau) = \varepsilon(\gamma) g_{v\gamma}(\tau).$$

Since the case of composite level structures comes with additional complications, we will assume for simplicity that the level $n = p \geq 5$ is prime, and consider congruence subgroups $\Gamma \geq \Gamma(p)$. Our goal is to construct explicit relations in the Jacobian using elements of the function field of X_Γ with cuspidal divisors, constructed from Siegel units. Concretely, we wish to construct units of the form

$$u_M(\tau) := \prod_{v \in \left(\frac{1}{p}\,\mathbf{Z}\,/\,\mathbf{Z}\right)^2 - \{0\}} g_v(\tau)^{M_v},$$

where M_v is a collection of integers that is chosen such that $u_M(\tau)$ defines an element of $\mathcal{O}(Y(p))^\times$. Such collections of integers are fully characterised by the following criteria:

1. for all $\gamma \in \Gamma$ we have $M_{v\gamma} = M_v$, i.e. M_v is constant on Γ-orbits,
2. the sum of all M_v is divisible by 12,

3. the following congruences modulo p are satisfied:

$$\sum_{v=(v_1,v_2)} M_v v_1^2 \equiv \sum_{v=(v_1,v_2)} M_v v_2^2 \equiv \sum_{v=(v_1,v_2)} M_v v_1 v_2 \equiv 0 \pmod{p}.$$

The above conditions may be solved in any example by straightforward linear algebra over $\mathbf{Z}$, and reduction to the Smith normal form will reveal the structure of the torsion subgroup of $\mathrm{Jac}(X_\Gamma)$ generated by the cusps. Concretely, we see that equivalently, we wish to determine an integer M_O for every Γ-orbit O of primitive row vectors in $\mathbf{F}_p^2$, which correspond to the cusps as we computed them in § 2.1. For every such orbit O, we now also wish to keep track of the data

$$s_1(O) := \sum_{(a,b)\in O} a^2, \qquad s_2(O) := \sum_{(a,b)\in O} b^2, \qquad s_{12}(O) := \sum_{(a,b)\in O} ab \qquad \pmod{p}$$

where it suffices to calculate these integers modulo p. We also wish to keep track of the divisor of the function $g_O = \prod_{v\in O} g_v$, which is given explicitly by

$$\mathrm{div}(g_O) = \sum_{c \text{ cusp}} \frac{\mathrm{width}(c)}{2} \sum_{v\in O} B_2\left((vA_c)_1\right) \cdot (c)$$

where for any cusp c we denote A_c for any matrix such that $A_c\infty = c$.

Let us do this explicitly for $\Gamma = \Gamma_0(p)$. As before, we have two orbits O and the corresponding cusp data is given by

Orbit O	Amount	$\lvert O\rvert$	Width	$s_1(O)$	$s_2(O)$	$s_{12}(O)$	$\mathrm{div}(g_O)$
$\{(0,*)\}$	1	$p-1$	1	0	0	0	$\frac{p-1}{12}(\infty) - \frac{p-1}{12}(0)$
$\{(x,*) : x \neq 0\}$	1	$p(p-1)$	p	0	0	0	$\frac{1-p}{12}(\infty) + \frac{p-1}{12}(0)$

We see that the only non-trivial linear condition on the pair of integers $m(O_1)$ and $m(O_2)$ that we need to solve for is that $m(O_1)(p-1)$ is divisible by 12, and likewise for $m(O_2)$. We find that a fundamental solution is given by $m(O_1)$ equal to the *denominator* of $(p-1)/12$, making the order of the cuspidal divisor $(\infty) - (0)$ equal to the *numerator* of $(p-1)/12$.

Going back to our discussion in the introduction, we recall that the modular curve $X_0(11)$ was an elliptic curve whose torsion subgroup was generated by a pair of cusps, and was of order 5, which is indeed the numerator of $10/12 = 5/6$. Carrying out this procedure for a number of small examples, we find the following table for the torsion subgroup of the Jacobian of some modular curves of small level:

p	11	13	17	19	23
$J_0(p)(\mathbf{Q})_{\mathrm{tors}}$	$\mathbf{Z}/5\mathbf{Z}$	1	$\mathbf{Z}/4\mathbf{Z}$	$\mathbf{Z}/3\mathbf{Z}$	$\mathbf{Z}/11\mathbf{Z}$
$J_1(p)(\mathbf{Q})_{\mathrm{tors}}$	$\mathbf{Z}/5\mathbf{Z}$	$\mathbf{Z}/19\mathbf{Z}$	$\mathbf{Z}/8\mathbf{Z} \times \mathbf{Z}/73\mathbf{Z}$	$\mathbf{Z}/9\mathbf{Z} \times \mathbf{Z}/487\mathbf{Z}$	$\mathbf{Z}/11\mathbf{Z} \times \mathbf{Z}/37181\mathbf{Z}$

Note that one additional justification is necessary: whereas the method sketched above only successfully determines the structure of the torsion subgroup generated by the cusps, it turns out that in all these examples, this subgroup is equal to the full torsion group on the Jacobian. This type of statement was conjectured by Ogg [Ogg75] and proved by Mazur [Maz78] for $X_0(p)$ and by Ohta for $X_1(p)$ [CES03, Oht13].

2.2 Torsion points of order 11

Since $X_1(11)$ is an elliptic curve, a simple 2-isogeny already told us that the rank was zero, and all the torsion was cuspidal, so we concluded that no elliptic curves $E/\mathbf{Q}$ can have a rational point of order 11. We will revisit this example using a descent by 5-isogeny instead. This allows us to explore some of the fundamental ideas in Mazur–Tate [MT73], and ultimately Mazur [Maz72, Maz77a, Maz78].

Both $X_0(11)$ and $X_1(11)$ are elliptic curves defined over $\mathbf{Q}$, with minimal Weierstraß models:

$$\begin{array}{rcrcl} X_1(11) & : & y^2+y & = & x^3-x^2 \\ X_0(11) & : & y^2+y & = & x^3-x^2-10x-20 \end{array} \tag{2.3}$$

The curve $X_1(11)$ has cuspidal 5-torsion point by § 2.1, which defines the isogeny

$$\phi : X_1(11) \longrightarrow X_0(11)$$

corresponding to the forgetful map on moduli problems that sends the point P of order 11 to the subgroup $\langle P \rangle$ of order 11. This 5-torsion point is 'meaningful', in the sense that it is generated by cusps and similar torsion is available on other modular curves. We will first spread out the geometry of this isogeny over $\mathbf{Z}$.

Néron models. Denote the Néron models of $X_0(11)$ and $X_1(11)$ over $\mathbf{Z}$ by $\mathscr{X}_0(11)$ and $\mathscr{X}_1(11)$ respectively. The only prime of bad reduction is 11, where Tate's algorithm finds that the reduction has Kodaira types I_5 and I_1 respectively. We may visualise the Néron models over $\mathbf{Z}_{11}$ as follows.

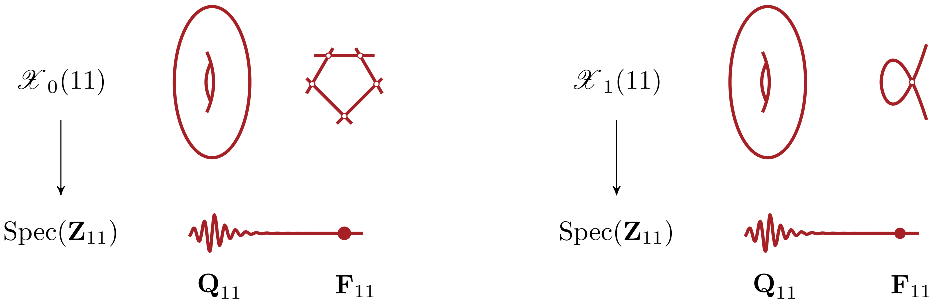


Figure 2.1: The Néron models $\mathscr{X}_0(11)$ and $\mathscr{X}_1(11)$ over $\operatorname{Spec}(\mathbf{Z}_{11})$.

Kernels. The pair of dual 5-isogenies between $X_0(11)$ and $X_1(11)$ extend to a pair of morphisms between the Néron models. Their kernels are group schemes, whose generic fibres are isomorphic to the group schemes $\mathbf{Z}/5\mathbf{Z}$ and μ_5 respectively. We will determine the kernels as group schemes over $\mathbf{Z}$, though in first instance only over $\mathbf{Z}[1/11]$, where these kernels are automatically finite flat group schemes.

- Looking at the coordinates of the torsion points in the minimal Weierstraß models, we find

  $$\begin{array}{rcl} X_1(11)(\mathbf{Q})_{\text{tors}} & = & \{0,(0,0),(0,-1),(1,0),(1,-1)\} \quad \simeq \quad \mathbf{Z}/5\mathbf{Z} \\ X_0(11)(\mathbf{Q})_{\text{tors}} & = & \{0,(5,5),(5,-6),(16,60),(16,-61)\} \quad \simeq \quad \mathbf{Z}/5\mathbf{Z} \end{array}$$

  Note that on $X_1(11)$ the torsion is injective when reduced modulo every prime, including 5, so the kernel of the forgetful isogeny $\mathscr{X}_1(11) \to \mathscr{X}_0(11)$ is $\mathbf{Z}/5\mathbf{Z}$ over $\mathbf{Z}[1/11]$. Via the Weil pairing, we find that the kernel of the dual isogeny $\mathscr{X}_0(11) \to \mathscr{X}_1(11)$ is isomorphic to μ_5 over $\mathbf{Z}/5\mathbf{Z}$.

- The second proof does not rely on explicit equations. The kernels of the isogenies over $\mathbf{Z}[1/11]$ are finite flat group schemes. Since they agree with the finite flat group schemes $\mathbf{Z}/5\mathbf{Z}$ and μ_5 on the generic fibre $\mathbf{Q}$, and since the ramification index of the base is smaller than $5-1=4$, the classification theorem of Oort–Tate [TO70] implies that the kernels are isomorphic to $\mathbf{Z}/5\mathbf{Z}$ and μ_5 over $\mathbf{Z}[1/11]$.

It is in fact also true over $\mathbf{Z}$ that the kernels are isomorphic to $\mathbf{Z}/5\mathbf{Z}$ and μ_5 respectively. The reader who is tempted to undertake the 5-descent that follows over $\mathbf{Z}$ is cautioned not to forget that the map $\mathscr{X}_1(11) \to \mathscr{X}_0(11)$ is surjective over $\mathbf{Z}[1/11]$ but not over $\mathbf{Z}$, the cokernel being the Néron component group in the special fibre at 11, as we will discover shortly by a simple local argument.

Descent by 5-isogeny

Over the base $\mathbf{Z}[1/11]$ we are now faced with a pair of exact sequences

$$\begin{array}{ccccccccc} 1 & \longrightarrow & \mathbf{Z}/5\mathbf{Z} & \longrightarrow & \mathscr{X}_1(11) & \xrightarrow{\phi} & \mathscr{X}_0(11) & \longrightarrow & 1 \\ 1 & \longrightarrow & \mu_5 & \longrightarrow & \mathscr{X}_0(11) & \xrightarrow{\widehat{\phi}} & \mathscr{X}_1(11) & \longrightarrow & 1 \end{array}$$

of sheaves in the fppf topology. We use a descent for each of these exact rows in turn, estimating the resulting fppf Selmer groups by Kummer theory. By the defining properties of Néron models, we have

$$\begin{array}{lllll} \mathrm{H}^0_{\mathrm{fppf}}(\mathbf{Z}[1/11], \mathscr{X}_0(11)) & = & \mathscr{X}_0(11)(\mathbf{Z}[1/11]) & = & X_0(11)(\mathbf{Q}) \\ \mathrm{H}^0_{\mathrm{fppf}}(\mathbf{Z}[1/11], \mathscr{X}_1(11)) & = & \mathscr{X}_1(11)(\mathbf{Z}[1/11]) & = & X_1(11)(\mathbf{Q}). \end{array}$$

1. We begin with the isogeny of $\mathbf{Z}[1/11]$-abelian schemes

   $$\phi : \mathscr{X}_1(11) \to \mathscr{X}_0(11),$$

   which has kernel $\mathbf{Z}/5\mathbf{Z}$. The long exact sequence in cohomology, taken in either the fppf or étale cohomology since the group schemes appearing are smooth [Mil80, III.3.9], gives an injection

   $$\frac{X_0(11)(\mathbf{Q})}{\phi(X_1(11)(\mathbf{Q}))} \hookrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/11], \mathbf{Z}/5\mathbf{Z}) = \mathrm{H}^1_{\text{ét}}(\mathbf{Z}[1/11], \mathbf{Z}/5\mathbf{Z}) \simeq \mathbf{F}_5,$$

   where the isomorphism is because there is a unique cyclic degree 5 extension of $\mathbf{Q}$ unramified outside 11, namely $\mathbf{Q}(\zeta_{11})^+$. The torsion on $X_0(11)(\mathbf{Q})$ gives the same lower bound, so we find

   $$\frac{X_0(11)(\mathbf{Q})}{\phi(X_1(11)(\mathbf{Q}))} \simeq \mathbf{Z}/5\mathbf{Z}\,.$$

2. The dual isogeny of $\mathbf{Z}[1/11]$-abelian schemes

   $$\hat{\phi} : \mathscr{X}_0(11) \longrightarrow \mathscr{X}_1(11)$$

   has kernel μ_5. Using $\mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/11], \mathbf{G}_m) = \mathrm{Pic}(\mathbf{Z}[1/11]) = 0$, the long exact sequence in flat cohomology associated to the Kummer short exact sequence gives us an isomorphism

   $$\mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/11], \mu_5) \simeq \mathbf{F}_5\,.$$

   We need to further sharpen this bound, since we are trying to show that the group $X_1(\mathbf{Q})/\hat{\phi}(X_0(\mathbf{Q}))$, which injects into it, is trivial. This will come from controlling the image at the final frontier: The prime 11. Indeed, the descent sequence gives us the following commutative diagram

$$\begin{tikzcd}
X_0(\mathbf{Q}) \arrow[r] \arrow[d] & X_1(\mathbf{Q}) \arrow[r] \arrow[d] & \mathrm{H}^1_{\mathrm{fppf}}\big(\mathbf{Z}[1/11], \mu_5\big) \arrow[d, "(**)"] & \simeq \mathbf{F}_5 \\
X_0(\mathbf{Q}_{11}) \arrow[r, "(*)"] & X_1(\mathbf{Q}_{11}) \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}\big(\mathbf{Q}_{11}, \mu_5\big) & \simeq \mathbf{F}_5^2
\end{tikzcd}$$

The triviality of the group $X_1(\mathbf{Q})/\hat{\phi}(X_0(\mathbf{Q}))$ would finally follow, if we could show that the map $(*)$ is surjective, and the map $(**)$ is injective. Let us check these in turn.

- Let us check first that $(*)$ is surjective. If we denote M_0 and M_1 to be the kernels of the reduction maps of $\mathscr{X}_0(\mathbf{Z}_{11})$ and $\mathscr{X}_1(\mathbf{Z}_{11})$ modulo 11, then the map $(*)$ fits into the following commutative diagram with exact rows:

$$\begin{tikzcd}
0 \arrow[r] & M_0 \arrow[r] \arrow[d, "\hat{\phi}"] & \mathscr{X}_0(\mathbf{Z}_{11}) \arrow[r] \arrow[d, "(*)"] & \mathscr{X}_0(\mathbf{F}_{11}) \arrow[r] \arrow[d, "\hat{\phi}"] & 0 \\
0 \arrow[r] & M_1 \arrow[r] & \mathscr{X}_1(\mathbf{Z}_{11}) \arrow[r] & \mathscr{X}_1(\mathbf{F}_{11}) \arrow[r] & 0
\end{tikzcd}$$

  Note that to prove that $(*)$ is surjective, it suffices to show that its left and right flanking maps are surjective, by the snake lemma. For the left map, we note that $[5] = \phi \circ \hat{\phi}$ is an isomorphism on the pro-11 group M_1, and therefore so is $\hat{\phi}$. Surjectivity of the right map

$$\mathscr{X}_0(11)(\mathbf{F}_{11}) \longrightarrow \mathscr{X}_1(11)(\mathbf{F}_{11})$$

  can be checked directly, by observing that it factors through the induced map on minimal Weierstraß models, which can be computed explicitly and shown to have trivial kernel, hence trivial cokernel since the special fibres are both cyclic groups of order 10.

- Finally, we note that the map $(**)$ is identified via the long exact sequence of flat cohomology associated to the Kummer exact sequence with the natural map

$$\pm 11^{\mathbf{Z}} / \pm 11^{5\,\mathbf{Z}} \longrightarrow \mathbf{Q}_{11}^{\times} / (\mathbf{Q}_{11}^{\times})^5.$$

  This map is clearly injective, since 11 is not a fifth power in $\mathbf{Q}_{11}^{\times}$.

The two descent sequences above have therefore concluded that we have a pair of isomorphisms

$$\frac{X_0(11)(\mathbf{Q})}{\phi(X_1(11)(\mathbf{Q}))} \simeq \mathbf{Z}/5\mathbf{Z}, \qquad \frac{X_1(11)(\mathbf{Q})}{\hat{\phi}(X_0(11)(\mathbf{Q}))} \simeq 1$$

from which we conclude in the usual way that both $X_1(11)$ and $X_0(11)$ have rank zero. If we compare this method to that of a descent by 2-isogeny, we should acknowledge that there was a price to pay in the amount of theoretical background required, but the benefits are tremendous. First, we did the entire argument by hand, without need for computational assistance. Second, the method has promise for generalisation to other modular curves, if we manage to eliminate steps that make reference to the explicit equations.

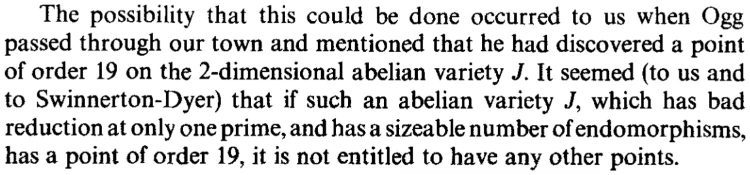
Remark. It is good to remind ourselves that any argument that is omitted is shorter than an argument that is not. The reader who is under the impression that the above 5-descent is more laborious than a 2-descent would have been, is invited to work through all the details of said 2-descent.

2.3 Torsion points of order 13

For the modular curve $X_1(11)$ we now proved twice that the set of rational point consists entirely of the 5 rational cusps, using a 2-descent and a 5-descent. The second proof is more amenable to generalisation, though clearly there are some obstacles to overcome. To get us a little closer to the general argument, we will use similar techniques to prove the following theorem, due to Mazur–Tate [MT73].

Theorem 5. *There are no elliptic curves $E_{/\mathbf{Q}}$ with a rational point of order* 13.

To prove this theorem, we will exploit the existence of a rational point of order 19 on the Jacobian $A := J_1(13)$ of $X_1(13)$. This was originally proved by Ogg [Ogg71], using techniques similar to the explicit relations constructed from Siegel units, which we used to make the table at the end of § 2.1. The announcement of this result was the impetus for the work of Mazur–Tate [MT73], who write:



> The possibility that this could be done occurred to us when Ogg passed through our town and mentioned that he had discovered a point of order 19 on the 2-dimensional abelian variety *J*. It seemed (to us and to Swinnerton-Dyer) that if such an abelian variety *J*, which has bad reduction at only one prime, and has a sizeable number of endomorphisms, has a point of order 19, it is not entitled to have any other points.

The modular curve $X_1(13)$

We begin by recalling a number of facts about the arithmetic and geometry of the modular curve $X_1(13)$ and its Jacobian $A := J_1(13)$, most of which were discussed in § 2.1. The curve $X_1(13)$ has genus 2 and a model over $\mathbf{Q}$ given by

$$X_1(13) : y^2 + (x^3 + x + 1)y = x^5 + x^4.$$

It has a total of 12 cusps, six of which are rational over $\mathbf{Q}$, and six of which are defined over $\mathbf{Q}(\zeta_{13})^+$, the maximal real subfield of the cyclotomic field $\mathbf{Q}(\zeta_{13})^+$. Our goal is to show that the set $X_1(13)(\mathbf{Q})$ consists precisely of the six rational cusps, and nothing else. The argument proceeds in two steps:

1. Use a 19-descent to show that $A(\mathbf{Q}) \simeq \mathbf{Z}/19\mathbf{Z}$, i.e. A has rank zero,
2. Show that this implies that $X_1(13)(\mathbf{Q}) = \{\text{six rational cusps}\}$.

The automorphism group

To carry out the 19-descent, we first want to understand the structure of the 19-torsion on A. To help us do this, we break up this module into smaller pieces, exploiting the structure of the endomorphism ring. The curve $X_1(13)$ has many automorphisms, induced by the maps on enhanced elliptic curves (E, p)

$$\begin{array}{lclcllll} \langle a\rangle & : & (E,p) & \longmapsto & (E, ap) & a \in (\mathbf{Z}/13\mathbf{Z})^\times & \in \mathrm{Aut}_{\mathbf{Q}}(X_1(13)) \\ w_\zeta & : & (E,p) & \longmapsto & (E/\langle p\rangle, q) & \langle p, \widetilde{q}\rangle_{\text{Weil}} = \zeta \in \mu_{13} & \in \mathrm{Aut}_{\mathbf{Q}(\zeta_{13})^+}(X_1(13)) \end{array}$$

Here, the notation $\widetilde{q}$ is used for any point on E whose image on $E/\langle p\rangle$ is equal to q. Note that the condition on the Weil pairing determines $\widetilde{q}$ up to multiples of p, which makes its image well-defined.

The elements $\langle a\rangle$, which are called *diamond operators*, form a group Γ which is isomorphic to $(\mathbf{Z}/13\mathbf{Z})^\times/\pm 1$ and is generated by the element $\langle 2\rangle$ of order 6. The elements w_ζ are involutions, indexed by the elements of $\mu_{13} - \{1\}$ modulo the action of inversion, i.e. there are six of these involutions. They all lift the Atkin–Lehner involution w_{13} on $X_0(13)$. We check that these automorphisms satisfy the following relations:

$$\begin{cases} w_\zeta \,\langle a\rangle\, w_\zeta &= \langle a^{-1}\rangle \\ \langle a\rangle\, w_\zeta &= w_{\zeta^a} \end{cases}$$

so that we exhibited two groups of automorphisms

$$\begin{aligned} \Gamma &\leq \operatorname{Aut}_{\mathbf{Q}} X_1(13) \\ \Delta := \Gamma \rtimes C_2 &\leq \operatorname{Aut}_{\overline{\mathbf{Q}}} X_1(13) \end{aligned}$$

The group Γ is cyclic of order 6, and the group Δ is dihedral of order 12. In fact, these constitute the full automorphism groups over $\mathbf{Q}$ and $\overline{\mathbf{Q}}$ respectively, though we shall not use this fact. Explicitly, by the equivariance of the Weil pairing, the action of an element of the Galois group $g \in G_{\mathbf{Q}}$ satisfies

$$w_\zeta^g = w_{\zeta^g} = \langle \chi_{cyc}(g)\rangle\, w_\zeta,$$

where $\chi_{cyc}(g)$ is the cyclotomic character, i.e. the image of g under

$$G_{\mathbf{Q}} \longrightarrow \operatorname{Gal}(\mathbf{Q}(\zeta_{13})^+/\mathbf{Q}) \simeq (\mathbf{Z}/13\mathbf{Z})^\times/\{\pm 1\}.$$

The structure of the 19-torsion

The argument of Mazur–Tate centers around the structure of the Galois module $V := J_1(13)[19]$ of 19-torsion points on the Jacobian of $X_1(13)$, for which we use the shorthand $A := J_1(13)$. We will use the endomorphism algebra of A to break up V into smaller pieces that are easier to understand. Note that we constructed an automorphism of order 6 of $X_1(13)$ defined over $\mathbf{Q}$, so that we have

$$\mathbf{Z}[\zeta_6] = \mathbf{Z}\left[\frac{1+\sqrt{-3}}{2}\right] \subset \operatorname{End}_{\mathbf{Q}}(A)$$

The ring $\mathbf{Z}[\zeta_6]$ is a principal ideal domain, and the prime 19 splits into two prime elements $19 = \pi \cdot \overline{\pi}$, each of which can be seen as a rational endomorphism of A. Taking the induced endomorphism on the Néron model $\mathscr{A}$ over $\mathbf{Z}[1/13]$, which is an abelian variety, we obtain the short exact sequence

$$0 \to \mathscr{A}[\pi] \longrightarrow \mathscr{A} \xrightarrow{\pi} \mathscr{A} \to 0$$

and likewise for $\overline{\pi}$. Both $\mathscr{A}[\pi]$ and $\mathscr{A}[\overline{\pi}]$ are rank two, and we have $\mathscr{A}[19] = \mathscr{A}[\pi] \oplus \mathscr{A}[\overline{\pi}]$. The rational point $T \in A(\mathbf{Q})$ of order 19 maps injectively into $\mathscr{A}(\mathbf{F}_\ell)$ for every prime $\ell \neq 13$, including 19. To see why, note that its closure is a finite flat group scheme which agrees with the constant group scheme $\mathbf{Z}/19\mathbf{Z}$ over $\mathbf{Q}$, so it must be isomorphic to it over $\mathbf{Z}[1/13]$ by the classification theorem of Oort–Tate [TO70]. Changing the roles of π and $\overline{\pi}$ if necessary, we can assume that

$$\mathbf{Z}/19\mathbf{Z} \subset \mathscr{A}[\overline{\pi}].$$

The module $\mathscr{A}[\pi]$ will now be filtered into two pieces whose flat cohomology groups we can compute using Kummer theory, as we did in the simpler case of $X_1(11)$. To this end, we define the subgroup

$$\mathscr{L} := w_\zeta(\mathbf{Z}/19\mathbf{Z}) \subset \mathscr{A}[\pi]$$

the image of the rational torsion on A under the involution w_ζ. This group is independent of the choice of $\zeta \in \mu_{13}$, since changing the root of unity amount to translating the torsion $\mathbf{Z}/19\mathbf{Z}$ with an element of Γ, which is an endomorphism defined over $\mathbf{Q}$ so it preserves the torsion. There is an exact sequence

$$0 \longrightarrow \mathscr{L} \longrightarrow \mathscr{A}[\pi] \xrightarrow{\langle T, \cdot \rangle_{\mathrm{Weil}}} \mu_{19} \longrightarrow 1 \tag{2.4}$$

of finite flat group schemes over $\mathbf{Z}[1/13]$. This requires justification, given in two steps:

- the modules $\mathscr{A}[\pi]$ and $\mathscr{A}[\overline{\pi}]$ are self-orthogonal with respect to the Weil pairing, so that the Weil pairing with a rational point T in $A(\mathbf{Q})$ of order 19 is necessarily a surjection onto μ_{19},
- the Weil pairing with T kills the subgroup $\mathscr{L}$, since $\mathscr{L}$ is not isomorphic to μ_{19} and there can be at most one such submodule of $\mathscr{A}[\pi]$, which must be the kernel of the Weil pairing by the first point.

We note for future reference that the group scheme $\mathscr{L}$ becomes trivial over the extension $\mathbf{Q}(\zeta_{13})^+$.

The 19-descent

We are now ready to perform a 19-descent on A, to show that $A(\mathbf{Q})/\pi A(\mathbf{Q}) = 0$ and conclude from there that the rank is zero. For the Néron model $\mathscr{A}$ of A over $\mathbf{Z}[1/13]$, we have the short exact sequence

$$0 \longrightarrow \mathscr{A}[\pi] \longrightarrow \mathscr{A} \xrightarrow{\pi} \mathscr{A} \longrightarrow 0 \qquad \text{over } \mathbf{Z}[1/13].$$

The long exact sequence in flat cohomology over both $\mathbf{Z}[1/13]$ and $\mathbf{Q}_{13}$ gives rise to a commutative diagram with exact rows, reminiscent of our arguments in the case of 11-torsion, which in this case is

$$\begin{tikzcd}
A(\mathbf{Q}) \arrow[r] \arrow[d] & A(\mathbf{Q}) \arrow[r] \arrow[d] & \mathrm{H}^1_{\mathrm{fppf}}\big(\mathbf{Z}[1/13], \mathscr{A}[\pi]\big) \arrow[d, "(**)"] \\
A(\mathbf{Q}_{13}) \arrow[r, "(*)"] & A(\mathbf{Q}_{13}) \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}\big(\mathbf{Q}_{13}, \mathscr{A}[\pi]\big)
\end{tikzcd}$$

As before, we will argue that $(*)$ is surjective, and $(**)$ is injective, using very similar arguments.

- The surjectivity of $(*)$ follows by a very similar argument. Namely, we use the commutative diagram whose rows are the short exact sequences that furnish the filtration on the local points of the Néron model, given by

$$\begin{tikzcd}
0 \arrow[r] & M \arrow[r] \arrow[d, "\pi"] & \mathscr{A}(\mathbf{Z}_{13}) \arrow[r] \arrow[d, "(*)"] & \mathscr{A}(\mathbf{F}_{13}) \arrow[r] \arrow[d, "\pi"] & 0 \\
0 \arrow[r] & M \arrow[r] & \mathscr{A}(\mathbf{Z}_{13}) \arrow[r] & \mathscr{A}(\mathbf{F}_{13}) \arrow[r] & 0
\end{tikzcd}$$

  The map induced on M is a factor of multiplication by 19, which must therefore be an isomorphism on the pro-13 group M. To show that the map $\pi : \mathscr{A}(\mathbf{F}_{13}) \to \mathscr{A}(\mathbf{F}_{13})$ is surjective, it suffices to show that it is injective, since it is an endomorphism of a finite module. By the snake lemma, it therefore suffices that the map $(*)$ itself, namely $\pi : \mathscr{A}(\mathbf{Z}_{13}) \to \mathscr{A}(\mathbf{Z}_{13})$, is *injective*. This is true, since the kernel of this morphism is precisely

$$\operatorname{Ker}\left(A(\mathbf{Q}_{13}) \longrightarrow A(\mathbf{Q}_{13})\right) = A[\pi]^{D_{13}}$$

where $D_{13} \leq G_{\mathbf{Q}}$ is a decomposition group above the prime 13. This is shown because taking D_{13}-invariants yields a short exact sequence of modules

$$1 \longrightarrow \mathscr{L}^{D_{13}} \longrightarrow A[\pi]^{D_{13}} \longrightarrow \mu_{19}^{D_{13}}$$

and we can argue that both flanking terms must vanish. Indeed, for $\mathscr{L}^{D_{13}}$ we note that it cannot be isomorphic to $\mathbf{F}_{19}$ since this would imply that 13 splits completely in the cyclotomic field $\mathbf{Q}(\zeta_{13})$, over which $\mathscr{L}$ becomes trivial. This is not the case, as 13 is totally ramified. Similarly, 13 does not split completely in the cyclotomic field $\mathbf{Q}(\zeta_{19})$, so that μ_{19} has trivial D_{13}-invariants.

- The injectivity of $(**)$ is slightly more involved, and exploits the filtration (2.4) of $\mathscr{A}[\pi]$. From the Hochschild–Serre spectral sequence applied to the finite étale extension

  $$\mathbf{Z}[1/13] \longrightarrow \mathbf{Z}[1/13, \zeta_{13}]$$

  with Galois group $G := \operatorname{Gal}(\mathbf{Q}(\zeta_{13})/\mathbf{Q}) \simeq (\mathbf{Z}/13\mathbf{Z})^{\times}$ we obtain an exact sequence

  $$0 \longrightarrow \mathrm{H}^1(G, \mathbf{Z}/19\mathbf{Z}) \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/13], \overline{\mathscr{L}}) \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/13, \zeta_{13}], \mathbf{Z}/19\mathbf{Z})^{G}.$$

  Both of the flanking terms are easily seen to be trivial:

  - Since the action of G is trivial, the group $\mathrm{H}^1(G, \mathbf{Z}/19\mathbf{Z})$ is equal to

    $$\operatorname{Hom}(\mathbf{Z}/13\mathbf{Z}, \mathbf{Z}/19\mathbf{Z}) = 0.$$

  - Since $\mathbf{Q}(\zeta_{13})$ has no $\mathbf{Z}/19\mathbf{Z}$-extensions that are unramified at all primes $\ell \neq 13$, we have

    $$\mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/13, \zeta_{13}], \mathbf{Z}/19\mathbf{Z}) = 1.$$

  We conclude from the filtration (2.4) that there is a commutative diagram

  $$\begin{tikzcd}
  0 \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Z}[1/13], \mathscr{A}[\pi]) \arrow[r] \arrow[d, "(**)"] & \mathrm{H}^1_{\mathrm{fppf}}\left(\mathbf{Z}[1/13], \mu_{19}\right) \arrow[d] \\
  & \mathrm{H}^1_{\mathrm{fppf}}(\mathbf{Q}_{13}, \mathscr{A}[\pi]) \arrow[r] & \mathrm{H}^1_{\mathrm{fppf}}\left(\mathbf{Q}_{13}, \mu_{19}\right)
  \end{tikzcd}$$

  whose upper horizontal arrow is an *injection*, by the triviality of the flat cohomology of $\mathscr{L}$ we just proved. To show that $(**)$ is injective, it is therefore enough to show that the right vertical arrow is injective. This follows by essentially the same argument we saw before. Namely, this map has a very concrete description by Kummer theory; it is the natural map induced by inclusion:

  $$\pm 13^{\mathbf{Z}} / \pm 13^{19\,\mathbf{Z}} \longrightarrow \mathbf{Q}_{13}^{\times} / (\mathbf{Q}_{13}^{\times})^{19}.$$

  This map is clearly injective, since 13 is not a nineteenth power in $\mathbf{Q}_{13}^{\times}$.

From the descent argument, we therefore conclude the triviality of the quotient

$$A(\mathbf{Q})/\pi(A(\mathbf{Q})) = 0.$$

Note that $A(\mathbf{Q})$ is a module over $\operatorname{End}_{\mathbf{Q}}(A) \supset \mathbf{Z}[\zeta_6]$ which is a principal ideal domain. The structure theorem of finitely generated modules implies $A(\mathbf{Q})$ is a torsion group. It follows that $A(\mathbf{Q}) = \mathbf{Z}/19\mathbf{Z}$.

The rational points on $X_1(13)$

Finally, we need to conclude that the set of rational points on $X_1(13)$ is fully accounted for by the 6 rational cuspidal points, which generate the Mordell–Weil group of A. We have two options here:

1. In the course of Samuel Le Fourn, we have learned about the method of Chabauty. This curve is a perfect example to try out this method for yourself. It has genus two, and a plane model

   $$X_1(13) : y^2 + (x^3 + x + 1)y = x^5 + x^4.$$

   We proved that the rank of its Jacobian is equal to $r = 0 < g = 2$. Since this is not the topic of this mini-course, it is an exercise to choose your favourite prime of good reduction (for instance, why not $p = 2$?), write down a basis of holomorphic differentials, and apply the method of Chabauty.

2. We may instead opt for a more "modular" approach that does not rely on explicit equations, but rather uses the moduli interpretation of $Y_1(13)$ as classifying elliptic curves (E, p) with a rational point of order 13. Let $\mathscr{E}$ be its Néron model, then reduction modulo 3 gives a sequence

   $$1 \longrightarrow \mathbf{Z}_3 \longrightarrow \mathscr{E}(\mathbf{Z}_3) \longrightarrow \mathscr{E}(\mathbf{F}_3) \longrightarrow 1$$

   so that $\mathscr{E}(\mathbf{F}_3)$ must have a point of order 13. If E has good or additive reduction modulo 2, then this is impossible, since $\mathscr{E}(\mathbf{F}_3)$ is an extension of a cyclic group of order 3 by a finite group of order at most 4, by Tate's algorithm. Therefore E must have semi-stable reduction.

   The point $x \in X_1(13)(\mathbf{Q})$ corresponding to (E, p) must therefore reduce to a cusp modulo 3. This must be one of the 6 rational cusps, since the non-rational cusps reduce to points over $\mathbf{F}_{27}$, not $\mathbf{F}_3$. Since the group $\Gamma \subset \operatorname{Aut}_{\mathbf{Q}}(X_1(13))$ acts simply transitively on the rational cusps, we may assume without loss of generality that x reduces to the cusp ∞, and therefore the element

   $$(x) - (\infty) \in A(\mathbf{Q})$$

   is contained in the pro-3 group $A^1(\mathbf{Q}_3)$, which is torsion free since $\mathbf{Q}_3$ has ramification index $e = 1 < 3 - 1$. At the same time, it defines a point in $A(\mathbf{Q})$, which we already showed is torsion. Both statements can only be true if the divisor $(x) - (\infty)$ is trivial, i.e. if $x = \infty$.

2.4 Torsion points of order $\ell \geq 17$

The general theorem of Mazur [Maz78] extends these arguments considerably. It is instructive to continue in the same way we have done, adding an additional technical layer of the proof at every step. We will merely point out one important key step here, and invite the reader to continue the story with $\ell = 17, 19, \ldots$.

For instance, for $\ell = 17$ we would be facing the task of determining the rational points on the modular curve $X_1(17)$, which has genus 5. Its Jacobian factors up to $\mathbf{Q}$-isogeny as

$$J_1(17) \sim J_0(17) \times A, \qquad \text{where } X_0(17) \ : \ y^2 + xy + y = x^3 - x^2 - x - 14$$

where A is a simple abelian variety of dimension 4. The first step is to prove that the Jacobian $J_1(17)$ has rank zero. If we were to carry out the above strategy most faithfully, we would be tempted to try and use a 2-descent for $X_0(17)$ and a 73-descent for the abelian variety A. The reader is invited to do this.

On the other hand, we could observe that we do not really need to deal with the abelian 4-fold A at all. It suffices to show that $X_0(17)$ has rank zero, using essentially the same argument as we saw above. Indeed, any rational point $x \in Y_1(17)$ must reduce to a cusp modulo 3. Its image $y \in Y_0(17)$ has the same property, and we may assume without loss of generality that it is ∞. We conclude just as before that $y = \infty$, and therefore no elliptic curve over $\mathbf{Q}$ can have a rational point of order 17.

Continuing this strategy, we need one more key ingredient: we can of course not expect the rank of the modular Jacobians $J_0(\ell)$ to be zero when ℓ is a large prime. For instance, $J_0(37)$ is isogenous to a product of two elliptic curves, one of which has rank 1. Therefore, the best one can hope to find in general is a quotient of genus zero. Mazur finds the so-called *Eisenstein quotient*

$$J_0(\ell) \longrightarrow J_{\text{eis}}$$

which is small enough to have rank zero, but large enough to remember enough about the curve $X_0(\ell)$. More precisely, for a scheme S we say that a morphism of S-schemes $f : X \longrightarrow Y$ is a formal immersion at $x \in X(S)$ if the induced map on complete local rings

$$f^* : \widehat{\mathscr{O}}_{Y,f(x)} \longrightarrow \widehat{\mathscr{O}}_{X,x}$$

is surjective. In fact, the argument above is easily extended to show that the only rational points on $X_1(\ell)$ are cusps, as long as one finds a rank zero quotient $J_0(p) \longrightarrow A$ for which the map $\mathscr{X}_0(p) \longrightarrow \mathscr{A}$ induced by the Abel–Jacobi map defined by the cusp ∞ is a formal immersion at ∞.

The difficult task that Mazur faced is to show that the Eisenstein quotient $A = J_{\text{eis}}$ satisfies these properties. To show that it has rank zero, he proceeds using a descent argument similar to what we encountered in our small examples, using the *Shimura subgroup* of $J_0(\ell)$ generated by $(0) - (\infty)$, which has order

$$n = \text{Numerator}\left(\frac{p-1}{12}\right).$$

Mazur performs an ℓ-descent, for $\ell \mid n$, to show that the rank of the Eisenstein quotient is zero. The general arguments are more sophisticated than they were in our small examples, but after our brief foray into special cases, our hope is that the reader will feel more confident reading the papers of Mazur [Maz77a, Maz78].

Chapter 3

CM points and the Birch–Swinnerton-Dyer conjecture

We discuss how a study of the arithmetic properties of Heegner points leads to a proof of the Birch–Swinnerton-Dyer conjecture for mdoular abelian varieties over $\mathbf{Q}$ of analytic rank at most one, by combining the works of Gross–Zagier [GZ85, GZ86] and Kolyvagin [Kol89]. The proofs of these works would deserve entire lecture courses by themselves, and indeed Kolyvagin's proof was discussed in detail by Castella at the 2021 ICTS Workshop "Elliptic curves and the special values of L-functions" [ICT21].

Since a large audience overlap between this year's ICTS workshop and the 2021 ICTS workshop [ICT21] is to be expected, these notes will not go into the mechanics of the subtle descent argument in the proof of Kolyvagin. Instead, we will emphasise the work of Gross–Zagier, and the consequences of Kolyvagin's work in the context of the previous chapter, i.e. for the construction of the 'winding quotient'.

3.1 The theory of complex multiplication

After our first source of special points provided by cusps, we will study a second source provided by points of *complex multiplication* (CM). The study of their fields of rationality and arithmetic properties goes back to the works of Eisenstein and Kronecker. We briefly recall a few of the basic points of the classical theory following Borel et. al. [BCH$^+$57], and refer also to the beautiful treatment of Weil [Wei76].

Singular moduli

The central actors in the theory of complex multiplication are the *singular moduli*, which are the special values at imaginary quadratic arguments τ in the Poincaré–Lobachevsky upper half plane

$$\mathfrak{H} := \{\tau \in \mathbf{C} \ : \ \mathrm{Im}(\tau) > 0\} = \{q \in \mathbf{C} \ : \ |q| < 1\}$$

with $q = \exp(2\pi i \tau)$, of the modular j-invariant

$$\begin{aligned} j(q) \quad &= \quad \Big(1 + 240 \sum_{n \geq 1} \frac{n^3 q^n}{1 - q^n}\Big)^3 \div \Big(q \prod_{n \geq 1} (1 - q^n)^{24}\Big) \qquad \in q^{-1}\,\mathbf{Z}[\![q]\!] \\ &= \quad q^{-1} + 744 + 196884q + 21493760q^2 + \ldots \end{aligned}$$

which has *integral* Fourier expansion. The j-invariant is holomorphic on $\mathfrak{H}$ with a simple pole at the cusp ∞. It is invariant under the action of $\mathrm{SL}_2(\mathbf{Z})$ by linear fractional transformations on the argument τ. To study the modular polynomials that it gives rise to, we will choose an auxiliary prime p and let

$$M_p \ := \ \mathrm{SL}_2(\mathbf{Z}) \backslash \{\gamma \in \mathrm{Mat}_{2\times 2}(\mathbf{Z}) \ : \ \det(\gamma) = p\} \quad = \bigsqcup_{j \in \{0, \ldots, p-1, \infty\}} \mathrm{SL}_2(\mathbf{Z})\gamma_i$$

where we fix the convenient collection of $p+1$ left coset representatives given by the matrices

$$\begin{cases} \gamma_j & := \left(\begin{smallmatrix} 1 & i \\ & p \end{smallmatrix}\right) \qquad \text{with } j = 0, 1, \ldots, p-1 \\ \gamma_\infty & := \left(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\right) \end{cases}$$

Theorem 6. *There exists a unique polynomial $\Phi_p(X,Y) \in \mathbf{Z}[X,Y]$ such that*

$$\Phi_p(X, j(\tau)) = \prod_{\gamma \in M_p} (X - j(\gamma\tau)). \tag{3.1}$$

It satisfies $\Phi_p(X,Y) = \Phi_p(Y,X)$ and the leading coefficient of the polynomial $\Phi_p(X,X)$ is equal to ± 1.

Proof. Let $P(X)$ be the polynomial on the right hand side of (3.1). Its coefficients are holomorphic functions on $\mathfrak{H}$ which are elementary symmetric polynomials in the set $j(\gamma\tau)$ with $\gamma \in M_p$. The action of an element in $\mathrm{SL}_2(\mathbf{Z})$ permutes this set, and thus preserves the coefficients of $P(X)$. These coefficients have fractional q-expansions, which by their invariance under translation must therefore be meromorphic at ∞. It follows that all coefficients of $P(X)$ lie in $\mathbf{C}[j]$. Note that

$$\exp\left(2\pi i \left(\frac{\tau + j}{p}\right)\right) = \zeta_p^j q^{1/p},$$

so as elements of $\mathbf{Z}[\zeta_p][\![q]\!]$ the coefficients of $P(x)$ are invariant under the Galois action of $\mathbf{Q}(\zeta_p)/\,\mathbf{Q}$, which permutes the set on which the coefficients are elementary symmetric polynomials. It follows that all coefficients of $P(X)$ lie in $\mathbf{Z}[j]$, i.e. are polynomials in $\mathbf{Z}[Y]$ evaluated at j. The first part of the theorem follows, substituting the coefficients of $P(X)$ with these polynomials.

The final statement follows from the observation that the leading term of the fractional q-expansion of $j(\tau) - j(\gamma\tau)$ is a root of unity in $\mathbf{Z}[\zeta_p]$. Therefore the same must hold for the product of all these expressions over $\gamma \in M_p$, so that the leading term of $\Phi_p(X,X)$ is a root of unity, and is contained in $\mathbf{Z}$, implying the statement. For the symmetry, which we will not use, we refer to [BCH$^+$57, III.4]. □

The richness of singular moduli is reflected in the gargantuan size of the coefficients of $\Phi_p(X,Y)$, whose representation requires $O(p^3 \log(p))$ bits. The slow history of their computation speaks to the difficulties this caused. The polynomial Φ_2 is classical and was known to Jacobi, Kronecker, and Joubert; it is

$$\begin{aligned} \Phi_2(X,Y) \quad = \quad & X^3 + Y^3 - 162000(X^2 + Y^2) + 1488XY(X+Y) - X^2Y^2 \\ & + 8748000000(X+Y) + 40773375XY - 157464000000000 \end{aligned}$$

Some notable landmarks on determining modular polynomials were

- Φ_3 was calculated by Smith [Smi78] in 1878,
- Φ_5 was calculated by Berwick [Ber16] in 1916,
- Φ_7 was computed by Herrmann [Her75] using a Siemens 2002 calculator in 1975,
- Φ_{11} was computed by Kaltofen–Yui [KY84] computed Φ_{11} using a VAX 11-780 in 1984.

Since then, there have been many spectacular computational advances, and the reader is invited to inspect the tables of modular polynomials for $p < 1000$ on the webpage [Sut], which was computed using the algorithm of Bröker–Lauter–Sutherland [BLS12]. It is not helpful to reproduce further examples of $\Phi_p(X,Y)$ in these notes, though the size of the restrictions to the diagonal $X = Y$ is more manageable, when suitably factorised. The first few examples are:

$$\begin{aligned}
\Phi_2(X,X) &= (X - 2^6 \cdot 5^3)(X + 3^3 \cdot 5^3)^2(X - 2^6 \cdot 3^3) \\
\Phi_3(X,X) &= (X - 2^6 \cdot 5^3)^2(X - 2^4 \cdot 3^3 \cdot 5^3)(X + 2^{15})^2 X \\
\Phi_5(X,X) &= (X - 2^3 \cdot 3^3 \cdot 11^3)^2(X - 2^6 \cdot 3^3)^2(X + 2^{15})^2(X + 2^{15} \cdot 3^3)^2(X^2 - 1264000X - 2^{12} \cdot 5^3 \cdot 11^3)
\end{aligned}$$

The factors of these polynomials are called *Hilbert class polynomials*, and they are of great significance. Let $\mathcal{O}$ be an imaginary quadratic order such that $p\,\mathcal{O} = \mathfrak{p}\overline{\mathfrak{p}}$ with principal prime ideals $\mathfrak{p}$ and $\overline{\mathfrak{p}}$. Note that for a fixed order $\mathcal{O}$, there is a positive density set of such primes p. In this case, we have for any invertible fractional $\mathcal{O}$-ideal $\mathfrak{a}$ that $\mathfrak{p}\mathfrak{a} \subset \mathfrak{a}$ is of index p, and $j(\mathfrak{p}\mathfrak{a}) = j(\mathfrak{a})$ since $\mathfrak{p}$ is principal. It follows that $j(\mathfrak{a})$ is a root of $\Phi_p(X,X)$ and is therefore an algebraic integer. Here are some examples:

$$\begin{array}{rclrclrcl}
j(\sqrt{-1}) &=& 2^6 \cdot 3^3 & j(\sqrt{-2}) &=& 2^6 \cdot 5^3 & j(\sqrt{-3}) &=& 2^4 \cdot 3^3 \cdot 5^3 \\
j\left(\frac{1+\sqrt{-3}}{2}\right) &=& 0 & j\left(\frac{1+\sqrt{-7}}{2}\right) &=& -3^3 \cdot 5^3 & j\left(\frac{1+\sqrt{-11}}{2}\right) &=& -2^{15}.
\end{array}$$

The singular moduli in these examples are all integers, and we observe that they are roots of several of the polynomials listed above, all for the reasons we stated. The singular modulus

$$j(\sqrt{-5}) = 282880\sqrt{5} + 632000$$

is an algebraic integer of degree two, and it is a root of the irreducible quadratic factor of $\Phi_5(X,X)$ listed above. To understand the field of definition of singular moduli in general, we need a refinement of the following classical result, known as the *Kronecker congruence.*

Theorem 7 (Kronecker)**.** *For any prime p, we have* $\Phi_p(X,Y) \equiv (X^p - Y)(X - Y^p) \pmod{p}$.

Proof. The prime p is totally ramified in the cyclotomic ring $\mathbf{Z}[\zeta_p]$, and the unique prime above it is principal generated by $(1 - \zeta_p)$. We have the following obvious congruence in $\mathbf{Z}[\zeta_p]$:

$$\exp\left(2\pi i\left(\frac{\tau + j}{p}\right)\right) = \zeta_p q^{1/p} \equiv q^{1/p} \bmod (1 - \zeta_p)\,\mathbf{Z}[\zeta_p]$$

from which it follows immediately that

$$\begin{aligned}
\Phi_p(X,j) &\equiv (X - j(q^{1/p}))^p(X - j(q^p)) & \bmod p\,\mathbf{Z}[\![q]\!][X] \\
&\equiv (X^p - j(q))(X - j(q)^p) & \bmod p\,\mathbf{Z}[\![q]\!][X]
\end{aligned}$$

This implies the desired congruence for the modular polynomial. □

Let $\mathfrak{a}$ be an invertible fractional ideal of the imaginary quadratic order $\mathcal{O}$, so that the singular modulus $j(\mathfrak{a})$ is an algebraic integer. The key to determining its field of definition will lie in a suitable refinement of the Kronecker congruence. We now know that for any $p\,\mathcal{O} = \mathfrak{p}\overline{\mathfrak{p}}$ we have

$$\big(j(\mathfrak{a})^p - j(\mathfrak{a}\mathfrak{p})\big) \cdot \big(j(\mathfrak{p}\mathfrak{a})^p - j(\mathfrak{a})\big) \quad \bmod p. \tag{3.2}$$

Thus any prime above p in the smallest field of definition of these singular moduli must divide one of the two factors in (3.2). However, we cannot yet say *which* of the two factors it divides. To do this, we make a brief study of some *elliptic units*, our goal being to prove that the first factor in (3.2) is divisible by $\overline{\mathfrak{p}}$.

Elliptic units

Elliptic units are CM values of the Siegel units defined in § 2.1. We will consider here only the special case of the eta quotients defined by

$$h_\gamma(\tau) := \det(\gamma)^{12}(c\tau+d)^{-12}\frac{\Delta(\gamma\tau)}{\Delta(\tau)}, \qquad \text{where} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_p$$

When $p = 2$, these are essentially (the 24-th power of) the classical *Weber functions*, or the Ramanujan G- and g-functions, whose special values at CM points play a crucial role in Heegner's proof of the class number one problem.

Theorem 8. *There exists a unique polynomial* $\Upsilon_p(X,Y) \in \mathbf{Z}[X,Y]$ *such that*

$$\Upsilon_p(X,j) = \prod_{\gamma\in M_p}(X - h_\gamma).$$

It satisfies $\Upsilon_p(0,Y) = p^{12}$.

Proof. The first part is an exercise. For the second part, computing the constant coefficient shows

$$\Upsilon_p(0,Y) = (-1)^p \prod h_\gamma = (-1)^{p+1}\cdot p^{12}\prod_{j=0}^{p-1}\zeta_p^j + O(q) = p^{12} + O(q).$$

Note at the same time that this product is $\mathrm{SL}_2(\mathbf{Z})$-invariant and holomorphic everywhere (including at infinity) and must therefore be constant. The second part of the theorem follows. □

Unlike the modular polynomials, the polynomials Υ_p are of much lower height, reflecting the poorer and more straightforward arithmetic encoded in elliptic units. The first few examples[1] are:

$$\begin{aligned}
\Upsilon_2(X,Y) &= (X+2^4)^3 - XY \\
\Upsilon_3(X,Y) &= (X-3^2)^3(X-3^6) + 2^3\cdot 3^2X(X+21)Y - XY^2 \\
\Upsilon_5(X,Y) &= (X^2-8050X+5^4)^3 + 2^5\cdot 5^2X(X+25)(47X^2+269650X+29375)Y \\
&\quad - 2^2\cdot 5X(207X^2-254750X+129375)Y^2 + 2^3\cdot 3\cdot 5X(X+25)Y^3 - XY^4.
\end{aligned}$$

As before, we see that for any invertible fractional $\mathcal{O}$-ideal $\mathfrak{a}$, we have that $h_\gamma(\mathfrak{a})$ is an algebraic integer whenever $p\,\mathcal{O} = \mathfrak{p}\overline{\mathfrak{p}}$ is regular. Unlike singular moduli, the factorisation of these elliptic units is rather poor and trivial outside p, by the second statement in the above theorem. The following theorem makes this more precise, determining their principal ideals in the ring of algebraic integers

$$\langle h_\gamma(\mathfrak{a})\rangle \lhd \overline{\mathbf{Z}}.$$

We fix some notation first. Pick any $\mathbf{Z}$-bases for the ideals $\mathfrak{a}$ and $\mathfrak{p}\mathfrak{a}$, then these bases are related by an integral matrix of determinant p, whose left coset representative $\gamma(\mathfrak{p}) \in M_p$ is independent of the chosen bases. Likewise, define the matrix $\gamma(\overline{\mathfrak{p}}) \in M_p$.

[1] I would like to thank Jenny Roberts, Sven Cats, and Andrei Seymour-Howell for providing me with these examples.

Theorem 9. *Suppose $p\,\mathcal{O} = \mathfrak{p}\overline{\mathfrak{p}}$ regular, then*

$$\langle h_{\gamma(\mathfrak{p})}(\mathfrak{a})\rangle = \overline{\mathfrak{p}}^{12}, \qquad \langle h_{\gamma(\overline{\mathfrak{p}})}(\mathfrak{a})\rangle = \mathfrak{p}^{12}$$

and for any $\gamma \in M_p$ different from $\gamma(\mathfrak{p}), \gamma(\overline{\mathfrak{p}})$ we have $\langle h_\gamma(\mathfrak{a})\rangle = (1)$.

Proof. Let $f \geq 1$ be the smallest integer such that $\mathfrak{p}^f = (\alpha)$ in $\mathcal{O}$, then we have

$$\left\langle \underbrace{\left(p^{12}\frac{\Delta(\mathfrak{p}^f\mathfrak{a})}{\Delta(\mathfrak{p}^{f-1}\mathfrak{a})}\right)}_{:=\lambda_f(\mathfrak{p})\,\in\,\overline{\mathbf{Z}}}\cdots\underbrace{\left(p^{12}\frac{\Delta(\mathfrak{p}^2\mathfrak{a})}{\Delta(\mathfrak{p}\mathfrak{a})}\right)}_{:=\lambda_2(\mathfrak{p})\,\in\,\overline{\mathbf{Z}}}\cdot\underbrace{\left(p^{12}\frac{\Delta(\mathfrak{p}\mathfrak{a})}{\Delta(\mathfrak{a})}\right)}_{:=\lambda_1(\mathfrak{p})\,\in\,\overline{\mathbf{Z}}}\right\rangle = \langle p^{12f}\alpha^{-12}\rangle = \overline{\mathfrak{p}}^{12f}.$$

Note that the algebraic number $\lambda_i := \lambda_i(\mathfrak{p})$ must divide both $\overline{\mathfrak{p}}^{12f}$ and $(p)^{12}$, and therefore it divides $\overline{\mathfrak{p}}^{12}$. On the other hand, we have $\langle\lambda_1\lambda_2\cdots\lambda_f\rangle = \overline{\mathfrak{p}}^{12f}$, so it follows that $\langle\lambda_i\rangle = \overline{\mathfrak{p}}^{12}$ for all i. The theorem now follows from the observation, proved in Theorem 8, that

$$h_{\gamma(\mathfrak{p})}(\mathfrak{a})\cdot h_{\gamma(\overline{\mathfrak{p}})}\cdot\prod_{\gamma\neq\gamma(\mathfrak{p}),\gamma(\overline{\mathfrak{p}})} h_\gamma(\mathfrak{a}) = (-1)^{p+1}p^{12},$$

since all factors are algebrac integers, and the first two factors are equal to $\lambda_1(\mathfrak{p})$ and $\lambda_1(\overline{\mathfrak{p}})$ respectively, so that their product already generates the ideal $(p)^{12}$, and the other factors must be units. □

Ring class fields

We are now in a situation where we have constructed two sources of algebraic integers, the arithmetically rich *singular moduli*, and on the other hand the *elliptic units* whose factorisation was completely determined and is always a p-unit. This is illustrated in the following example:

$$\begin{array}{lcl|lcl} j(\sqrt{-14}) & = & 2^3\left(323+228\sqrt{2}+(231+161\sqrt{2})\sqrt{2\sqrt{2}-1}\right)^3 & \operatorname{Nm} j(\sqrt{-14}) & = & 2^{24}\cdot 11^6\cdot 17^3\cdot 41^3 \\ h_\gamma(\sqrt{-14}) & = & 2^{-6}(\sqrt{2}+1+\sqrt{2\sqrt{2}-1})^{12} \qquad \text{for } \gamma := \left(\begin{smallmatrix}1&\\&2\end{smallmatrix}\right) & \operatorname{Nm} h_\gamma(\sqrt{-14}) & = & 2^{24} \end{array}$$

This computation was done by Weber [Web98, Section 144]. We are now ready to refine the Kronecker congruence proved in Theorem 7, and to do so we define our third (and last) modular polynomial, whose task will be to connect the collections of singular moduli and elliptic units.

Theorem 10. *There exists a unique polynomial $\Psi_p(X,Y,Z) \in \mathbf{Z}[X,Y,Z]$ such that*

$$\Psi_p(X,Y,j) = \sum_{\gamma\in M_p}(X - j(\gamma\tau))\prod_{\delta\neq\gamma}(Y - h_\delta).$$

It satisfies the congruence $\Psi_p(Z^p,Y,Z) \equiv 0 \pmod{p}$.

Proof. The first part of the theorem is proved using similar methods to what we have previously seen, and is left as an exercise. For the second part, using the congruence

$$\exp\left(2\pi i\left(\frac{\tau+j}{p}\right)\right) = \zeta_p^j q^{1/p} \equiv q^{1/p} \bmod (\zeta_p - 1)$$

we find the following set of congruences of $q^{1/p}$-series

$$\left\{\begin{array}{ccccccccl} j(\gamma_0\tau) & \equiv & j(\gamma_1\tau) & \equiv & \dots & \equiv & j(\gamma_{p-1}\tau) & & \operatorname{mod}(\zeta_p-1)\,\mathbf{Z}[\zeta_p][\![q^{1/p}]\!] \\ h_{\gamma_0} & \equiv & h_{\gamma_1} & \equiv & \dots & \equiv & h_{\gamma_{p-1}} & & \operatorname{mod}(\zeta_p-1)\,\mathbf{Z}[\zeta_p][\![q^{1/p}]\!] \end{array}\right.$$

so that we may conclude from the first part of the theorem that

$$\Psi_p(X,Y,j) \equiv (X-j(q^p))(Y-h_{\gamma_0}).$$

This is a stronger congruence than the required statement, which follows upon setting $X=j^p$. □

To provide just one example, we mention that when $p=2$ we have

$$\Psi_2(X,Y,Z) = Z^3-(Y^2+48Y+2256)Z^2+(1488Y^2+67326Y+1106688-X)Z+3(Y+16)^2(X-54000)$$

The congruence we just proved for the polynomials Ψ_p allows us to sharpen the Kronecker congruence of Theorem 7 using the factorisation we proved in Theorem 9. Many sources are fond of proving this using Deuring's geometric language, but we will give the classical German proof here.

Theorem 11. *Let $\mathcal{O}\subset K$ be an imaginary quadratic order, with $p\,\mathcal{O}=\mathfrak{p}\bar{\mathfrak{p}}$ regular, then*

$$j(\mathfrak{a})^p \equiv j(\mathfrak{a}\bar{\mathfrak{p}}) \pmod{\mathfrak{p}}.$$

Proof. Substituting the triple $(X,Y,Z)=\big(j(\mathfrak{a})^p, h_{\gamma(\mathfrak{p})}, j(\mathfrak{a})\big)$ into the polynomial Ψ_p we find

$$\big(j(\mathfrak{a})^p - j(\mathfrak{p}\mathfrak{a})\big)\cdot \prod_{\gamma\neq\gamma(\bar{\mathfrak{p}})} \underbrace{\big(h_{\gamma(\bar{\mathfrak{p}})}-h_\gamma\big)}_{\not\equiv 0 \pmod{\mathscr{P}}} \equiv 0 \pmod{\mathfrak{p}}.$$

Since all the factors in the product, except possibly the first one, are not contained in $\mathscr{P}$ for any prime $\mathscr{P}\,|\,\mathfrak{p}$ by Theorem 9, the first factor is contained in $\mathscr{P}$ for any such $\mathscr{P}$, so that the theorem follows. □

We are ready to determine the field of rationality of singular moduli, and will show that $j(\mathfrak{a})$ for invertible ideals $\mathfrak{a}\lhd\mathcal{O}$ generates the *ring class field* $K_{\mathcal{O}}$ of the imaginary quadratic order $\mathcal{O}$, defined as the finite abelian extension of K corresponding via global class field theory to the idèlic quotient

$$\operatorname{Pic}(\mathcal{O}) \simeq K^\times\backslash \widehat{K}^\times/\widehat{\mathcal{O}}^\times, \qquad \text{where } \widehat{K}=\widehat{\mathcal{O}}\otimes\mathbf{Q},\ \ \widehat{\mathcal{O}}=\prod_\ell \mathcal{O}\otimes\mathbf{Z}_\ell\,.$$

This extension is also characterised by the fact that it is finite and unramified over K outside the singular primes of the ring $\mathcal{O}$, and that the set of primes p that split completely in $K_{\mathcal{O}}/\,\mathbf{Q}$ is precisely the set of primes for which $p\,\mathcal{O}=\mathfrak{p}\bar{\mathfrak{p}}$ with $\mathfrak{p}$ principal.

Corollary 1. *Suppose $\mathfrak{a}\lhd\mathcal{O}$ is regular. Then $K(j(\mathfrak{a}))$ is the ring class field of $\mathcal{O}$.*

Proof. Let $K_{\mathcal{O}}$ be the ring class field of $\mathcal{O}$. Take p to be any prime for which $p\,\mathcal{O}=\mathfrak{p}\bar{\mathfrak{p}}$, and excluding finitely many primes p if necessary, we have

$$\begin{array}{rcl} p \text{ splits completely in } K_{\mathcal{O}}/\,\mathbf{Q} & \Longleftrightarrow & \mathfrak{p},\bar{\mathfrak{p}} \text{ are principal} \\ & \Longrightarrow & \left\{\begin{array}{lllll} j(\mathfrak{a}) & = & j(\mathfrak{p}\mathfrak{a}) & \equiv & j(\mathfrak{a})^p \pmod{\bar{\mathfrak{p}}} \\ j(\mathfrak{a}) & = & j(\bar{\mathfrak{p}}\mathfrak{a}) & \equiv & j(\mathfrak{a})^p \pmod{\mathfrak{p}} \end{array}\right. \\ & \Longrightarrow & p \text{ splits completely in } K(j(\mathfrak{a}))/\,\mathbf{Q} \end{array}$$

where the last implication follows since the Frobenius element of any prime above p in $K(j(\mathfrak{a}))$ is trivial. For all these implications to hold, we have to exclude the finitely many primes p that divide either $\mathrm{disc}(\mathcal{O})$ or the index of $\mathcal{O}_K[j(\mathfrak{a})]$ in the ring of integers of $K(j(\mathfrak{a}))$. It follows from the Chebotarev density theorem that $K(j(\mathfrak{a}))$ is contained in $K_{\mathcal{O}}$.

We leave the converse inclusion, which also uses Chebotarev density, as an exercise for the reader. □

3.2 CM points on modular curves

The considerations above belong to the classical era of complex multiplication. Further results that we have not included here – but may be proved with little extra work – concern the fields of rationality of CM values of Siegel units describing the x-coordinates of torsion points, and the action of the Galois group on them. This yields more refined information than the eta quotients h_γ considered above.

These results of the theory of complex multiplication may be fruitfully summarised and strengthened in adèlic language, providing us with a concrete way to determine the fields of rationality of CM points on various modular curves. We briefly discuss the formulation of Shimura [Shi71], and illustrate it on a few concrete examples, by making explicit computations of CM points on a few small modular curves.

Shimura reciprocity

Consider an open compact subgroup U of $\mathrm{GL}_2(\widehat{\mathbf{Z}})$. There is an associated modular curve Y_U, whose set of complex points is described by an analytic uniformisation from the half planes $\mathfrak{H}^\pm = \mathbf{C} - \mathbf{R}$ via

$$Y_U(\mathbf{C}) = \mathrm{GL}_2(\mathbf{Q}) \,\backslash\, \mathfrak{H}^\pm \times \mathrm{GL}_2(\widehat{\mathbf{Q}}) \,/\, U \tag{3.3}$$

The actions of the groups that we quotient out by are

- $\mathrm{GL}_2(\mathbf{Q})$ acts on $\mathfrak{H}^\pm$ by linear fractional transformations, and on $\mathrm{GL}_2(\widehat{\mathbf{Q}})$ by left multiplication,
- the compact open subgroup U acts trivially on $\mathfrak{H}^\pm$, and on $\mathrm{GL}_2(\widehat{\mathbf{Q}})$ by right multiplication.

By strong approximation for SL_2, the set of connected components of Y_U is canonically identified

$$\det : \pi_0(Y_U) \xrightarrow{\sim} \widehat{\mathbf{Z}}^\times / \det(\mathrm{U})$$

For a compact open U with surjective determinant, the curve Y_U is therefore connected, and it may be shown that the curve is defined over $\mathbf{Q}$ by its *canonical model.* In this canonical model, the coordinates of points with complex multiplication are algebraic, and defined over a finite abelian extension of the CM field. The Shimura reciprocity law describes how the Galois group acts on these points.

Suppose $\tau \in \mathcal{H}^\pm$ has complex multiplication by an order $\mathcal{O}$ in an imaginary quadratic field K. Then for every $\alpha \in K^\times$, there is a unique matrix $\eta_\tau(\alpha) \in \mathrm{GL}_2(\mathbf{Q})$ with the property

$$\eta_\tau(\alpha) \cdot \begin{pmatrix} \tau \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha\tau \\ \alpha \end{pmatrix}.$$

When extended naturally to the ring of finite adèles, we obtain a canonical map

$$\begin{array}{rcl} \widehat{\eta}_\tau \ : \ K^\times \backslash \widehat{K}^\times & \longrightarrow & \mathrm{GL}_2(\mathbf{Q}) \,\backslash\, \mathfrak{H}^\pm \times \mathrm{GL}_2(\widehat{\mathbf{Q}}) \,/\, U \ = \ Y_U(\mathbf{C}) \\ s & & [\tau, \eta_\tau(s)] \end{array} \tag{3.4}$$

The following result is commonly referred to as the Shimura reciprocity law. It summarises and generalises various statements obtained in the classical era of complex multiplication concerning the Galois action on the set of CM points on modular curves. For its proof, we refer to Shimura [Shi71, § 6].

Theorem 12. *The image of* $\widehat{\eta}_\tau$ *is contained in* $Y_U(K^{\mathrm{ab}})$ *and the action of* $\mathrm{Gal}(K^{\mathrm{ab}}/\,\mathbf{Q})$ *is described by*

- *The map* $\widehat{\eta}_\tau$ *is equivariant for the action of* $\mathrm{Gal}(K^{\mathrm{ab}}/K)$, *if we make the latter act on* $K^\times \backslash \widehat{K}^\times$ *via multiplication by the* <u>*inverse*</u> *of the global Artin map from class field theory.*
- *The map* $\widehat{\eta}_\tau$ *transforms under complex conjugation* σ *via the rule*

$$\widehat{\eta}_\tau(s)^\sigma = \widehat{\eta}_{\overline{\tau}}(s^{-1}).$$

Some explicit computations

To render Shimura reciprocity more explicit, we treat a number of explicit examples. The adèlic description of the law may be made more explicit by observing that when $\tau \in \mathfrak{H}^\pm$ is a CM point, there is a uniquely determined triple of integers $(a, b, c) \in \mathbf{Z}^3$ that are coprime and satisfy

$$\tau = \frac{-b + \sqrt{\Delta}}{2a}$$

where $\Delta = b^2 - 4ac$ is the discriminant of the order $\mathcal{O} = \mathbf{Z} + \mathbf{Z}\, a\tau$ by which τ has complex multiplication. The map $\eta_\tau : K^\times \to \mathrm{GL}_2(\mathbf{Q})$ in the statement of Shimura reciprocity allows the explicit description

$$\eta_\tau : x + ya\tau \longmapsto \begin{pmatrix} x - by & -cy \\ ay & x \end{pmatrix}, \qquad \text{for } (x, y) \in \mathbf{Q}^2\,. \tag{3.5}$$

From this description we recover several classical statements about singular moduli and Heegner points on modular curves. As an illustration of the general statement, we discuss a few specific level structures with explicit examples in the remainder of this chapter.

1. The modular curve $X(1)$**.** The statements about singular moduli we proved in our discussion of the classical era of complex multiplication in § 3.1 follow upon taking the maximal open subgroup

$$U = \mathrm{GL}_2(\widehat{\mathbf{Z}})$$

whose associated modular curve is $Y(1)$. The canonical model is defined over $\mathbf{Q}$ and has function field $\mathbf{Q}(j)$. The kernel of the map (3.4) is the set of idèles for which the completion of (3.5) is integral at all places. This is exactly the unit group of $\widehat{\mathcal{O}}$, whose fixed field is the ring class field of $\mathcal{O}$, recovering Corollary 1.

In fact, this may be sharpened so that also the action of the Galois group $\mathrm{Pic}(\mathcal{O})$ on the CM point defined by τ becomes transparent. Suppose that the lattice $\mathfrak{a} = \mathbf{Z} + \mathbf{Z}\,\tau$ is an invertible fractional $\mathcal{O}$-ideal, and $\mathfrak{b}$ is any other such ideal. Take any idèle $s_\mathfrak{b}$ representing the class of $\mathfrak{b}$ in the Picard group $\mathrm{Pic}(\mathcal{O})$. Using strong approximation for $\mathrm{GL}_2(\widehat{\mathbf{Q}})$ we may factorise it as

$$\eta_\tau(s_\mathfrak{b}) = \gamma_\mathfrak{b}\delta_\mathfrak{b}, \qquad \gamma_\mathfrak{b} \in \mathrm{GL}_2(\mathbf{Q}), \delta_\mathfrak{b} \in \mathrm{GL}_2(\widehat{\mathbf{Z}}).$$

Now note that $\gamma_\mathfrak{b}$ represents a matrix of base change between $\mathfrak{a}$ and $\mathfrak{a}\mathfrak{b}$. It follows that

$$[\tau, \eta_\tau(s_\mathfrak{b})] \sim [\tau', 1]$$

where $\mathfrak{a}\mathfrak{b} = \mathbf{Z} + \mathbf{Z}\,\tau'$. We deduce the classical statements proved above, namely

$$j(\mathfrak{a}\mathfrak{b}) = j(\mathfrak{a})^{\mathrm{Art}^{-1}(\mathfrak{b})}, \qquad \text{and} \qquad j(\mathfrak{a})^{\sigma} = j(\mathfrak{a}^{\sigma}),\ \sigma \in \mathrm{Gal}(\mathbf{C}\,/\,\mathbf{R}).$$

In other words, all singular moduli of CM points for a fixed order $\mathcal{O}$ form a full set of algebraic conjugates over K, for the simply transitive action of $\mathrm{Pic}(\mathcal{O})$ provided by the theory of complex multiplication.

Example. Consider $\mathcal{O}$ the quadratic order of discriminant $\Delta = -23$, which has $\mathrm{Pic}(\mathcal{O}) \simeq \mathbf{Z}\,/3\mathbf{Z}$. The three singular moduli may be computed numerically from the q-expansion, and are approximately

$$\begin{array}{rcl} j\left(\frac{-1+\sqrt{-23}}{2}\right) & \approx & -3493225.6999699 \\ j\left(\frac{-1+\sqrt{-23}}{4}\right) & \approx & 737.8499850 + 1764.0189386i \\ j\left(\frac{-1-\sqrt{-23}}{4}\right) & \approx & 737.8499850 - 1764.0189386i \end{array}$$

Note that we may recover their minimal polynomial by evaluating all elementary symmetric polynomials, the result of which must be an integer. We obtain the Hilbert class polynomial

$$x^3 + 3491750x^2 - 5151296875x + 5^9 \cdot 11^3 \cdot 17^3 = 0.$$

2. The modular curves $X_0(n)$. Much of our understanding of the Birch–Swinnerton-Dyer conjecture has come from the properties of *Heegner points*. We define them here, and return to their arithmetic properties found by Gross–Zagier and Kolyvagin in § 3.3. Consider the open compact subgroup

$$U = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\widehat{\mathbf{Z}}) \ : \ b \in n\widehat{\mathbf{Z}} \right\}$$

The corresponding modular curve is $X_0(n)$, which is defined over $\mathbf{Q}$. Suppose $\tau \in \mathcal{H}^{\pm}$ is a CM point with associated order $\mathcal{O} = \mathbf{Z} + \mathbf{Z}\,a\tau$. The kernel of $\widehat{\eta}_\tau$ consists of the suborder $\mathcal{O}_n = \mathcal{O} \cap (\mathbf{Z} + \mathbf{Z}\,n\tau)$. When $\mathcal{O}_n = \mathcal{O}$ (i.e. when $n|a$ we say that the image of τ on $Y_0(n)$ is a *Heegner point*. From the Shimura reciprocity law, we see that all Heegner points are defined over the ring class field $K_{\mathcal{O}}/K$.

Note that for a given value of $n > 0$, the set of Heegner points of discriminant $\Delta < 0$ may be empty, since $n|a$ implies that every prime divisor $\ell | n$ splits in the order $\mathcal{O}$ of discriminant $\Delta = b^2 - 4ac \equiv 0 \pmod{\ell}$. In fact, the set of Heegner points on $Y_0(N)$ of discriminant Δ is non-empty if and only if every prime divisor of n splits in $\mathcal{O}$. The latter condition is usually called the *Heegner hypothesis*.

Example. Let us explore some Heegner points for the example

$$X_0^+(58) = E : \ y^2 + xy = x^3 - x^2 - x + 1$$

which has conductor 58 and j-invariant $-3^3 \cdot 19^3/2^2 \cdot 29$. Using a 2-descent we discover that $E(\mathbf{Q})$ is torsion-free and has rank one, its Mordell–Weil group being generated by the point $P_0 = (0, 1)$.

Like every elliptic curve over $\mathbf{Q}$, the curve E is modular. We will however need a more explicit form of this statement, and will find an explicit pair of functions $x(\tau)$ and $y(\tau)$ on the upper half plane that are invariant under the group

$$\Gamma := \left\langle \Gamma_0(58)\,,\ \begin{pmatrix} 0 & -1/\sqrt{2} \\ \sqrt{2} & 0 \end{pmatrix}\,,\ \begin{pmatrix} 0 & -1/\sqrt{29} \\ \sqrt{29} & 0 \end{pmatrix} \right\rangle$$

such that

$$\begin{cases} y^2 + xy & = & x^3 - x^2 - x + 1, \\ dx/(2y + x) & = & 2\pi i f(\tau) d\tau, \end{cases}$$

where f is the cusp form of weight two associated to E, which has Fourier expansion

$$f(q) = q - q^2 - 3q^3 + q^4 - 3q^5 + 3q^6 - 2q^7 - q^8 + 6q^9 + 3q^{10} + \dots$$

From this data, the power series $x(\tau)$ and $y(\tau)$ may be computed recursively, yielding the following series

$$\begin{array}{rcl} x(\tau) & = & q^{-2} + q^{-1} + 3 + 3q + 7q^2 + 7q^3 + 14q^4 + \dots \\ y(\tau) & = & -q^{-3} - 2q^{-2} - 5q^{-1} - 8 - 16q - 24q^2 - 44q^3 + \dots \end{array}$$

With the uniformisation in hand, we may now compute the values of $(x(\tau), y(\tau)) \in E(\mathbf{C})$ numerically, and attempt to recognise them as algebraic numbers. We begin with $\Delta = -23$ where find for instance the Heegner form $[116, 21, 1] \in Q_{-23}$ and we compute the corresponding Heegner point to be

$$\tau = \frac{-21 + \sqrt{-23}}{232} \longmapsto (x(\tau), y(\tau)) = (0.8774\ldots - i0.7448\ldots, -0.2150\ldots + i1.3071\ldots).$$

Using algebraic recognition routines such as LLL, one finds that both coordinates indeed appear to be defined over the number field $\mathbf{Q}(\alpha)$ where $\alpha^3 - \alpha^2 + 1 = 0$. This defines the Hilbert class field over $K = \mathbf{Q}(\sqrt{-23})$ as predicted by CM theory.

In fact, there are 3 orbits on Q_{-23} for the action of Γ, each give a corresponding Heegner point. We know that individually, they are defined over the Hilbert class field of $\mathbf{Q}(\sqrt{-23})$, but their sum is rational over K. The curve E has sign $w_{58} = 1$ for the Fricke involution, which implies that the sum of the three Heegner points is also fixed under complex conjugation, and therefore rational over $\mathbf{Q}$. We compute that the sum of the corresponding points is

$$\begin{array}{rcl} (0.8774\ldots - i0.7448\ldots, -0.215\ldots + i1.3071\ldots) & + & \\ (0.8774\ldots + i0.7448\ldots, -0.215\ldots - i1.3071\ldots) & + & \\ (-0.7548\ldots, -0.5698\ldots) & = & (1, 0) \ = \ -2P_0 \end{array}$$

This is very encouraging, as it shows that CM theory is capable of producing rational points of infinite order on elliptic curves, as opposed to our experiences with cusps.

Could it be that Heegner points are always non-trivial? Alas, the reality is clearly more subtle, as we can see by considering instead $\Delta = -71$. Here we find for instance the Heegner form $[174, 25, 1] \in Q_{-71}$ and the associated Heegner point

$$\tau = \frac{-25 + \sqrt{-71}}{348} \longmapsto (x(\tau), y(\tau)) = (-0.3448\ldots - i1.0787\ldots, 2.0250\ldots + i0.9148\ldots)$$

which is a non-trivial point defined over the Hilbert class field of $K = \mathbf{Q}(\sqrt{-71})$, which is an extension of degree $7 = |\mathrm{Cl}_K|$ obtained by adjoining a root α of the polynomial

$$\alpha^7 + \alpha^6 - \alpha^5 - \alpha^4 - \alpha^3 + \alpha^2 + 2\alpha - 1 = 0.$$

This time, there are 7 orbits for Γ, the sum of the corresponding points is 0 in the Mordell–Weil group $E(\mathbf{Q})$. It appears that for some discriminants, the Heegner point construction succeeds in finding a non-trivial rational point, whereas for other discriminants it yields trivial or torsion points. This leads us to the natural question

> **Q:** Heegner points provide a systematic construction of rational points on (modular) elliptic curves. But when are the resulting points non-trivial?

In fact, we might sharpen this question further, and ask about the *position* of Heegner points in modular Jacobians. For isntance, let us denote the Heegner point of discriminant $-d$ by P_d, and recall that $E(\mathbf{Q})$ is free of rank one, generated by P_0. Set

$$P_d = b_d P_0, \qquad b_d \in \mathbf{Z}$$

where the sequence of integers encodes the position of the various Heegner points we construct this way. Continuing to compute as above, we find the following generating series:

$$\sum_{d>0} b_d q^d = q^4 + 2q^7 - q^{16} + q^{20} - 2q^{23} - q^{24} - 2q^{28} - 2q^{36} + 3q^{52} - 4q^{63} + \dots \tag{3.6}$$

We will return to all of these questions in § 3.3 with the work of Gross and Zagier, which determines the position of Heegner points in modular Jacobians, and in particular the question of whether they are trivial. It follows from the work of Gross–Kohnen–Zagier that the generating series (3.6) is the q-expansion of a modular form of weight $3/2$, which is attached to the elliptic curve E under the Shimura correspondence.

3.3 The work of Gross–Zagier

The origins of the groundbreaking work of Gross and Zagier are well documented [Zag83, Gro22] and are highly recommended reading. A central goal in the collaboration was to study the prime factorisations of differences of singular moduli $j(\tau_1) - j(\tau_2)$, starting during a momentous week in September 1982 for CM points of the same discriminant [Gro22], and later for coprime discriminants [Zag83]. The prime factorisations of differences of singular moduli are extraordinarily rich. Let us consider the example

$$\begin{aligned} j\left(\frac{1+\sqrt{-67}}{2}\right) - j\left(\frac{1+\sqrt{-163}}{2}\right) &= -2^{15}\cdot 3^3\cdot 5^3\cdot 11^3 + 2^{18}\cdot 3^3\cdot 5^3\cdot 23^3\cdot 29^3 \\ &= \quad 2^{15}\cdot 3^7\cdot 5^3\cdot 7^2\cdot 13\cdot 139\cdot 331 \end{aligned}$$

The primes appearing in the factorisation all have the property that they are inert in both of the quadratic fields $\mathbf{Q}(\sqrt{-67})$ and $\mathbf{Q}(\sqrt{-163})$, and they all divide a positive integer of the form

$$M_n := \frac{67\cdot 163 - n^2}{4} \in \mathbf{Z}_{>0}\,.$$

Note that for any discriminants $\Delta_1, \Delta_2 < 0$ the set $S(\Delta_1, \Delta_2)$ of positive integers M_n of this form is finite. Gross and Zagier prove a remarkable elementary formula for the norm of differences of singular moduli, which makes this more precise. They consider the case where Δ_1, Δ_2 are coprime fundamental discriminants. Let ℓ be any prime that is not inert in the real quadratic field

$$F = \mathbf{Q}(\sqrt{\Delta_1\Delta_2}).$$

Then at least one of the discriminants $\Delta \in \{\Delta_1, \Delta_2\}$ is not divisible by ℓ, and we define

$$\varepsilon(\ell) = \left(\frac{\Delta}{\ell}\right), \qquad \text{and} \qquad \varepsilon(m) = \prod_i \varepsilon(\ell_i)^{e_i}, \qquad P(m) := \prod_{\substack{ab=m\\ a,b>0}} a^{\varepsilon(b)}$$

for any integer $m = \prod_i \ell_i^{e_i} > 0$ that is only divisible by primes ℓ_i that are not inert in F. The function P is always equal to a prime power. The following result is proved in [GZ85, Theorem 1.3].

Theorem 13. *With the above definition, we have*

$$\mathrm{Nm}_{\mathbf{Q}}(j(\tau_1) - j(\tau_2))^2 = \pm \prod_{M_n \in S(\Delta_1, \Delta_2)} P(M_n) \tag{3.7}$$

By far the most remarkable thing about this formula is its proof, or rather, proof*s*. Gross and Zagier give two independent proofs of their factorisation formula, each using very different techniques:

- **Algebraic proof:** There is a pair of CM elliptic curves E_1 and E_2 associated to τ_1 and τ_2. Since the reduction map is injective on endomorphism rings, these two curves can only be isomorphic modulo a prime above p if their endomorphism ring is of rank 4. In this case, their common reduction is supersingular, and we obtain a pair of embeddings

  $$\mathcal{O}_1, \mathcal{O}_2 \ \hookrightarrow\ R \subset B_{p\infty}$$

  of the quadratic orders of discriminants Δ_1 and Δ_2 into a maximal order R of the definite quaternion algebra $B_{p\infty}$ ramified at p. The p-adic valuation of the difference of singular moduli is then determined purely algebraically in terms of the embeddings.

- **Analytic proof:** Associated to the genus character χ of F that cuts out the biquadratic extension L/F obtained by adjoining τ_1 or τ_2 to F is a real analytic family of Hilbert Eisenstein series

  $$E_s(z_1, z_2) \in M_{1,1}^{\mathrm{an}}(\mathrm{SL}_2(\mathcal{O}_F))$$

  considered by Hecke in his famous work [Hec24], depending on an analytic variable s. Gross–Zagier observe that since $E_s(z_1, z_2)$ vanishes at $s = 0$, its first order derivative with respect to s retains the modularity properties. They compute the Fourier coefficients of

  1. the first order derivative with respect to s, at $s = 0$,
  2. its diagonal restriction,
  3. its holomorphic projection.

  The Fourier coefficients consist of two different contributions. On the one hand, the logarithm

  $$\log \mathrm{Nm}_{\mathbf{Q}}\left(j(\tau_1) - j(\tau_2)\right)$$

  and on the other hand, a finite algebraic contribution equal to the logarithm of the right hand side of (3.7). The relationship between these two quantities follows, since the holomorphic modular form

  $$\mathrm{Proj}_{\mathrm{hol}}\left(\frac{\partial}{\partial s} E_s(z, z) \mid_{s=0}\right) \in M_2(\mathrm{SL}_2(\mathbf{Z})) = \{0\}$$

  necessarily has all of its Fourier coefficients equal to zero, from which (3.7) follows.

These two independent proofs each have a very different and complementary role to play in the subsequent work of Gross–Zagier [GZ86]. The factorisation formula may be interpreted as a calculation of the Néron local heights of the Heegner divisors P_{Δ_1} and P_{Δ_2} on the modular curve $X(1)$ of level $\mathrm{SL}_2(\mathbf{Z})$.

The logaritm of the norm of differences of singular moduli is the term that arises from the evaluation of a Green's function, and therefore constitutes the contribution at the infinite prime, whereas the multiplicity of $\log(p)$ in the logarithm of the right hand side of (3.7) is precisely the intersection number of these two divisors, apparent in the algebraic proof. The fact that they sum up to zero is just a reflection of the fact that the global Néron height is trivial, since $X(1)$ has trivial Jacobian.

This reinterpretation in terms of global Néron heights of Heegner divisors generalises to all modular curves $X_0(N)$, and is the subject of the subsequent papers [GZ86] on Heegner divisors with the same discriminant, and [GKZ87] on Heegner divisors with coprime discriminants. The computation of the local heights follows similar techniques to those in level one. The series that appears in the analytic proof is adapted to higher level, and acts as the holomorphic kernel for the linear functional

$$\left[\, f \;\mapsto\; \mathrm{L}'(f/K,1) \,\right] \qquad \in\; \mathrm{Hom}_{\mathbf{C}}(S_2(\Gamma_0(N)),\mathbf{C}).$$

These arguments are used in [GZ86] to establishing the following landmark result.

Theorem 14 (Gross–Zagier)**.** *Let f be a newform of weight two on $\Gamma_0(N)$, and K an imaginary quadratic field of discriminant Δ such that every prime divisor of N splits in K. Consider the Heegner divisor*

$$P_K := \mathrm{Tr}_{H/K}((P)-(\infty)) \in J_0(N)(K)$$

for some Heegner point P of discriminant Δ on $X_0(N)$, defined over the Hilbert class field H/K. Then

$$\mathrm{L}'(f/K,s) = \frac{8\pi^2 \cdot \langle f, f\rangle_{\mathrm{Weil}}}{\sqrt{|\Delta|}\cdot|\mathrm{Cl}_K|\cdot|\mu_K|^2}\cdot \widehat{h}(P_{K,f})$$

where $P_{K,f} \in A_f(K)\otimes\mathbf{C}$ is the f-isotypical component of P_K in $J_0(N)(K)\otimes\mathbf{C}$.

The consequences for the Birch–Swinnerton-Dyer conjecture are evident. For elliptic curves $E/\,\mathbf{Q}$ we find that E has analytic rank one over K if and only if $E(K)$ contains a Heegner point of infinite order. Interestingly, when $E(K)$ is of rank at least two, all Heegner points are torsion!

3.4 The work of Kolyvagin

The next big step in our understanding of the Mordell–Weil groups of elliptic curves $E/\,\mathbf{Q}$ came from the work of Kolyvagin. He established the following result, providing an important complement to the work of Gross and Zagier on Heegner points. Let E be an elliptic curve over $\mathbf{Q}$ of conductor N, and $P_K \in E(K)$ the image under a modular parametrisation $J_0(N) \to E$ of a Heegner point over an imaginary quadratic field K as above. The following result was proved in [Kol89].

Theorem 15 (Kolyvagin)**.** *Suppose P_K is of infinite order. Then $\mathrm{rk}_{\mathbf{Z}}E(K) = 1$ and $\text{Ш}(E/K)$ is finite.*

The proof of this theorem is via a p-descent argument, for a suitably chosen p. Kolyvagin uses the non-triviality of the Heegner point P_K, as well as the fact that it is part of a norm-compatible system of Heegner points of larger conductor, to systematically construct enough singular classes in the dual Selmer group. The bound on the size of the Selmer group is then obtained via an application of duality.

Remark. During early stages of the planning, my goal was to discuss this wonderful proof in more detail. This plan came to an abrupt halt when I learned that an excellent treatment was recently given at ICTS by Castella, during a workshop in 2021. The overlap in the audiences is likely very large, so I will not say anything more about the mechanics of Kolyvagin's proof.

Consequences for BSD

The work of Kolyvagin, combined with that of Gross and Zagier, implies the Birch–Swinnerton-Dyer conjecture in analytic rank at most one, as follows. Let E be an elliptic curve over $\mathbf{Q}$ of conductor N, and assume that the analytic rank at most 1. Choose an imaginary quadratic field $K/\mathbf{Q}$ of discriminant $\Delta < 0$ such that the Heegner hypothesis is satisfied, i.e. all primes dividing N are split in $K/\mathbf{Q}$, and such that

$$\mathrm{ord}_{s=1}\,\mathrm{L}(E/K,s) = 1.$$

The fact that such a K can be chosen (in infinitely many ways) is an analytic result due to ??. The analytic rank of E/K being exactly equal to one implies, by the theorem of Gross–Zagier, that the Heegner points P_K are non-torsion in $E(K)$. Now observe that

$$E(K) \otimes \mathbf{Q} = \Big(E(\mathbf{Q}) \oplus E^{(\Delta)}(\mathbf{Q})\Big) \otimes \mathbf{Q}\,. \tag{3.8}$$

where $E^{(\Delta)}$ denotes the quadratic twist of E associated to K. The theorem of Kolyvagin implies that $E(K)$ is of rank one, and the Heegner point P_K generates a subgroup of finite index. To determine the projection of P_K to the two factors on the right hand side of (3.8), we use the Atkin–Lehner operator w_N to conclude

$$\begin{array}{rcl} w_N P_K = P_K & \Rightarrow & \mathrm{rk}_{\mathbf{Z}} E(\mathbf{Q}) = 1 \\ w_N P_K = -P_K & \Rightarrow & \mathrm{rk}_{\mathbf{Z}} E(\mathbf{Q}) = 0 \end{array}$$

Similar considerations hold for all modular abelian varieties $A_f/\mathbf{Q}$, see [KL89]. They imply that all such abelian varieties with $\mathrm{L}(A_f,1) \neq 0$ must be of algebraic rank zero, as predicted by BSD.

Uniform boundedness of torsion

A natural question to ask is whether the methods of Mazur extend to number fields of higher degree. This was explored by Kamienny and Mazur [Kam92b, Kam92a, KM95], see also Edixhoven [Edi95]. The results remained in first instance limited to particular fields, such as $K/\mathbf{Q}$ quadratic. An important step forward came in the work of Merel [Mer96], who showed that the torsion is uniformly bounded in the strongest possible sense:

Theorem 16 (Merel)**.** *Let E be an elliptic curve defined over a number field K. The size of the torsion subgroup of $E(K)$ is bounded by a constant depending only on the degree of K over $\mathbf{Q}$.*

The reader who wishes to study the proof of this theorem is best served with the original paper [Mer96]. The key innovation is to replace the Eisenstein quotient with the technically simpler *winding quotient*, which is defined as follows. Recall from our discussion of Manin–Drinfeld that the path $\{0 \to \infty\}$ in $\mathcal{H}$ defines an element of $\mathrm{H}_1(X_0(N),\mathbf{Q})$ which we will denote by c_{wind}. The winding quotient is defined by

$$J_{\mathrm{wind}} := J_0(N)/I_{\mathrm{wind}}J_0(N), \qquad I_{\mathrm{wind}} = \mathrm{Ann}_{\mathbb{T}}(c_{\mathrm{wind}}).$$

Note that the analytic rank of the winding quotient is zero, since any simple $\mathbf{Q}$-isogeny factor A_f of the winding quotient has by definition the property that its central L-value satisfies

$$\mathrm{L}(f,1) = 2\pi \int_0^\infty f(it)t^{s-1}dt \neq 0.$$

The winding quotient, as compared to the Eisenstein quotient, has the following key technical differences in the context of the study of torsion points using the strategy of Mazur which we discussed in § 2.

- The intricate flat descent arguments of Mazur are replaced by the works of Gross–Zagier [GZ85, GZ86] and Kolyvagin [Kol89] which imply that the rank of the winding quotient is zero. The non-triviality of the winding quotient is easily shown by analytic arguments.
- The formal immersion criterion of Mazur–Kamienny translates into a statement about the linear independence of certain modular symbols. Merel shows this by giving explicit expressions for the intersection numbers of these modular symbols, see Rebolledo [Reb09] for a detailed discussion.

Remark. The winding quotient is tremendously useful in this context, but it should be pointed out that the (smaller) Eisenstein quotient is of great independent interest, and continues to be intensely studied in the context of modularity lifting theorems, see for instance the recent works [WWE20, Lec21].

Appendix **A**

Appendix

A.1 Exercises

1. Let f be a newform of weight two on $\Gamma_0(N)$. Show that

$$\mathrm{L}^*(f,s) := \left(\frac{\sqrt{N}}{2\pi}\right)^s \cdot \Gamma(s) \cdot \mathrm{L}(f,s) = \int_0^\infty f\left(\frac{it}{\sqrt{N}}\right) t^s \frac{dt}{t}.$$

2. Carry out the procedure in § 2.1 to show that $X_1(11)$ is an elliptic curve with Weierstraß equation

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

3. Let E be an elliptic curve over $\mathbf{Q}$ such that the $G_{\mathbf{Q}}$-module $E[2]$ is irreducible. Consider the morphism

$$\varphi \ : \ \mathrm{H}^1(\mathbf{Q}, E[2]) \longrightarrow \mathrm{H}^1(K, \mu_2) \simeq K^\times/(K^\times)^2$$

constructed in (**??**). Show that φ is injective, and has image equal to

$$\mathrm{Im}(\varphi) = \mathrm{Ker}\left(K^\times/(K^\times)^2 \xrightarrow{\mathrm{Nm}} \mathbf{Q}^\times/(\mathbf{Q}^\times)^2\right)$$

Hint: Show that the map φ arises in the long exact sequence in cohomology associated to an appropriately defined short exact sequence of $G_{\mathbf{Q}}$-modules of the form

$$1 \longrightarrow E[2] \longrightarrow \mathrm{Ind}_K^{\mathbf{Q}}(\mu_2) \longrightarrow \mu_2 \longrightarrow 1.$$

4. Determine the Mordell–Weil group of the curve

$$E_N \ : \ y(N-y) = x^3 - x.$$

for the case $N = 6$ appearing in the work of Diophantus. Show that the rank is at least two for all but finitely many integer values of N, and find examples where it is larger than two.

5. Determine all rational solutions to

$$E \ : \ y^2 + xy + y = x^3 - x^2 - x - 14.$$

Bonus: Find all rational elliptic curves with a rational subgroup of order 17.

6. Prove that there are no elliptic curves over $\mathbf{Q}$ with a rational point of order 17.

Hint: First deduce it from the previous exercise. Then prove it using a descent on $X_1(17)$ in the style of Mazur–Tate. You may use that $J_1(17)$ has a rational point of order 73.

A.2 The flat topology

The arguments of Mazur take place in the *flat topology*. Anyone who wants to understand the fine print of these techniques should consult Milne [Mil80]. In these notes, we will take a pedestrian approach to the inherent technicalities, and content ourselves with using the formal cohomological framework, along with one or two black boxes. Treating the cohomological formalism as a given allows one to already appreciate some of the beautiful ideas in the general case.

Motivation. In the previous chapter, we showed that there are no elliptic curves over $\mathbf{Q}$ with a rational point of order 11, using an explicit 2-descent. But how does this approach generalise to find rational points on $X_1(\ell)$ for primes $\ell > 11$? A careful examination of the structures we encountered in the 2-descent makes us desire for an alternative approach that has the following features:

- The 2-descent involved some cubic extension over which we cannot expect good control in general. The general descent argument should involve structures that have 'meaning', in the sense of the moduli problem, so as to generalise to other modular curves.
- The descent argument should yield sharp bounds so as to avoid having to write down explicit equations for twists. Clearly, this will not be a fruitful approach for primes $\ell > 11$, so we will most likely need to engage with what happens at the finite set of bad primes S.

The ideas of Mazur achieve these goals in the following way. The first point is addressed by performing a descent with respect to a canonical class of rational points, which in the case of $X_1(11)$ is accounted for by the rational 5-torsion. The second point is resolved by working with flat cohomology groups, making the cohomological framework interact with the finite set of bad places S. In these notes, we will avoid the additional technical complications that arise at primes of bad reduction, settling for the primes dividing the order of the isogeny. This will be sufficient for the specific examples that we treat here.

The flat topology. Mazur replaces the étale topology by the finer *flat topology* which is better equipped for dealing with group schemes of order p in characteristic p. The flat topology is a Grothendieck topology, where the coverings of a scheme S are given by families of morphisms

$$\{\varphi_i : T_i \longrightarrow S\}$$

where each morphism φ is flat and locally of finite presentation, and their images cover S in the sense that $S = \bigcup_i \varphi_i(T_i)$. This notion of coverings satisfied the axioms of a Grothendieck topology [Mil80, Chapter II.1] and very important theorem of Grothendieck [Mil80, Theorem I.2.17] implies that whenever G is a commutative group scheme over S then the functor defined by

$$T \mapsto \mathrm{Hom}_S(T, G)$$

is a sheaf of abelian groups with respect to the flat topology defined above. This theorem is central for the practical usefulness of the flat site, and it gives a mechanism whereby short exact sequences of commutative group schemes give rise to long exact sequences in flat cohomology via the cohomological framework developed by Grothendieck, see [Mil80, Chapter III].

Kummer theory. We mention one key fact about the flat topology that is used in the descent arguments on the Jacobians of $X_1(11)$ and $X_1(13)$ below. We often need to control cohomology with values in μ_p when analysing the torsion of modular Jacobians. To do this, we may use that the Kummer sequence

$$1 \longrightarrow \mu_n \longrightarrow \mathbf{G}_m \xrightarrow{\cdot n} \mathbf{G}_m \longrightarrow 1$$

of abelian group schemes over any base scheme S is *exact* in the flat topology. To see why it is surjective, let U be any S-scheme with a global section $u \in \mathrm{Hom}_S(U, G) = \Gamma(U, \mathcal{O}_U^\times)$. Choose an affine Zariski covering of U by open sets $\mathrm{Spec}(A_i)$, and let $u_i \in A_i^\times$ be the restriction of u to this open subset. For each such open set, there is a covering in the flat topology given by

$$\mathrm{Spec}\, A_i[T]/(T^n - u_i) \longrightarrow \mathrm{Spec}\, A_i \tag{A.1}$$

and note that the restriction (= pullback) of u_i to this covering is in the image of the n-th power map, since it is the n-th power of the unit T. This shows that the Kummer sequence is indeed right exact.

Remark. Note that the Kummer sequence is not generally exact in the étale topology. The problem with the above argument is that the covering (A.1) is not a covering in the étale topology. When n is invertible on S, Hensel's lemma for the polynomial $T^n - u_i$ does imply that the covering (A.1) is étale, and therefore the Kummer sequence is exact in such cases. Note that we already used this fact in our discussion of the weak Mordell–Weil theorem, and it was precisely our desire to use this fact that caused us to add the primes dividing n to the finite set S of bad places. Perhaps this strengthens our faith that we may include these bad places, at the cost of working with the formalism of flat cohomology.

An important consequence of the exactness of the Kummer sequence, which we will use several times in the arguments to follow, is that when the base scheme is the spectrum of the ring $\mathcal{O}_S$ of S-integers in a number field K, its first flat cohomology group with coefficients in μ_p can be computed in terms of arithmetic invariants of the ring, as follows:

Lemma 1. *Suppose S is a finite set of primes in a number field K, and n is any integer. Then we have a short exact sequence*

$$1 \longrightarrow \mathcal{O}[1/S]^\times/(\mathcal{O}[1/S]^\times)^n \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathcal{O}[1/S], \mu_n) \longrightarrow \mathrm{Cl}(\mathcal{O}[1/S])[n] \longrightarrow 1.$$

Proof. By the exactness of the Kummer sequence in the flat topology, we obtain a long exact sequence in flat cohomology from which we extract the five-term sequence

$$\mathrm{H}^0_{\mathrm{fppf}}(\mathcal{O}[1/S], \mathbf{G}_m) \xrightarrow{(-)^n} \mathrm{H}^0_{\mathrm{fppf}}(\mathcal{O}[1/S], \mathbf{G}_m) \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathcal{O}[1/S], \mu_n) \longrightarrow \mathrm{H}^1_{\mathrm{fppf}}(\mathcal{O}[1/S], \mathbf{G}_m) \xrightarrow{(-)^n} \mathrm{H}^1_{\mathrm{fppf}}(\mathcal{O}[1/S], \mathbf{G}_m)$$

The lemma now follows from the observations that

$$\begin{array}{rclcl} \mathrm{H}^0_{\mathrm{fppf}}(\mathcal{O}[1/S], \mathbf{G}_m) & = & \mathcal{O}[1/S]^\times & & \\ \mathrm{H}^1_{\mathrm{fppf}}(\mathcal{O}[1/S], \mathbf{G}_m) & = & \mathrm{H}^1_{\text{ét}}(\mathcal{O}[1/S], \mathbf{G}_m) & = & \mathrm{Pic}(\mathcal{O}[1/S]) \end{array}$$

where the latter equalities follow from the fact that flat cohomology agrees with étale cohomology when valued in the sheaf $\mathbf{G}_m$, and they both compute the Picard group of the base [Mil80, Theorem III.4.9]. The Picard group of a number ring is its class group, consisting of invertible fractional ideals modulo principal ones. This proves the lemma. □

Bibliography

[BCDT01] C. Breuil, B. Conrad, F. Diamond, and R. Taylor. On the modularity of elliptic curves over **q**. *J. Amer. Math. Soc.*, 14(4):843–939, 2001. ↑7.

[BCH$^+$57] A. Borel, S. Chowla, C.S. Herz, K. Iwasawa, and J.-P. Serre. *Seminar on complex multiplication*, volume 21 of *Lecture Notes in Math.* Springer-Verlag, 1957. ↑29, 30.

[Ber16] W.E.H. Berwick. An invariant modular equation of the fifth order. *Quarterly J. Math*, 47:94–103, 1916. ↑30.

[BLS12] R. Bröker, K. Lauter, and D. Sutherland. Modular polynomials via isogeny volcanoes. *Math. Comp.*, 81:1201–1231, 2012. ↑31.

[BSD65] B. Birch and H.P.F. Swinnerton-Dyer. Notes on elliptic curves II. *J. Reine Angew. Math.*, 165(218):79–108, 1965. ↑6, 7.

[Cas65] J.W.S. Cassels. Arithmetic on curves of genus one. viii. on conjectures of birch and swinnerton-dyer. *J. Reine Angew. Math.*, 217:180–199, 1965. ↑6.

[CES03] B. Conrad, B. Edixhoven, and W. Stein. $j_1(p)$ has connected fibres. *Doc. Math.*, 8:331–408, 2003. ↑19.

[dF59] P. de Fermat. Letter to Pierre de Carcavi. 14 August 1659. ↑2.

[DR73] P. Deligne and M. Rapoport. Les schémas de modules de courbes elliptiques. In W. Kuyk, editor, *Modular forms in one variable II*, volume 349 of *LNM*, pages 143–316. Springer-Verlag, 1973. ↑14.

[Dri73] V.G. Drinfel'd. Two theorems on modular curves. *Akad. Nauk CCCP*, 7(2):83–84, 1973. ↑17.

[Edi95] B. Edixhoven. Rational torsion points of elliptic curves over number fields (after Kamienny and Mazur). *Astérisque*, 227:209–227, 1995. ↑42.

[Fer70] S. Fermat. *Diophanti Alexandrini Arithmeticorum Libri Sex: cum commentariis C.G. Bacheti et observationibus D.P. de Fermat.* 1670. ↑2.

[FLS$^+$01] E.V. Flynn, F. Leprévost, E.F. Schaefer, W.A. Stein, M. Stoll, and J. Wetherell. Empirical evidence for the Birch and Swinnerton-Dyer conjecture for modular jacobians of genus 2 curves. *Math. Comp.*, 70(236):1675–1697, 2001. ↑7.

[GKZ87] B. Gross, W. Kohnen, and D. Zagier. Heegner points and derivatives of L-series II. *Math. Ann.*, 278:497–562, 1987. ↑11, 41.

[Gro22] B. Gross. Working with Don. 2022. ↑39.

[GZ85] B. Gross and D. Zagier. On singular moduli. *J. Reine Angew. Math.*, 355:191–220, 1985. ↑11, 29, 40, 43.

[GZ86] B. Gross and D. Zagier. Heegner points and derivatives of L-series. *Invent. Math.*, 84(2):225–320, 1986. ↑11, 29, 40, 41, 43.

[Hec24] E. Hecke. Analytische Funktionen und Algebraische Zahlen. Zweiter Teil. *Abhandlungen aus dem Mathematischen Universität*, 3:213–236, 1924. ↑40.

[Her75] O. Herrmann. Über die Berechnung der Fourierkoeffizienten der Funktion $j(\tau)$. *J. Reine Angew.*

Math., 274/275:187–195, 1975. ↑30.

[ICT21] ICTS. Workshop: Elliptic curves and the special values of l-functions. In *https://www.icts.res.in/program/ECL2021*, August 2021. ↑29.

[Kam92a] S. Kamienny. Torsion points on elliptic curves and q-coefficients of modular forms. *Invent. Math.*, 109(221–229), 1992. ↑42.

[Kam92b] S. Kamienny. Torsion points on elliptic curves over fields of higher degree. *Int. Math. Res. Not.*, (6):129–133, 1992. ↑42.

[KL81] D. Kubert and S. Lang. *Modular units*. Grundlehren der mathematischen Wissenschaften. Springer-Verlag, New York, 1981. ↑18.

[KL89] V. A. Kolyvagin and D. Yu. Logachëv. Finiteness of the Shafarevich-Tate group and the group of rational points for some modular abelian varieties. *Algebra i Analiz*, 1(5):171–196, 1989. ↑42.

[KM85] N. Katz and B. Mazur. *Arithmetic moduli of elliptic curves*, volume 108 of *Annals of Math. Studies*. Princeton University Press, 1985. ↑14.

[KM95] S. Kamienny and B. Mazur. Rational torsion of prime order in elliptic curves over number fields. *Astérisque*, 228(3):81–100, 1995. ↑42.

[Kol89] V. Kolyvagin. Finiteness of e(q) and x(e,q) for a class of Weil curves. *Math. USSR-Izv.*, 32(3):523–541, 1989. ↑29, 41, 43.

[KY84] E. Kaltofen and N. Yui. On the modular equation of order 11. In *Third MACSYMA User's Conference*, pages 472–485. General Electric, 1984. ↑30.

[Lec21] E. Lecouturier. Higher Eisenstein elements, higher Eichler formulas and rank of Hecke algebras. *Invent. Math.*, 223:485–595, 2021. ↑43.

[Man72] Yu. I. Manin. Parabolic points and zeta functions of modular curves. *Izv. Akad. Nauk. CCCP*, 36(1):19–66, 1972. ↑17.

[Maz72] B. Mazur. Rational points of abelian varieties with values in towers of number fields. *Invent. Math.*, 18:183–266, 1972. ↑20.

[Maz77a] B. Mazur. Modular curves and the Eisenstein ideal. *IHÉS Publ. Math.*, 47:33–186, 1977. ↑11, 12, 20, 28.

[Maz77b] B. Mazur. Rational points on modular curves. In *Modular functions in one variable V*, volume 601 of *Lecture Notes in Math.*, pages 107–148. Springer-Verlag, 1977. ↑11.

[Maz78] B. Mazur. Rational isogenies of prime degree (with an appendix by D. Goldfeld). *Invent. Math.*, 44(2):129–162, 1978. ↑11, 12, 19, 20, 27, 28.

[Mer96] Loï c Merel. Bornes pour la torsion des courbes elliptiques sur les corps de nombres. *Invent. Math.*, 124(1-3):437–449, 1996. ↑42.

[Mes81] J.-F. Mestre. Corps euclidiens, unités exceptionelles et courbes elliptiques. *J. Number Theory*, 13:123–137, 1981. ↑14.

[Mil80] J. Milne. *Étale cohomology*. Princeton University Press, 1980. ↑5, 21, 45, 46.

[Mil06] J.S. Milne. *Elliptic Curves*. BookSurge Publishers, 2006. ↑5.

[Mor22] L. J. Mordell. On the rational solutions of the indeterminate equations of the third and fourth degrees. *Proc. Cambridge Phil. Soc.*, 21:179–192, 1922. ↑4.

[MT73] B. Mazur and J. Tate. Points of order 13 on elliptic curves. *Invent. Math.*, 22:41–49, 1973. ↑20, 23.

[Ogg71] A. Ogg. Rational points on finite order on elliptic curves. *Invent. Math.*, 12:105–111, 1971. ↑23.

[Ogg75] A. Ogg. Diophantine equations and modular forms. *Bull. Amer. Math. Soc.*, 81:14–27, 1975. ↑19.

[Oht13] M. Ohta. Eisenstein ideals and the rational torsion subgroups of modular Jacobian varieties. *J. Math. Soc. Japan*, 65(3):733–772, 2013. ↑19.

[Reb09] M. Rebolledo. Merel's theorem on the boundedness of the torsion of elliptic curves. *Clay Mathematics Proceedings*, 8:71–82, 2009. ↑43.

[Shi71] G. Shimura. *Introduction to the arithmetic theory of automorphic functions.* Publications of the Math. Soc. of Japan, 1971. ↑35, 36.

[Sil09] J. Silverman. *The arithmetic of elliptic curves, 2nd edition*, volume 106 of *GTM.* Springer-Verlag, 2009. ↑5.

[Smi78] H.J.S. Smith. On the singularities of the modular equations and curves. *Proc. London Math. Soc.*, IX(1):242–276, 1878. ↑30.

[Sut] ↑31.

[Tat66] J. Tate. On the conjecture of Birch and Swinnerton-Dyer and a geometric analog. *Séminaire Bourbaki*, 306:415–440, 1966. ↑6, 7.

[Tat72] J. Tate. Algorithm for determining the type of a singular fiber in an elliptic pencil. In B. Birch and W. Kuyk, editors, *Modular functions in one variable IV*, pages 33–52. Springer-Verlag, 1972. ↑9.

[TO70] J. Tate and F. Oort. Group schemes of prime order. *Ann. Sci. ENS*, 3(1):1–21, 1970. ↑21, 24.

[TW95] R. Taylor and A. Wiles. Ring-theoretic properties of certain Hecke algebras. *Ann. of Math. (2)*, 141(3):553–572, 1995. ↑7.

[vB21] R. van Bommel. Efficient computation of bsd invariants in genus 2 efficient computation of bsd invariants in genus 2 efficient computation of BSD invariants in genus 2. In *Arithmetic Geometry, Number Theory, and Computation*, Simons Symp., pages 237–258, 2021. ↑7.

[vB22] R. van Bommel. Numerical verification of the birch and swinnerton-dyer conjecture for hyperelliptic curves of higher genus over $\mathbb{Q}$ up to squares. *Experiment. Math.*, 31(1):138–145, 2022. ↑7.

[Web98] H. Weber. *Lehrbuch der Algebra, vol 3.* AMS Chelsea Publishing, 1898. ↑33.

[Wei29] A. Weil. L'arithmétique sur les courbes algébriques. *Acta Mathematica*, 52(1):281–315, 1929. ↑4.

[Wei76] A. Weil. *Elliptic functions according to Eisenstein and Kronecker.* Classics in Mathematics. Springer-Verlag, Berlin, 1976. ↑29.

[Wil95] A. Wiles. Modular elliptic curves and Fermat's Last Theorem. *Ann. of Math. (2)*, 141(3):443–551, 1995. ↑7.

[WWE20] P. Wake and C. Wang-Erickson. The rank of Mazur's Eisenstein ideal. *Duke Math. J.*, 169(1):31–115, 2020. ↑43.

[Zag83] D. Zagier. Letter to gross, February 1983. ↑39.