Chapter 5

L-functions

In this chapter we define the L-functions attached to elliptic curves and modular forms, and we investigate when an elliptic curve and a modular form could have the same L-function.

5.1. The L-function of an elliptic curve

Let E be an elliptic curve over $\mathbb{Q}$ given by a minimal model (as in Definition 2.6.3):

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with coefficients $a_i \in \mathbb{Z}$. For p a prime in $\mathbb{Z}$ of good reduction for $E/\mathbb{Q}$, we define N_p as the number of points in the reduction of the curve modulo p, i.e., the number of points in $E(\mathbb{F}_p)$. In other words, N_p is the number of points in

$$\{O\} \cup \{(x,y) \in {\mathbb{F}_p}^2 \colon y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6 \equiv 0 \bmod p\}$$

where O is the point at infinity (see Section 2.6 and, in particular, Hasse's theorem 2.6.11). Also, let $a_p = p + 1 - N_p$. We define the

local factor at p of the L-series to be

$$L_p(T) = \begin{cases} 1 - a_pT + pT^2, \text{ if } E \text{ has good reduction at } p, \\ 1 - T, \text{ if } E \text{ has split multiplicative reduction at } p, \\ 1 + T, \text{ if } E \text{ has non-split multiplicative reduction at } p, \\ 1, \text{ if } E \text{ has additive reduction at } p. \end{cases}$$

Definition 5.1.1. The L-function of the elliptic curve E is defined to be

$$L(E, s) = \prod_{p \geq 2} \frac{1}{L_p(p^{-s})},$$

where the product is over all primes $p \geq 2$ and $L_p(T)$ is the local factor defined above. $L(E, s)$ is sometimes called the Hasse-Weil L-function of $E/\mathbb{Q}$.

Remark 5.1.2. The product that defines $L(E, s)$ converges and gives an analytic function for all $\Re(s) > 3/2$. This follows from Hasse's bound (Theorem 2.6.11), which implies that $|a_p| \leq 2\sqrt{p}$. However, far more is true. Indeed, mathematicians conjectured that $L(E, s)$ should have an analytic continuation to the whole complex plane and that it must satisfy a functional equation relating the values of $L(E, s)$ and $L(E, 2-s)$. For the precise functional equation see Theorem 5.1.9 below.

Example 5.1.3. Let $E/\mathbb{Q}$ be the elliptic curve with equation

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

This is a minimal model for $E/\mathbb{Q}$, and its discriminant is $\Delta_E = -11^5$. Therefore, $p = 11$ is the only prime of bad reduction for $E/\mathbb{Q}$, and the reduction is split multiplicative (see the discussion about E_3 in Example 2.6.7). Therefore,

$$L(E, s) = \left(\frac{1}{1 - 11^{-s}}\right) \cdot \prod_{\substack{p \geq 2 \\ p \neq 11}} \frac{1}{1 - a_p p^{-s} + p^{1-2s}}.$$

When expanded, the L-series attached to E has the form

$$L(E, s) = 1 - \frac{2}{2^s} - \frac{1}{3^s} + \frac{2}{4^s} + \frac{1}{5^s} + \frac{2}{6^s} - \frac{2}{7^s} - \frac{2}{9^s} - \frac{2}{10^s} + \frac{1}{11^s} + \cdots.$$

In general, one can always write $L(E, s) = \sum_{n \geq 1} a_n n^{-s}$, where the a_n are characterized in Proposition 5.1.5 below. ■

Example 5.1.4. Let $E/\mathbb{Q} : y^2 = x^3 - 11x^2 + 385$. The curve E has bad additive reduction at 2 and 11, split multiplicative at 5 and non-split multiplicative at 7 and 461. Thus, by definition

$$\begin{aligned} L(E,s) &= \left((1-5^{-s})(1+7^{-s})(1+461^{-s})\right)^{-1} \\ &\quad \cdot \prod_{\substack{\text{primes } p \\ p\neq 2,5,7,11,461}} \frac{1}{1-a_p p^{-s}+p^{1-2s}} \\ &= 1 - \frac{2}{3^s} + \frac{1}{5^s} - \frac{1}{7^s} + \frac{1}{9^s} + \frac{2}{13^s} - \frac{2}{15^s} - \frac{5}{17^s} + \frac{2}{21^s} \cdots . \end{aligned}$$

Proposition 5.1.5. *Let $E/\mathbb{Q}$ be an elliptic curve, and let $L(E,s)$ be its L-function. Define Fourier coefficients a_n for all $n \geq 1$ as follows. Let $a_1 = 1$. If $p \geq 2$ is prime, we define*

$$a_p = \begin{cases} p+1-N_p & \text{if } E \text{ has good reduction at } p; \\ 1 & \text{if } E \text{ has split multiplicative reduction at } p; \\ -1 & \text{if } E \text{ has non-split multiplicative reduction at } p; \\ 0 & \text{if } E \text{ has additive reduction at } p. \end{cases}$$

If $n = p^r$ for some $r \geq 1$, we define a_{p^r} recursively using the relation

$$a_p \cdot a_{p^r} = a_{p^{r+1}} + p \cdot a_{p^{r-1}} \quad \text{if } E/\mathbb{Q} \text{ has good reduction at } p$$

and $a_{p^r} = (a_p)^r$ if $E/\mathbb{Q}$ has bad reduction at p. Finally, if $(m,n) = 1$, then we define $a_{mn} = a_m \cdot a_n$. Then the L-function of E can be written as the series

$$L(E,s) = \sum_{n\geq 1} \frac{a_n}{n^s}.$$

The proof is left as an exercise (Exercise 5.7.2).

Remark 5.1.6. Notice that the recurrence formula $a_p \cdot a_{p^r} = a_{p^{r+1}} + p \cdot a_{p^{r-1}}$ (and $a_{p^r} = (a_p)^r$ in the bad reduction case) is strikingly similar to the recurrence relation defining the Hecke operators T_{p^r} for $k = 2$, and also the recurrence relation satisfied by the eigenvalues of an eigenform (see Definition 4.4.8, Remark 4.4.13 and Exercise 4.5.16). This is one of the first pieces of evidence that the L-function of an elliptic curve may be connected to a modular form.

Before we write down the functional equation for $E/\mathbb{Q}$, we need one more ingredient: the conductor of $E/\mathbb{Q}$. For each prime $p \in \mathbb{Z}$, we define the quantity f_p as follows:

$$f_p = \begin{cases} 0, \text{ if } E \text{ has good reduction at } p, \\ 1, \text{ if } E \text{ has multiplicative reduction at } p, \\ 2, \text{ if } E \text{ has additive reduction at } p, \text{ and } p \neq 2, 3, \\ 2 + \delta_p, \text{ if } E \text{ has additive reduction at } p = 2 \textit{ or } 3, \end{cases}$$

where δ_p is a technical invariant (see [Sil94], Ch. IV, §10; the invariant δ_p describes whether there is wild ramification in the action of the inertia group at p of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the Tate module $T_p(E)$).

Definition 5.1.7. The *conductor* $N_{E/\mathbb{Q}}$ of $E/\mathbb{Q}$ is defined to be

$$N_{E/\mathbb{Q}} = \prod_p p^{f_p},$$

where the product is over all primes and the exponents f_p are defined as above.

Example 5.1.8. Let us see some examples of conductors.

(1) Let $E/\mathbb{Q}$: $y^2 + y = x^3 - x^2 + 2x - 2$. The primes of bad reduction for E are $p = 5$ and 7. The reduction at $p = 5$ is additive, while the reduction at $p = 7$ is multiplicative. Hence $N_{E/\mathbb{Q}} = 25 \cdot 7 = 175$.

(2) As we saw above, the curve $y^2 + y = x^3 - x^2 - 10x - 20$ has split multiplicative reduction at $p = 11$ and the reduction is good elsewhere. Thus, the conductor is 11.

(3) The curves $E_A : y^2 + y = x^3 - x$ and $E_B : y^2 + y = x^3 + x^2 - 23x - 50$ are two non-isomorphic curves with conductor equal to 37.

Theorem 5.1.9 (Functional equation)**.** *The L-series $L(E, s)$ has an analytic continuation to the entire complex plane, and it satisfies the following functional equation. Define*

$$\Lambda(E, s) = (N_{E/\mathbb{Q}})^{s/2}(2\pi)^{-s}\Gamma(s)L(E, s),$$

where $N_{E/\mathbb{Q}}$ is the conductor of E and $\Gamma(s) = \int_0^\infty t^{s-1}e^{-t}dt$ is the Gamma function. Then

$$\Lambda(E,s) = w \cdot \Lambda(E, 2-s) \quad \textit{with } w = \pm 1.$$

The number $w = w(E/\mathbb{Q})$ in the functional equation is usually called the *root number* of E, and it has an important conjectural meaning (see the next section on the Birch and Swinnerton-Dyer conjecture). Theorem 5.1.9 was proved in 1999, since it follows from the Taniyama-Shimura-Weil conjecture 5.4.5, which was proved by work of Wiles, Taylor-Wiles, and Breuil, Conrad, Diamond and Taylor.

5.2. The Birch and Swinnerton-Dyer conjecture

Figure 1. Bryan Birch (left) and Sir Peter Swinnerton-Dyer (right). Photograph courtesy of William Stein.

Conjecture 5.2.1 (Birch and Swinnerton-Dyer)**.** *Let E be an elliptic curve over $\mathbb{Q}$, and let $L(E,s)$ be the L-function attached to E. Then:*

(1) *$L(E,s)$ has a zero at $s=1$ of order equal to the rank R_E of $E(\mathbb{Q})$. In other words, the Taylor expansion of $L(E,s)$ at $s=1$ is of the form*

$$L(E,s) = C_0 \cdot (s-1)^{R_E} + C_1 \cdot (s-1)^{R_E+1} + C_3 \cdot (s-1)^{R_E+2} + \cdots$$

where C_0 is a non-zero constant.

(2) *The residue of $L(E,s)$ at $s=1$, i.e., the coefficient C_0, has a concrete expression in terms of invariants of $E/\mathbb{Q}$. More concretely,*

$$C_0 = \lim_{s\to 1} \frac{L(E,s)}{(s-1)^{R_E}} = \frac{|Ш| \cdot \Omega_E \cdot \mathrm{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E_{\mathrm{torsion}}(\mathbb{Q})|^2}.$$

The invariants that appear in the conjectural formula for the residue are listed below:

- R_E is the (free) rank of $E(\mathbb{Q})$ (see Section 2.7).
- $\Omega_E = \int_{E(\mathbb{R})} \left|\frac{dx}{y}\right|$ is either the real period or twice the real period of a minimal model for E, depending on whether $E(\mathbb{R})$ is connected.
- $|Ш|$ is the order of the Shafarevich-Tate group of $E/\mathbb{Q}$ (we defined the 2-torsion of Sha, $Ш_2$, in Section 2.11).
- $\mathrm{Reg}(E/\mathbb{Q})$ is the elliptic regulator of $E(\mathbb{Q})$, as in Definition 2.8.4.
- $|E(\mathbb{Q})_{\mathrm{torsion}}|$ is the number of torsion points on $E/\mathbb{Q}$, including the point at infinity $\mathcal{O}$ (see Section 2.5).
- c_p is an elementary local factor, equal to the cardinality of $E(\mathbb{Q}_p)/E_0(\mathbb{Q}_p)$, where $E_0(\mathbb{Q}_p)$ is the set of points in $E(\mathbb{Q}_p)$ whose reduction modulo p is non-singular in $E(\mathbb{F}_p)$. Notice that if p is a prime of good reduction for $E/\mathbb{Q}$, then $c_p = 1$, so $c_p \neq 1$ only for finitely many primes p. The number c_p is called the *Tamagawa number* of E at p.

In 1974 ([Tat74], p. 198), John Tate wrote about the BSD conjecture: "*This remarkable conjecture relates the behavior of a function L at a point where it is not at present known to be defined ($s=1$) to the order of a group* (Ш) *which is not known to be finite!*" Tate is referring to the fact that, when the conjecture was first proposed, the analytic continuation of $L(E,s)$ was not known, and we did not know whether Ш was ever finite (nowadays we know many examples where Ш is finite, but it is still not known for all elliptic curves).

Example 5.2.2. Let $E/\mathbb{Q}$ be an elliptic curve. By Theorem 5.1.9, the function $L(E,s) = \sum_{n\geq 1} a_n n^{-s}$ has an analytic continuation to $\mathbb{C}$. In particular, if we restrict our attention to real values t, then $L(E,t)$

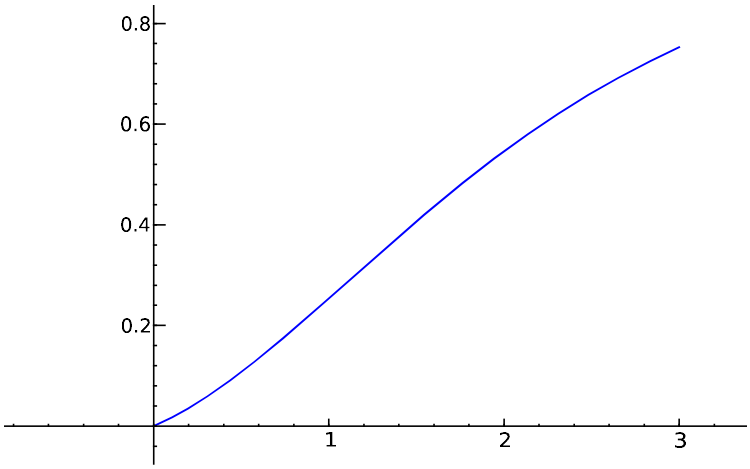


Figure 2. $L(E_0, t)$ for $E_0 : y^2 + y = x^3 - x^2 - 10x - 20$ and $-1 \leq t \leq 3$.

is a real-valued function. Since $L(E, s)$ is analytic, $L(E, t)$ should be continuous and (infinitely) differentiable. Let E_r, for $r = 0, 1, 2$ and 3, be elliptic curves defined by

$$\begin{aligned} E_0 &: y^2 + y = x^3 - x^2 - 10x - 20, \quad E_1 : y^2 + y = x^3 - x \\ E_2 &: y^2 + y = x^3 + x^2 - 2x, \quad E_3 : y^2 + y = x^3 - 7x + 6. \end{aligned}$$

The reader can check that the rank of E_r is precisely r. In Figures 2 through 5 we show the graphs of $L(E_r, t)$ for $-1 \leq t \leq 3$. Notice that the function $L(E_r, t)$ seems to have a zero of order r at $t = 1$, in agreement with the BSD conjecture. ■

Example 5.2.3. Let $E/\mathbb{Q} : y^2 = x^3 - 1156x$. Recall that in Examples 2.10.4 and 2.11.2 we calculated $R_E = 2$, $E(\mathbb{Q})_{\text{torsion}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $Ш_2 = \{(1, 1)\}$ (here $Ш_2$ is just the 2-torsion of Ш). A non-trivial calculation yields $Ш = Ш_2 = \{(1, 1)\}$. Figure 6 provides the values of all the invariants that appear in the BSD conjecture. Thus,

$$\frac{|Ш| \cdot \Omega_E \cdot \text{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E(\mathbb{Q})_{\text{torsion}}|^2} = 6.3851519548\ldots.$$

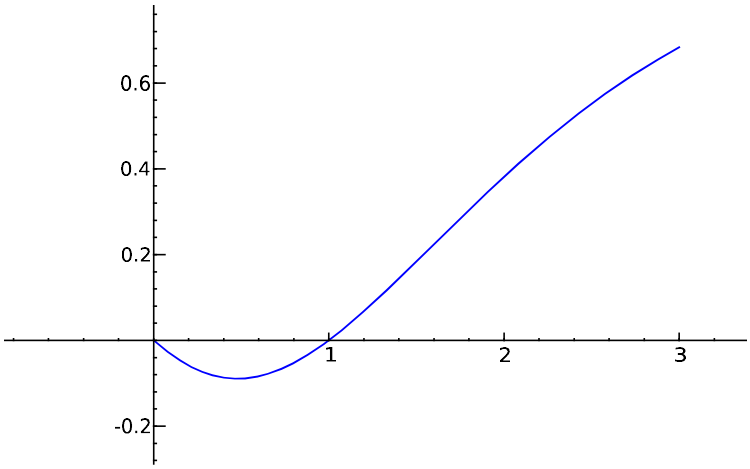


Figure 3. $L(E_1, t)$ for $E_1 : y^2 + y = x^3 - x$ and $-1 \leq t \leq 3$.

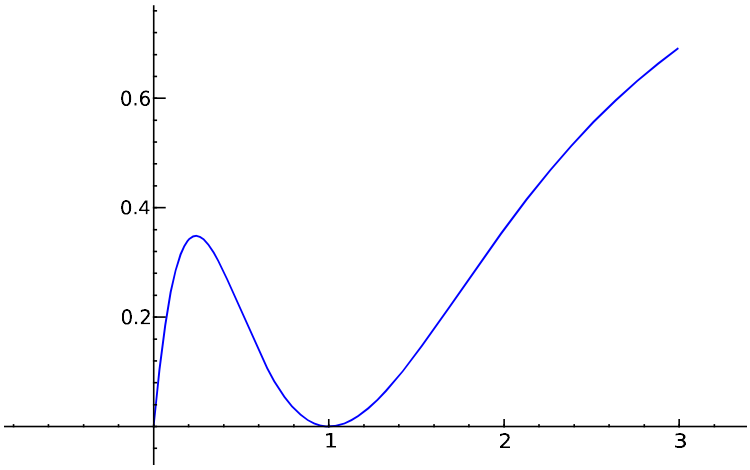


Figure 4. $L(E_2, t)$ for $E_2 : y^2 + y = x^3 + x^2 - 2x$ and $-1 \leq t \leq 3$.

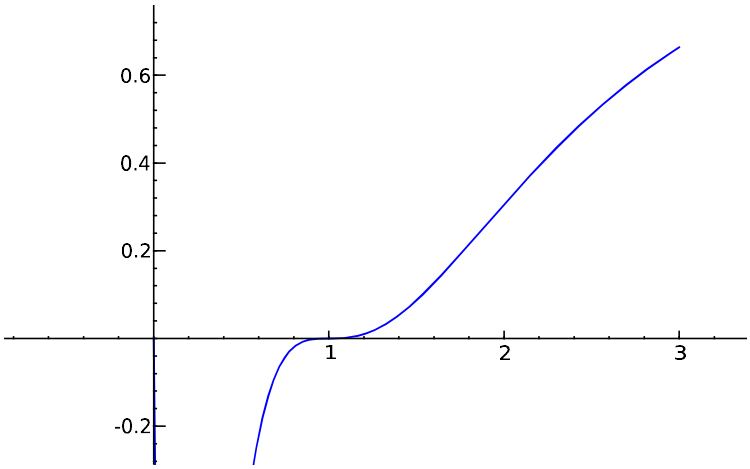


Figure 5. $L(E_3, t)$ for $E_3 : y^2 + y = x^3 - 7x + 6$ and $-1 \leq t \leq 3$.

$E/\mathbb{Q} : y^2 = x^3 - 1156x$	
R_E	2, $\langle P = (-16, 120), Q = (-2, 48)\rangle$
$\lvert Ш\rvert$	1
Ω_E	$0.8993583214\ldots$
$\mathrm{Reg}(E/\mathbb{Q})$	$\det \mathcal{H}(\{P, Q\}) = 7.0996751824\ldots$
$E(\mathbb{Q})_{\mathrm{torsion}}$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong \langle (0,0), (34,0)\rangle$
$\prod_{p \geq 2} c_p$	$c_2 \cdot c_{17} = 4 \cdot 4$

Figure 6. BSD data for the curve $E/\mathbb{Q} : y^2 = x^3 - 1156x$.

We can also calculate the value $L(E, 1)$ and the values of the derivatives $L'(E, 1)$ and $L''(E, 1)$; i.e., we can approximate numerically these values. For instance, one can use Sage (see Appendix A.3). For a technical description of the algorithms involved, see [**Dok04**]. Once we have calculated these values, we can write the first few terms

$E/\mathbb{Q} : y^2 = x^3 - 6724x$	
R_E	0
$\lvert Ш\rvert$	4
Ω_E	$0.5791156343\ldots$
$\mathrm{Reg}(E/\mathbb{Q})$	$\det \mathcal{H}(\{\}) = 1$
$E(\mathbb{Q})_{\mathrm{torsion}}$	$\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \cong \langle (0,0), (81,0) \rangle$
$\prod_{p \geq 2} c_p$	$c_2 \cdot c_{41} = 4 \cdot 4$

Figure 7. BSD data for the curve $E/\mathbb{Q} : y^2 = x^3 - 6724x$.

of the Taylor expansion of $L(E,s)$ around $s = 1$.

$$\begin{aligned} L(E,s) &\approx 9.508 \cdot 10^{-24} - (2.374 \cdot 10^{-23}) \cdot (s-1) \\ &\quad + (6.3851519548) \cdot (s-1)^2 + \cdots . \end{aligned}$$

Therefore, our approximate calculation suggests that $L(E,s)$ has a zero of order 2 at $s = 1$ and the residue is $6.3851519548\ldots$, in perfect agreement with the BSD conjecture (at least up to the given precision). ■

Example 5.2.4. Let $E/\mathbb{Q} : y^2 = x^3 - 6724x$. Recall that Examples 2.10.5 and 2.11.3 *suggest* that $R_E = 0$, $E(\mathbb{Q})_{\mathrm{torsion}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ and $\lvert Ш_2\rvert = 4$. A non-trivial calculation reveals that R_E is indeed 0 and $\lvert Ш\rvert = \lvert Ш_2\rvert = 4$. Figure 7 provides the values of all the invariants that appear in the BSD conjecture. Thus,

$$\frac{\lvert Ш\rvert \cdot \Omega_E \cdot \mathrm{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{\lvert E(\mathbb{Q})_{\mathrm{torsion}}\rvert^2} = 2.3164625374\ldots .$$

We can approximate the first few terms of the Taylor expansion of $L(E,s)$ around $s = 1$.

$$\begin{aligned} L(E,s) &\approx 2.3164625374 - (7.8248271660) \cdot (s-1) \\ &\quad + (25.7352635691) \cdot (s-1)^2 + \cdots . \end{aligned}$$

Therefore, our approximate calculation suggests that $L(E,s)$ does not vanish at $s = 1$, and $L(E,1) = 2.3164625374\ldots$, again in perfect agreement with the BSD conjecture. ■

The following is an easy consequence of the BSD conjecture (Exercise 5.7.3). Recall that the root number of E is the sign in the functional equation of $L(E, s)$.

Conjecture 5.2.5 (Parity Conjecture)**.** *The root number of E, denoted by $w = w(E/\mathbb{Q})$, indicates the parity of the rank of the elliptic curve; i.e., $w = 1$ if and only if the rank R_E is even, and $w = -1$ iff the rank is odd. Equivalently,*

$$w = (-1)^{\operatorname{ord}_{s=1} L(E,s)} = (-1)^{\operatorname{rank}(E(\mathbb{Q}))}$$

or $\operatorname{ord}_{s=1} L(E, s) \equiv \operatorname{rank}(E(\mathbb{Q})) \bmod 2$.

See Exercise 5.7.3.

Definition 5.2.6. Let $E/\mathbb{Q}$ be an elliptic curve, and let $L(E, s)$ be the L-function attached to E. The *analytic rank* of $E(\mathbb{Q})$ is defined to be the order of vanishing of $L(E, s)$ at $s = 1$, i.e.,

$$\operatorname{rank}_{\mathrm{an}}(E/\mathbb{Q}) := \operatorname{ord}_{s=1} L(E, s).$$

In other words, $\operatorname{rank}_{\mathrm{an}}(E/\mathbb{Q})$ is the order of the zero of $L(E, s)$ at $s = 1$.

Thus, the first part of the BSD conjecture is the statement that the analytic rank equals the (algebraic) free rank of the Mordell-Weil group $E(\mathbb{Q})$.

Example 5.2.7. Let $E/\mathbb{Q} : y^2 = x^3 - 157^2 x$. Recall that Proposition 1.1.3 says that the rational points on $E/\mathbb{Q}$ with $y \neq 0$ give right triangles of area 157, so if we find a single non-trivial point on E we prove that $n = 157$ is a congruent number (as defined in Example 1.1.2).

Comparing values of $\Lambda(s)$ and $\Lambda(2 - s)$, we calculate the root number $w = w(E/\mathbb{Q}) = -1$. Thus, the parity conjecture suggests that $E(\mathbb{Q})$ has odd rank, therefore ≥ 1, and so $E(\mathbb{Q})$ must be infinite. However, a computer search only yields the trivial 2-torsion points $(0, 0)$, $(157, 0)$ and $(-157, 0)$. We can calculate values of $L(E, s)$ and its derivatives at $s = 1$ and write down an approximate Taylor expansion:

$$L(E, s) \approx (11.4259445007) \cdot (s - 1) - (49.9773214816) \cdot (s - 1)^2 + \cdots .$$

Hence, the BSD conjecture suggests that $R_E = 1$ and

$$\frac{|Ш| \cdot \Omega_E \cdot \mathrm{Reg}(E/\mathbb{Q}) \cdot \prod_p c_p}{|E(\mathbb{Q})_{\mathrm{torsion}}|^2} = 11.4259445007\ldots. \tag{5.1}$$

If we believe that $R_E = 1$ and we write P for a generator of $E(\mathbb{Q})$ modulo torsion, then one can show that $Ш_2$ must be trivial (and, in fact, Ш is trivial as well, but this is much tougher to prove). Some other invariants are easy to calculate:

$$\Omega_E = 0.4185259488\ldots, \ \prod_{p\geq 2} c_p = c_2 \cdot c_{157} = 2 \cdot 4, \ |E(\mathbb{Q})_{\mathrm{torsion}}| = 4.$$

However, $\mathrm{Reg}(E/\mathbb{Q}) = \langle P, P\rangle = 2 \cdot \widehat{h}(P)$ is difficult to calculate because we do not know P (here $\widehat{h}$ is the canonical height). But we can solve for $\mathrm{Reg}(E/\mathbb{Q})$ in Eq. (5.1) and obtain

$$\mathrm{Reg}(E/\mathbb{Q}) = 2 \cdot \widehat{h}(P) = 54.6008892938\ldots$$

and $\widehat{h}(P) = 27.3004446469\ldots$. That's a huge height! Recall that

$$\widehat{h}(P) \approx \frac{1}{2} \log \max\{\mathrm{num}(x(P)), \mathrm{den}(x(P))\}$$

and so $\max\{|\mathrm{num}(x(P))|, |\mathrm{denom}(x(P))|\} \approx e^{54.6} \approx 5.157 \cdot 10^{23}$. This calculation gives us a rough idea of the size of the numerator and denominator of the x coordinate. With the help of homogeneous spaces, and looking for points in the correct height range, we can succeed at finding P. Its coordinates $P = (x(P), y(P))$ are:

$$x(P) = -\frac{16613623166818526754080 4}{2825630694251145858025},$$

$$y(P) = \frac{167661624456834335404812111469782006}{150201095200135518108761470235125}$$

and the canonical height of P is precisely $27.3004446469\ldots$, as predicted by the Birch and Swinnerton-Dyer conjecture. ■

There has been a great amount of research on the BSD conjecture, but the progress in the general case over $\mathbb{Q}$ is minimal (a lot is known about BSD for elliptic curves over function fields). The conjecture has been verified for many elliptic curves (for instance, see [**GJPST09**], [**Mil10**]), but there is little evidence in the form of proven theorems. The following result is the strongest piece of evidence proved to date.

Theorem 5.2.8 (Gross-Zagier, Kolyvagin)**.** *Let $E/\mathbb{Q}$ be an elliptic curve of algebraic rank R_E. Suppose that the analytic rank of $E/\mathbb{Q}$ is ≤ 1, i.e., $\mathrm{ord}_{s=1} L(E,s) \leq 1$. Then:*

(1) *The first part of BSD holds for $E/\mathbb{Q}$, i.e.,*

$$R_E = \mathrm{rank}(E(\mathbb{Q})) = \mathrm{rank}_{an}(E/\mathbb{Q}) = \mathrm{ord}_{s=1} L(E,s).$$

(2) *The Shafarevich-Tate group Ш associated to $E/\mathbb{Q}$ is finite.*

5.3. The L-function of a modular (cusp) form

Let $N, k \geq 1$ and let $f(z)$ be a cusp form of weight $2k$ for the congruence subgroup $\Gamma_0(N)$, i.e., $f(z) \in S_{2k}(\Gamma_0(N))$ in the notation of Section 4.2 (and, in particular, Prop. 4.2.3). For any $N \geq 1$, the matrix $T = (1,1;0,1)$ belongs to $\Gamma_0(N)$, and therefore $f(z) = f(z+1)$ for all $z \in \mathbb{H}$. Moreover, $f(z)$ is a cusp form and so f vanishes at all the cusps of $\mathbb{H}^*/\Gamma_0(N)$, and in particular it vanishes at ∞. Hence $f(z)$ has a q-expansion expression of the form

$$f(z) = \sum_{n \geq 1} a_n q^n,$$

where $q = e^{2\pi i z}$ for some coefficients $a_n \in \mathbb{C}$.

Definition 5.3.1. The L-function attached to a cusp form $f(z) = \sum_{n\geq 1} a_n q^n \in S_k(\Gamma_0(N))$ is defined by

$$L(f,s) = \sum_{n \geq 1} a_n n^{-s} = \sum_{n \geq 1} \frac{a_n}{n^s} = a_1 + \frac{a_2}{2^s} + \frac{a_3}{3^s} + \cdots.$$

Example 5.3.2. Let $N = 11$ and $k = 1$. The space $M_2(\Gamma_0(11))$ is a 2-dimensional $\mathbb{C}$-vector space with basis elements $\{f, g\}$ given in Example 4.2.11. In particular, $S_2(\Gamma_0(11))$ is generated by

$$f(q) \quad = \quad q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + O(q^{11}),$$

where $q = e^{2\pi i z}$. Hence the L-function associated to f is

$$L(f,s) = 1 - \frac{2}{2^s} - \frac{1}{3^s} + \frac{2}{4^s} + \frac{1}{5^s} + \frac{2}{6^s} - \frac{2}{7^s} - \frac{2}{9^s} - \frac{2}{10^s} + \cdots.$$

The very attentive reader might recognize these few terms as the first few terms in the L-function $L(E,s)$ that appeared in Example 5.1.3, where $E/\mathbb{Q}$ is the elliptic curve with equation $y^2 + y = x^3 - x^2 -$

$10x - 20$. Are they truly the same L-series? Further calculations show that all terms agree as we increase the precision. We will see that the Taniyama-Shimura-Weil conjecture 5.4.5, i.e., the modularity theorem, implies that $L(E, s) = L(f, s)$. Notice that the conductor of $E/\mathbb{Q}$ is precisely $N = 11$, as we saw in Example 5.1.8. ■

Example 5.3.3. Let $N = 37$ and $k = 1$. In Example 4.2.12 we described the space $S_2(\Gamma_0(37))$ with basis elements $\{f, g\}$ given by the q-expansions

$$\begin{aligned} f(q) &= q + q^3 - 2q^4 - q^7 - 2q^9 + 3q^{11} - 2q^{12} - 4q^{13} + O(q^{16}), \\ g(q) &= q^2 + 2q^3 - 2q^4 + q^5 - 3q^6 - 4q^9 - 2q^{10} + 4q^{11} + O(q^{12}). \end{aligned}$$

The L-functions attached to f and g are

$$\begin{aligned} L(f, s) &= 1 + \frac{1}{3^s} - \frac{2}{4^s} - \frac{1}{7^s} - \frac{2}{9^s} + \frac{3}{11^s} - \frac{2}{12^s} - \frac{4}{13^s} + \dots, \\ L(g, s) &= \frac{1}{2^s} + \frac{2}{3^s} - \frac{2}{4^s} + \frac{1}{5^s} - \frac{3}{6^s} - \frac{4}{9^s} - \frac{2}{10^s} + \frac{4}{11^s} + \dots. \end{aligned}$$

Now, let E_A and E_B be the elliptic curves of conductor 37 described in Example 5.1.8. Then

$$L(E_B, s) = 1 + \frac{1}{3^s} - \frac{2}{4^s} - \frac{1}{7^s} - \frac{2}{9^s} + \frac{3}{11^s} - \frac{2}{12^s} - \frac{4}{13^s} + \dots$$

and, indeed, we shall see that $L(f, s) = L(E_B, s)$. How about E_A?

$$L(E_A, s) = 1 - \frac{2}{2^s} - \frac{3}{3^s} + \frac{2}{4^s} - \frac{2}{5^s} + \frac{6}{6^s} - \frac{1}{7^s} + \frac{6}{9^s} + \frac{4}{10^s} - \frac{5}{11^s} + \dots$$

so $L(E_A, s) \neq L(g, s)$ or $L(f, s)$. Is there some form $F(z) \in S_k(\Gamma_0(37))$ such that $L(E_A, s) = L(F, s)$? If so, $F(q)$ must be a linear combination $\lambda \cdot f(q) + \mu \cdot g(q)$ for some $\lambda, \mu \in \mathbb{C}$. After a quick look at the first few coefficients of the q-expansions of f and g, and those of the series $L(E_A, s)$, one can check that, if some F works, then it must be $F(q) = f(q) - 2g(q)$, and indeed

$$(f - 2g)(q) = 1 - 2q^2 - 3q^3 + 2q^4 - 2q^5 + 6q^6 - q^7 + 6q^9 + 4q^{10} - 5q^{11} + O(q^{12})$$

and so

$$L(f - 2g, s) = 1 - \frac{2}{2^s} - \frac{3}{3^s} + \frac{2}{4^s} - \frac{2}{5^s} + \frac{6}{6^s} - \frac{1}{7^s} + \frac{6}{9^s} + \frac{4}{10^s} - \frac{5}{11^s} + \dots$$

Once again, we shall see that the Taniyama-Shimura-Weil conjecture implies the equality $L(f - 2g, s) = L(E_A, s)$. ■

5.4. The Taniyama-Shimura-Weil conjecture

In Examples 5.3.2 and 5.3.3, we have seen examples of elliptic curves $E/\mathbb{Q}$ of conductor N and modular forms $f \in S_2(\Gamma(N))$ such that the L-functions $L(E,s)$ and $L(f,s)$ seem to be identical.

Definition 5.4.1. We say that an elliptic curve $E/\mathbb{Q}$ is *modular* if there is a cusp form $f(z)$ such that

$$L(E,s) = L(f,s).$$

In the second half of the 20th century, many mathematicians grew increasingly interested in the question of whether every elliptic curve over $\mathbb{Q}$ is modular. However, early on, it was noticed that not every cusp form comes from an elliptic curve.

Notice that if E is modular and $L(E,s) = L(f,s) = \sum_{n\geq 1} a_n n^{-s}$, then a_p must equal $p+1-N_p$ when p is a prime of good reduction for E and, in general, a_n must coincide with those values defined in Proposition 5.1.5. Hence, for a given elliptic curve, there is a clear candidate for a cusp form f associated to the elliptic curve E.

Definition 5.4.2. Let $E/\mathbb{Q}$ be an elliptic curve. We define the *potential cusp form* associated to E to be a function $f_E : \mathbb{H} \to \mathbb{C}$ defined by its q-expansion

$$f_E(q) = \sum_{n\geq 1} a_n q^n,$$

where $q = e^{2\pi i z}$ and the a_n are defined in Proposition 5.1.5 (for instance, if $E/\mathbb{Q}$ has good reduction at p, then $a_p = p+1-N_p$).

It is *very far from clear* that f_E is a modular form. Let us suppose for a moment that f_E is indeed a modular form and $L(E,s) = L(f_E,s)$. What kind of modular form should f_E be?

(1) The examples suggest that, first of all, f_E must be a cusp form of weight 2 for $\Gamma_0(N)$, where $N = N_E$ is the conductor of $E/\mathbb{Q}$;

(2) If $L(E,s) = L(f_E,s)$, then, by the functional equation for $L(E,s)$ in Theorem 5.1.9, the L-function associated to f_E, that is $L(f_E,s)$, must also satisfy a functional equation;

(3) If $L(E,s) = L(f_E,s)$, then $L(f_E,s)$ must have an Euler product, since $L(E,s)$ has one. We say that $L(s) = \sum_{n\geq 1} a_n n^{-s}$ has an Euler product if it can be written as a product $L(s) = \prod_{p\geq 2} L_p(s)$ over all primes $p \geq 2$. Clearly, $L(E,s)$ is defined as an Euler product, so $L(f_E,s)$ must have an Euler product as well.

The work of Hecke characterizes which cusp forms in $S_2(\Gamma_0(N))$ satisfy a functional equation and which cusp forms have an Euler product. Recall that in Proposition 4.4.2 we defined ± 1-spaces of S_2 such that

$$S_2(\Gamma_0(N)) = S_2^+(\Gamma_0(N)) \oplus S_2^-(\Gamma_0(N)).$$

Theorem 5.4.3 (Hecke; [DS05], §5.10)**.** *Let $N, k \geq 1$ and $f(z) \in S_{2k}(\Gamma_0(N))$ be a cusp form such that $f(z)$ is an eigenvector for the operator w_N, i.e., $f(z) \in S_{2k}^{\varepsilon}(\Gamma_0(N))$ for $\varepsilon = +1$ or -1. Then $L(f,s)$ has an analytic continuation to $\mathbb{C}$. Moreover, if we define*

$$\Lambda(f,s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s),$$

where $\Gamma(s)$ is the Gamma function, then $\Lambda(f,s)$ satisfies the functional equation

$$\Lambda(f,s) = \varepsilon \cdot \Lambda(f, 2-s).$$

Recall (Definition 4.4.10) that we say that $f(z) = \sum_{n\geq 0} a_n q^n$ is an eigenform if f is an eigenvector **for all** Hecke operators T_n, $n \geq 1$, simultaneously. We say that $f(z)$ is a normalized eigenform if $a_1 = 1$.

Theorem 5.4.4 (Hecke; [DS05], §5.9)**.** *Let $N, k \geq 1$. Let $f(z)$ be a normalized eigenform of weight $2k$ for $\Gamma_0(N)$ such that $T_p(f) = \lambda_p \cdot f$ for every prime $p \geq 2$. Then $L(f,s)$ has an Euler product of the form*

$$L(f,s) = \prod_{p|N} \frac{1}{1-\lambda_p p^{-s}} \prod_{p\nmid N} \frac{1}{1-\lambda_p p^{-s} + p^{2k-1-2s}}.$$

Now we may use Hecke's theorems to narrow down which cusp forms may be associated to elliptic curves. Suppose that $E/\mathbb{Q}$ is an elliptic curve with conductor N and let us assume that the potential cusp form f_E associated to E is indeed a cusp form. Then f_E must verify the following properties:

(1) $f_E(z) \in S_2(\Gamma_0(N))$. The level of $f_E(z)$ should be precisely N and not lower; otherwise f would correspond to a curve of lower conductor. Thus, we require $f_E(z) \in S_2^{\text{new}}(\Gamma_0(N))$. Note that the functional equation of f_E determines N, the conductor/level.

(2) $f_E(z)$ must be in one of the ε-spaces of cusp forms, i.e.,

$$f_E \in S_2^{\text{new}}(\Gamma_0(N)) \cap S_2^{\varepsilon}(\Gamma_0(N))$$

for $\epsilon = +1$ or -1.

(3) $f_E(z)$ must be a normalized eigenform in $S_2^{\text{new}}(\Gamma_0(N))$, and it needs to be an eigenvector for w_N as well. Therefore, $f_E(z)$ is a normalized newform (Definition 4.4.18).

Taniyama, Shimura and Weil are credited with the following formulation of the modularity conjecture.

Conjecture 5.4.5 (Taniyama-Shimura-Weil)**.** *A series of the form $L(s) = \sum_{n\geq 1} a_n n^{-s}$ with $a_n \in \mathbb{Z}$ is the L-function $L(E,s)$ of an elliptic curve $E/\mathbb{Q}$ of conductor N if and only if $L(s) = L(f,s)$ is the L-function of a normalized newform of weight 2 for $\Gamma_0(N)$.*

The conjecture of Taniyama, Shimura and Weil was proved in several stages.

- Eichler and Shimura ([**Shi73**], Ch. 7, Thm. 7.14) showed one of the directions of the equivalence in the conjecture: if $f(z)$ is a normalized newform of weight 2 for $\Gamma_0(N)$, then there exists an elliptic curve $E_f/\mathbb{Q}$ such that $L(f,s) = L(E_f,s)$.
- Wiles [**Wil95**] and Taylor and Wiles [**TW95**] proved the Taniyama-Shimura-Weil conjecture when $E/\mathbb{Q}$ is *semistable* (i.e., if the conductor N_E is square-free or, equivalently, when $E/\mathbb{Q}$ does not have any primes of bad additive reduction). This was the case that was needed to finalize the proof of Fermat's last theorem (see Section 5.5).
- Finally, Breuil, Conrad, Diamond and Taylor [**BCDT01**] showed that the conjecture is true for all elliptic curves over $\mathbb{Q}$.

The Taniyama-Shimura-Weil conjecture is nowadays frequently called the modularity theorem. We conclude this section with an important equivalent formulation of the TSW conjecture:

Theorem 5.4.6 (Modularity theorem)**.** *Let $E/\mathbb{Q}$ be an elliptic curve of conductor N, and let $X_0(N)$ be given by an algebraic model over $\mathbb{Q}$ (see Remark 3.6.4). Then there is a surjective algebraic map of curves $\Psi_{E,N} : X_0(N) \to E$ defined over $\mathbb{Q}$. (The map $\Psi_{E,N}$ is called a modular parametrization of E.)*

5.5. Fermat's last theorem

Theorem 5.5.1. *The equation $x^n + y^n = z^n$ has no solutions in integers x, y, z with $xyz \neq 0$, whenever $n > 2$.*

Figure 8. Andrew J. Wiles (right) and his Ph.D. advisor, John H. Coates (left).

Suppose that n, u, v and w are integers such that $n > 2$, $uvw \neq 0$ and

$$u^n + v^n = w^n.$$

Therefore, either n is divisible by 4, i.e., $n = 4k$ with $k \geq 1$, and $(u^k)^4 + (v^k)^4 = (w^k)^4$, or there is a prime divisor $p \geq 3$ of n, with

$n = ph$ and $h \geq 1$, such that $(u^h)^p + (v^h)^p = (w^h)^p$. Fermat showed that the equation $x^4 + y^4 = z^4$ has no solutions $x, y, z \in \mathbb{Z}$ with $xyz \neq 0$, so we conclude that $x^p + y^p = z^p$ must have an integer solution for some prime $p \geq 3$ and $xyz \neq 0$.

Thus, let us suppose that $p \geq 3$ and $a^p + b^p = c^p$, with $a, b, c \in \mathbb{Z}$ and $abc \neq 0$. However, we know that this is not possible for $p = 3, 5$ or 7.

- Leonhard Euler is generally credited for the proof of the $p = 3$ case (although his solution, in 1770, had a major gap). Kausler (1802), Legendre (1823) and many others have also published proofs of this case.
- The case of $p = 5$ was first shown (independently) by Legendre and Dirichlet, around 1825.
- The proof of Fermat's last theorem for $p = 7$ is due to Lamé, published in 1839.

Hence, we may assume that $p \geq 11$. It is worth pointing out that, in 1846, Ernst Kummer proved Fermat's last theorem for *regular* primes. Not all primes are regular: we know that there are infinitely many irregular primes (the first few irregular primes are 37, 59, 67, 101, 103, 131, 149, ...), but it is widely believed that there are also infinitely many regular primes. In 1984, the proof of Mordell's conjecture (now known as Faltings' theorem; see the paragraph on *Higher degree* in Section 2.1) was announced which shows that, for a fixed $n > 2$, $x^n + y^n = z^n$ may have at most a finite number of relatively prime integer solutions.

The strategy that led to the first (correct) proof of Fermat's last theorem was layed out by Frey [**Fre86**] and Serre [**Ser87**]. Let $p \geq 11$ and suppose a, b, c are relatively prime integers with $a^p + b^p = c^p$ and $abc \neq 0$. In 1984, Frey discovered that the elliptic curve

$$E : y^2 = x(x - a^p)(x + b^p)$$

would be semistable with conductor $N_E = \prod_{\ell | abc} \ell$ (see Exercise 5.7.5) and would satisfy some other technical properties. Moreover, Frey claimed that such a curve $E/\mathbb{Q}$ could not be modular; i.e., there is no weight 2 normalized newform $f \in S_2(\Gamma_0(N_E))$ such that

$L(f,s) = L(E,s)$. The problem with the modularity of E was made precise by Serre, and Ribet [**Rib90**] proved in 1986 that, indeed, E cannot be modular.

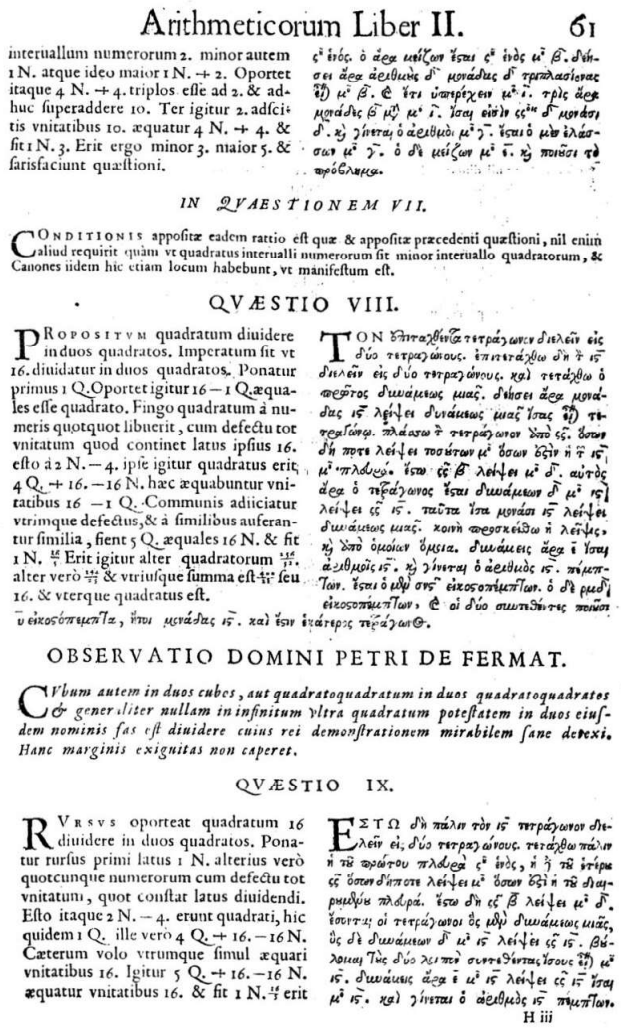
Finally, in 1995, Wiles [**Wil95**] and Taylor and Wiles [**TW95**] proved the Taniyama-Shimura-Weil conjecture 5.4.5 for all semistable elliptic curves $E/\mathbb{Q}$. Therefore, $E : y^2 = x(x-a^p)(x+b^p)$ would have to be modular if it existed. Hence, neither E nor the aforementioned solution (a,b,c) to $x^p + y^p = z^p$ can exist, and Fermat's last theorem holds.

5.6. Looking back and looking forward

The quest to find a proof of Fermat's last theorem lasted more than 350 years, and hundreds of mathematicians tried to attack the problem in many very different ways. It was simply a fantastic challenge that piqued the interest of essentially every mathematician from Fermat to Wiles. Still today, Fermat's last theorem captivates the imagination of math enthusiasts across the world. It is curious, though, that Fermat's last theorem has virtually no interesting consequences other than the statement itself.

However, the study of the solutions of such a simple equation ($x^n + y^n = z^n$) has been the driving force in developing an immense amount of extremely interesting mathematics. The statement of Fermat's last theorem may not have relevant corollaries, but the tools that were used in the proof are incredibly important and offer a vast range of very useful applications.

The final stages of the proof of Fermat's last theorem (as outlined in Section 5.5) represent one of the biggest triumphs of modern mathematics — not just because a 358-year-old problem was solved, but for the fundamental advances in the theory of elliptic curves and modular forms that were produced in order to verify Fermat's claim. This was no small enterprise; we have already briefly described the remarkable involvement of many important mathematicians (Shimura, Taniyama, Weil, Frey, Serre, Ribet, Wiles, Taylor, Breuil, Conrad, and Diamond, among many others). Just the proof of the modularity theorem (Theorem 5.4.6) occupies more than 200 pages of research articles (that's only counting [**Wil95**], [**TW95**] and [**BCDT01**]), and



Arithmeticorum Liber II. 61

interuallum numerorum 2. minor autem 1 N. atque ideo maior 1 N. + 2. Oportet itaque 4 N. + 4. triplos esse ad 2. & adhuc superaddere 10. Ter igitur 2. adsciti vnitatibus 10. æquatur 4 N. + 4. & fit 1 N. 3. Erit ergo minor 3. maior 5. & satisfaciunt quæstioni.

ς^ε ἑνός. ὁ ἄρα μείζων ἔσται ς^ε ἑνὸς μ^ε β̄. δεῖ-σει ἄρα ἀριθμὸν δ̄ μονάδας δ̄ τριπλασίονας ϛ̄δ̄ μ^ε β̄. ☞ ἔστι ὑπερέχειν μ^ε. ῑ. τρὶς ἄρα μονάδες β̄ μ^ε ῑ. ἴσαι εἰσὶν ις^ιν δ̄ μονάσι δ̄. καὶ γίνεται ὁ ἀριθμὸς μ^ε γ̄. ἔσται ὁ μὲν ἐλάσσων μ^ε γ̄. ὁ δὲ μείζων μ^ε ε̄. καὶ ποιοῦσι τὸ προβλημα.

IN QVAESTIONEM VII.

CONDITIONIS appositæ eadem ratio est quæ & appositæ præcedenti quæstioni, nil enim aliud requirit quàm vt quadratus interualli numerorum sit minor interuallo quadratorum, & Canones iidem hic etiam locum habebunt, vt manifestum est.

QVÆSTIO VIII.

PROPOSITVM quadratum diuidere in duos quadratos. Imperatum sit vt 16. diuidatur in duos quadratos. Ponatur primus 1 Q. Oportet igitur 16 — 1 Q. æquales esse quadrato. Fingo quadratum à numeris quotquot libuerit, cum defectu tot vnitatum quod continet latus ipsius 16. esto à 2 N. — 4. ipse igitur quadratus erit 4 Q. + 16. — 16 N. hæc æquabuntur vnitatibus 16 — 1 Q. Communis adiiciatur vtrimque defectus, & à similibus auferantur similia, fient 5 Q. æquales 16 N. & fit 1 N. $\frac{16}{5}$. Erit igitur alter quadratorum $\frac{256}{25}$. alter verò $\frac{144}{25}$ & vtriusque summa est $\frac{400}{25}$ seu 16. & vterque quadratus est.

ΤΟΝ δοθέντα τετράγωνον διελεῖν εἰς δύο τετραγώνους. ἐπιτετάχθω δὴ τὸν ις̄ διελεῖν εἰς δύο τετραγώνους. καὶ τετάχθω ὁ πρῶτος δυνάμεως μιᾶς. δεήσει ἄρα μονάδας ις̄ λείψει δυνάμεως μιᾶς ἴσας ϛ̄δ̄ τετραγώνῳ. πλάσω δὴ τετράγωνον ἀπὸ ς̄. ὅστων δὲ ἡ πρώτη λείψει τοσούτων μ^ε ὅσων ὀξεῖα ἡ ις̄ μ^ε πλευρά. ἔστω ς̄ β̄. λείψει μ^ε δ̄. αὐτὸς ἄρα ὁ τετράγωνος ἔσται δυνάμεων δ̄ μ^ε ις̄ λείψει ς̄ ις̄. ταῦτα ἴσα μονάσι ις̄ λείψει δυνάμεως μιᾶς. κοινὴ προσκείσθω ἡ λεῖψις, καὶ ἀπὸ ὁμοίων ὁμοῖα. δυνάμεις ἄρα ε̄ ἴσαι ἀριθμοῖς ις̄. καὶ γίνεται ὁ ἀριθμὸς ις̄. πέμπτων. ἔσται ὁ μὲν πρῶτος ις̄ εἰκοσοπέμπτων. ὁ δὲ δεύτερος εἰκοσοπέμπτων. ☞ οἱ δύο συντεθέντες ποιοῦσι τὸν εἰκοσόπεμπτα, ἤτοι μονάδας ις̄. καὶ ἔστιν ἑκάτερος τετράγωνος.

OBSERVATIO DOMINI PETRI DE FERMAT.

CVbum autem in duos cubos, aut quadratoquadratum in duos quadratoquadratos & generaliter nullam in infinitum vltra quadratum potestatem in duos eiusdem nominis fas est diuidere cuius rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet.

QVÆSTIO IX.

RVRSVS oporteat quadratum 16 diuidere in duos quadratos. Ponatur rursus primi latus 1 N. alterius verò quotcunque numerorum cum defectu tot vnitatum, quot constat latus diuidendi. Esto itaque 2 N. — 4. erunt quadrati, hic quidem 1 Q. ille verò 4 Q. + 16. — 16 N. Cæterum volo vtrumque simul æquari vnitatibus 16. Igitur 5 Q. + 16. — 16 N. æquatur vnitatibus 16. & fit 1 N. $\frac{16}{5}$ erit

ΕΣΤΩ δὴ πάλιν τὸν ις̄ τετράγωνον διελεῖν εἰς δύο τετραγώνους. τετάχθω πάλιν ἡ τῶ πρώτου πλευρὰ ς^ε ἑνὸς, ἡ δὲ τῆ ἑτέρα ς̄ ὅσων δεήσει λείψει μ^ε ὅσων ὀξεῖα ἡ τὸ διαιρεθὲν πλευρά. ἔστω δὴ ς̄ β̄ λείψει μ^ε δ̄. ἔσονται οἱ τετράγωνοι ὁ μὲν δυνάμεως μιᾶς, ὁ δὲ δυνάμεων δ̄ μ^ε ις̄ λείψει ις̄ ις̄. βούλομαι τοὺς δύο λείπειν συντεθέντας ἴσους ϛ̄δ̄ μ^ε ις̄. δυνάμεις ἄρα ε̄ μ^ε ις̄ λείψει ις̄ ις̄ ἴσαι μ^ε ις̄. καὶ γίνεται ὁ ἀριθμὸς ις̄ πεμπτῶν.

H iii

Figure 9. A 1670 edition of Diophantus' *Arithmetica*, which includes the original Greek text, a Latin translation, and Fermat's commentary: "Observatio Domini Petri de Fermat". In this page Fermat states his famous last theorem.

many books have been written to explain the brilliant mathematics developed for the proof (see [**CSS00**] for a graduate-level textbook).

Fermat's last theorem has been proved, but the broad areas of research that this book touches on (namely algebraic number theory, algebraic geometry and their intersection, arithmetic geometry) have seen an exponential growth over the last couple of centuries, and they continue to grow at a vigorous pace. Nowadays, there is an immense amount of research being done on elliptic curves, modular forms, and generalizations of the modularity theorem to other settings (abelian varieties, elliptic curves over number fields, etc.). Many questions remain unanswered; for instance,

- Are there elliptic curves over $\mathbb{Q}$ of arbitrarily high rank? See Conjecture 2.4.7 and the discussion in the same section.
- Is the Shafarevich-Tate group of an elliptic curve, $Ш(E/\mathbb{Q})$, always a finite group?
- Is the Birch and Swinnerton-Dyer conjecture true for all elliptic curves? See Conjecture 5.2.1 and Section 5.2. The Clay Mathematics Institute has offered a reward of one million dollars for a proof (or counterexample!) of this celebrated conjecture.

These are just three questions of great (huge!) interest to number theorists, but there are many other interesting questions and challenging problems being formulated as the reader stares at this page. The Preface to this book contains a list of suggested reading material so that the reader can continue to learn (more rigorously, and in depth) about elliptic curves, modular forms, and their L-functions.

5.7. Exercises

Exercise 5.7.1. Let $E/\mathbb{Q}$ be an elliptic curve and let $p \geq 2$ be a prime. Define $E^{\mathrm{ns}}(\mathbb{F}_p)$ to be the set of all non-singular points on $E(\mathbb{F}_p)$, and write $N_p^{\mathrm{ns}} = |E^{\mathrm{ns}}(\mathbb{F}_p)|$. For instance, if p is a prime of good reduction, then $E^{\mathrm{ns}}(\mathbb{F}_p) = E(\mathbb{F}_p)$ and $N_p^{\mathrm{ns}} = N_p = p + 1 - a_p$. Suppose that $E/\mathbb{Q}$ has bad reduction at p. Show that:

$$N_p^{\mathrm{ns}} = \begin{cases} p-1 & \text{if E has split multiplicative reduction at } p; \\ p+1 & \text{if E has non-split multiplicative reduction at } p; \\ p & \text{if E has additive reduction at } p. \end{cases}$$

Conclude that $L_p(p^{-1}) = \frac{N_p^{\mathrm{ns}}}{p}$ for every $p \geq 2$ (including good and bad primes), where the function $L_p(T)$ appears in Definition 5.1.1. (Hint: write $E : f(x,y) = 0$ and express $f(x,y) = ((y-y_0) - \alpha(x-x_0)) \cdot ((y-y_0) - \beta(x-x_0)) - (x-x_0)^3$ where (x_0, y_0) is the singular point for $E(\mathbb{F}_p)$. Exercise 2.12.11 shows that there is (at most) one singular point in $E(\mathbb{F}_p)$, at least for $p \geq 3$.)

Exercise 5.7.2. Prove Proposition 5.1.5. (Hint: $\frac{1}{1-x} = 1 + x + x^2 + \cdots = \sum_{n \geq 0} x^n$, and use the Fundamental Theorem of Arithmetic.)

Exercise 5.7.3. Prove the parity conjecture 5.2.5, assuming the Birch and Swinnerton-Dyer conjecture and the functional equation of $L(E,s)$. (Hint: use the Taylor expansion of $L(E,s)$ around $s = 1$.) Conclude that, if the root number $w(E/\mathbb{Q}) = -1$, then $E(\mathbb{Q})$ is infinite.

Exercise 5.7.4. Let $f(z) = \sum_{n \geq 1} a_n q^n$ be a cusp form in $S_k(\Gamma_0(N))$, and define the Mellin transform of $f(z)$ by

$$\widehat{f}(s) = \int_0^\infty f(iy) y^s \frac{dy}{y}.$$

Show that $\widehat{f}(s) = (2\pi)^{-s}\Gamma(s)L(f,s)$, where $\Gamma(s)$ is the Gamma function and $L(f,s)$ is the L-function attached to f. (You may ignore convergence issues and assume that integrals and infinite sums commute.)

Exercise 5.7.5. Let $p > 3$ be a prime and suppose that a, b, c are pairwise relatively prime integers such that $a^p + b^p = c^p$ and $abc \neq 0$. Let $E/\mathbb{Q}$ be the elliptic curve (Frey curve) defined by

$$E : y^2 = x(x - a^p)(x + b^p).$$

The goal of this exercise is to show that E is semistable with conductor $N_E = \prod_{\ell | abc} \ell$.

(1) Show that, after rearranging a, b and c if necessary, we can assume that $a \equiv 0 \bmod 2$ and $b \equiv c \equiv 1 \bmod 4$. (Hint: if $2|a$ and $b \equiv 3 \bmod 4$, consider $a^p + (-c)^p = (-b)^p$.)

(2) Calculate the discriminant Δ of $E/\mathbb{Q}$.

(3) Show that $E/\mathbb{Q}$ has good reduction at all primes ℓ that do not divide abc.

(4) Show that if $\ell \geq 3$ is a prime dividing abc, then $E/\mathbb{Q}$ has bad multiplicative reduction at ℓ.

(5) Show that $E/\mathbb{Q}$ has bad multiplicative reduction at $\ell = 2$. (Hint: use the following change of variables

$$x = \frac{X}{4}, \quad y = \frac{Y}{8} + \frac{3X}{8}$$

to find another model isomorphic to $E/\mathbb{Q}$. Show that this model has coefficients in $\mathbb{Z}$, and analyze the reduction at $\ell = 2$.)

(6) Conclude that the conductor of E is precisely $N_E = \prod_{\ell | abc} \ell$. (See Definition 5.1.7.)

Appendix A

PARI/GP and Sage

This appendix is meant as a brief introduction to the usage of the software packages PARI/GP and Sage, oriented to the study of elliptic curves and modular forms. The websites for these packages are:

- PARI/GP: http://pari.math.u-bordeaux.fr/
- Sage: http://www.sagemath.org/

but *notice* that you can call PARI/GP from Sage, so I would recommend simply installing Sage on your computer. I strongly recommend that you use the "notebook" option in Sage and interact with the software through your favorite internet browser (e.g. Firefox). Sage can also be found online (although the performance, usually, is slower than a local version on your computer):

- Sage online: http://www.sagenb.org/

Both packages have online manuals and specific sections on elliptic curves.

A.1. Elliptic curves

A.1.1. Definition of an Elliptic Curve. An elliptic curve is a plane curve E given by a Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with coefficients $a_1, \ldots, a_6$ in some field F. If the field is of characteristic different from 2 or 3, one can find an easier model of the form

$$y^2 = x^3 + Ax + B.$$

In order to work with elliptic curves using the software packages, we need to define the curves first:

- `GP > E = ellinit`$([a_1, a_2, a_3, a_4, a_6])$
- `Sage > E = EllipticCurve`$([a_1, a_2, a_3, a_4, a_6])$
- or `Sage > E = EllipticCurve`$([A, B])$.

Once we have defined an elliptic curve E, we can calculate basic quantities such as the discriminant, the j-invariant or any of the coefficients b_i or c_j (as defined in [Sil86], Ch. III, §1):

- In `GP`, type `E.disc`, `E.c4` or `E.j`,
- In `Sage`, type `E.discriminant()`, `E.c4()` or `E.j_invariant()`.

If the elliptic curve is given by a model of the form $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ but you would rather have a model $y^2 = x^3 + Ax + B$, use the command `E.integral_short_weierstrass_model()`.

Remark A.1.1. Perhaps **the two most useful Sage tricks** are the "Tab" key after an object and "?" after a command to get help. For instance, if we have defined an elliptic curve `E`, then typing

```
E.
```

followed by the "Tab" key displays all possible commands that one can use with an elliptic curve. This is very useful when we do not remember the exact syntax or we are wondering if Sage is capable of doing some particular operation on `E`. Similarly, if we want to know more about the usage of a particular command, then "`E.command_name?`" will display a help box. For example, if we input `E.discriminant?` then Sage tells us that this command returns the discriminant of `E` and provides a couple of examples for the user.

A.1.2. Basic operations. Let us start by using the addition on an elliptic curve. Let E be the curve given by $Y^2 = X^3 + 1$, and suppose we have initialized E as above. This curve has points $P = [0, 1]$ and

$Q = [-1, 0]$. Let us find $P + Q$ and $2P$ (the answers are $[2, -3]$ and $[0, -1]$ respectively). The commands are:

- In GP, the commands are `elladd(E,[0,1],[-1,0])` and, in order to find $2P$, one types `ellpow(E,[0,1],2)`;
- Sage: First we create points on the curve: `P = E([0,1]); Q = E([-1,0])` and now we can do addition: type `P+Q` and `P+P`, or calculate multiples by typing `2*P`, `3*P`, etc.

Notice that Sage will transform affine points to projective coordinates (e.g., `P = E([0,1])` returns `(0 : 1 : 1)` in Sage). If you want to find points on a curve (up to a given bound B on the height of the point), use `E.point_search(B)` in Sage.

A.1.3. Plotting. Here is an example of a 2D-plot with Sage:

```
E = EllipticCurve([0,0,0,0,1]);
Ep = plot(E, -1,2.5,thickness=2);
p1=(2,3); p2=(0,1); p3=(-1,0); p4=(0,-1); p5=(2,-3);
L1=line([p1,p3],rgbcolor=(1,0,0));
L2=line([p5,p3],rgbcolor=(1,0,0));
L3=line([p4,p3],rgbcolor=(1,0,0));
L4=line([p2,p5],rgbcolor=(1,0,0));
L5=line([p4,p1],rgbcolor=(1,0,0));
T1=text('P',[2,3.5]); T2=text('2P',[0.15,1.5]);
T3=text('3P',[-1,.5]); T4=text('4P',[0.15,-1.5]);
T5=text('5P',[2,-3.5]);
P=point([p1,p2,p3,p4,p5],pointsize=30,
rgbcolor=(0,0,0));
PLOT=Ep+T1+T2+T3+T4+T5+L1+L2+L3+L4+L5+P; show(PLOT)
```

The result is the graph that appears in Figure 3. The following is an alternative way to plot points on a curve:

```
Q = E(2,3);
Qplot = plot(Q, pointsize=30)+plot(2*Q, pointsize=30);
show(Qplot)
```

A.1.4. Good and bad reduction. Given a prime p and an elliptic curve $E/\mathbb{Q}$ given by a Weierstrass equation with integer coefficients, we can consider E as a curve over $\mathbb{Z}/p\mathbb{Z}$. The primes that divide the (minimal) discriminant are called bad primes or primes of bad reduction. In Sage, you can find the minimal model of an elliptic curve E by typing `E.minimal_model()`. For example, in Sage, the commands

```
E=EllipticCurve([0,5,0,0,35]);
prime_divisors(E.discriminant())
```

will return `[2,5,7,17]`. You may also use

```
factor(E.discriminant()).
```

Then one can use the command `kodaira_type()` to find out the precise type of reduction: `I0` is good reduction; `Ij`, where $j > 0$ is some positive number, means bad multiplicative reduction; `II, III, IV` or `Ij*`, for $j \geq 0$, or `II*,III*,IV*` mean additive reduction. For an explanation of the terminology of Kodaira symbols, see [**Sil86**], Appendix C, §15. For our example $E : y^2 = x^3 + 5x^2 + 35$, we obtain

`E.kodaira_type(2)` returns `II` (i.e., additive);

`E.kodaira_type(5)` returns `II` (i.e., additive);

`E.kodaira_type(7)` returns `I1` (i.e., multiplicative);

`E.kodaira_type(17)` returns `I2` (i.e., multiplicative);

`E.kodaira_type(11)` returns `I0` (i.e., good).

Note: if the equation is not minimal, some of the prime divisors of the discriminant may not be bad after all. For example,

> `E=EllipticCurve([0,0,0,0,15625]);`
>
> `prime_divisors(E.discriminant())` returns `[2,3,5]` but
>
> `E.kodaira_type(5)` returns `I0` (i.e., good).

This happened because the model $y^2 = x^3 + 15625$ is not minimal ($15625 = 5^6$); we should have used $y^2 = x^3 + 1$ instead.

If $E/\mathbb{Q}$ has good reduction at p, then E defines an elliptic curve over the finite field $\mathbb{Z}/p\mathbb{Z}$ and we can count the number of points modulo p (always including the extra point at infinity). N_p denotes this number of points while $a_p = p + 1 - N_p$. In GP, the command `ellap(E,p)` returns the coefficient a_p and `ellan(E,n)` returns an array with the first n coefficients a_k for $k = 1, \ldots, n$.

In Sage, the command `E.ap(p)` returns a_p while `E.an(n)` returns the nth coefficient (and only the nth), and `E.anlist(n)` provides a list of all the coefficients up to a_n. In Sage you can also directly find the number N_p by typing `E.Np(p)`.

The conductor of $E/\mathbb{Q}$ is another associated quantity that is very useful in practice:

- In Sage, type `E.conductor()`,
- In GP, type `ellglobalred(E)`.

The command `ellglobalred(E)` returns an array [conductor, global minimal model, product of local Tamagawa numbers]. In Sage, you can find a minimal model of an elliptic curve E by typing the command `E.minimal_model()`.

A.1.5. The torsion subgroup. It follows from the Mordell-Weil theorem that the torsion subgroup of an elliptic curve (over a number field) is a finite abelian group. Over $\mathbb{Q}$, a theorem of B. Mazur says that the torsion subgroup is one of the following: $\mathbb{Z}/n\mathbb{Z}$ with $1 \leq n \leq 10$ or $n = 12$, or $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2m\mathbb{Z}$ with $1 \leq m \leq 4$. One can compute the torsion subgroup as follows. The computation is easy, due to a theorem of Nagell and Lutz:

- In GP, the output of `elltors(E)` is a vector `[t, [n, m], [P,Q]]`, where t is the size of the torsion subgroup, which is isomorphic to $\mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$, generated by the points P and Q. If P is a torsion point, the command `ellorder(P)` provides the order of the element.
- In Sage, `E.torsion_order()` returns the order of the group, while `G = E.torsion_subgroup()` returns the group itself. Then `G.0` and `G.1` return generators for G.

Remark A.1.2. Even though the Nagell-Lutz theorem provides a simple algorithm to calculate the torsion subgroup of an elliptic curve, this method may not be very effective (at least when the discriminant is divisible by many primes). In general, there are better algorithms (for example, see [Dou98]).

A.1.6. The free part and the rank. It also follows from the Mordell-Weil theorem that the free part (here *free* is the opposite of torsion) of the group of points $E(K)$ on an elliptic curve (again over a number field K) is generated by a finite number of points P_1, P_2, ..., P_R of infinite order. The number R of generators (of infinite order) is called the rank of $E(K)$. There is no known algorithm that will always terminate and provide the rank and a set of generators. However, the so-called "descent algorithm" will terminate in certain cases (the descent procedure is an algorithm if Ш is finite, and we conjecture that Ш is always finite). The following commands compute lower and upper bounds for the rank and, in some cases, if they coincide, provide the rank of the curve. There are also commands to calculate generators; however, in many situations, the resulting points will only generate a group of finite index in $E(K)$ (the software will warn you when this may be the case). Some of the algorithms take an optional argument of a bound B.

In Sage, the command `E.selmer_rank_bound()` gives an upper bound of the rank, and `E.rank()`, `E.gens()` *try* to find, respectively, the rank and generators modulo torsion... but the computer may not succeed! When these commands are called, Sage is using an algorithm of Cremona in the background (see [Cre97]).

A.1.7. Heights and independence. In order to determine if a set of rational points is algebraically independent, we use a pairing arising from the canonical height. The following commands calculate the global Néron-Tate canonical height of a rational point P on a curve E:

In GP use `ellheight(E,P)`;

In Sage simply use `P.height()`, where `P` is a point on E.

If $S = \{P_1, \ldots, P_n\}$ is a set of rational points, we can test whether they are independent using the canonical height matrix. The height pairing of P and Q is defined by $\langle P, Q\rangle = h(P+Q) - h(P) - h(Q)$, where h is the canonical height on E. The height matrix relative to S is a matrix H whose coordinate ij is given by $\langle P_i, P_j\rangle$. The canonical height is a positive definite quadratic form on $E(\mathbb{Q})$ tensored with the reals. Thus, the determinant of H is non-zero if and only if the points in S are independent modulo torsion.

In GP use `S = [P1,P2,P3];`

`H=ellheightmatrix(E,S); matdet(H);`

In Sage use `E.height_pairing_matrix([P1,P2,P3])`,

where `P1, P2, P3` are points on E (previously defined). In GP, if `matdet(H)` returns 0, one can calculate generators for the kernel of H with `matker(H)`. Each element of the kernel represents a linear combination of points that adds up to a torsion point. In Sage, you may use `H.kernel()` for the same purpose.

A.1.8. Elliptic curves over $\mathbb{C}$. The period lattice of an elliptic curve $E/\mathbb{Q}$ can be found by typing

```
L=E.period_lattice()
```

and a basis for the period lattice is found simply using `L.basis()`. Using PARI/GP, one can start from a lattice and obtain the associated elliptic curve, as follows:

`L=[1,I]`;
`elleisnum(L,4)` returns $G_4(L)$,
  which equals `2268.8726415...`,
`elleisnum(L,6)` returns $G_6(L)$,
  which equals `-3.97...E-33`, i.e., 0,
thus, L corresponds to an elliptic curve
$y^2 = x^3 - (34033.089\ldots)x$.

The elliptic curve $y^2 = x^3 - (34033.089\ldots)x$ is isomorphic to $E/\mathbb{Q}$: $y^2 = x^3 - x$ over $\mathbb{C}$. Thus, $\mathbb{C}/\langle 1, i\rangle \cong E(\mathbb{C})$.

A.2. Modular forms

In this section, all commands we list are to be used in the Sage environment.

A.2.1. The modular group and congruence subgroups. The modular group and main congruence subgroups, defined for any $N > 0$ by

$$\begin{aligned}
\mathrm{SL}(2,\mathbb{Z}) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a,b,c,d \in \mathbb{Z},\ ad-bc=1 \right\}, \\
\Gamma_0(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : c \equiv 0 \bmod N \right\}, \\
\Gamma_1(N) &= \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) : a \equiv d \equiv 1 \bmod N \right\},
\end{aligned}$$

may be defined in Sage using `SL2Z`, `Gamma0(N)`, and `Gamma1(N)`, respectively. Alternatively, $\mathrm{SL}(2,\mathbb{Z})$ can also be defined as $\Gamma_0(1)$. Notice that those 2×2 matrices that define elements of congruence subgroups are stored in Sage as 4-dimensional row vectors. One can use the subcommand `.gens()` on any of the modular and congruence groups to find a set of matrices that generate (multiplicatively) the given group.

You can call the generators by using the suffix `[0]`, `[1]`, etc. Here are some examples:

`A = SL2Z([1,1,0,1]);`

`G = SL2Z.gens()` returns two matrices

$$G[0] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad G[1] = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

`H = Gamma0(3).gens()` returns six matrices

$$H[0] = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, H[1] = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, H[2] = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix},$$

$$H[3] = \begin{pmatrix} 1 & -1 \\ 3 & -2 \end{pmatrix}, H[4] = \begin{pmatrix} 2 & -1 \\ 3 & -1 \end{pmatrix}, H[5] = \begin{pmatrix} -2 & 1 \\ -3 & 1 \end{pmatrix}.$$

The genus of the modular curve $X_0(N)$ can be computed with the command `Gamma0(N).genus()`. Similarly, `Gamma(N).genus()` and `Gamma1(N).genus()` return the genus of $X(N)$ and $X_1(N)$, respectively.

A.2.2. Vector spaces of modular forms. Let Γ be a congruence subgroup of $\mathrm{SL}(2,\mathbb{Z})$ and define:

- $M_k(\Gamma)$, the $\mathbb{C}$-vector space of all modular forms for Γ of weight k;
- $S_k(\Gamma)$, the $\mathbb{C}$-vector space of all cusp forms for Γ of weight k.

Suppose you have already defined a congruence subgroup `G` (for example, `G = Gamma0(3)`) and are interested in forms of weight `k`. The vector spaces of modular forms and cusp forms can be defined in Sage by

`M=ModularForms(G,k)` or `ModularForms(G,k,prec=m)`

if you want q-series expansions up to q^m;

`S=CuspForms(G,k)` or `CuspForms(G,k,prec=m)`.

The precision is set to 6 by default. If you want to find the dimension or a basis, you can use the suffix `.dimension()` or `.basis()`,

respectively. Here is an example:

`M=ModularForms(Gamma0(3),4, prec=10);`

`M.dimension()` returns `2`;

`M.basis()` returns the forms:

$[1 + 240q^3 + 2160q^6 + 6720q^9 + O(q^{10}),$

$q + 9q^2 + 27q^3 + 73q^4 + 126q^5 + 243q^6$

$+344q^7 + 585q^8 + 729q^9 + O(q^{10})].$

The command `CuspForms(Gamma0(3),4,prec=10)` returns only the 0 vector space. Notice that even though the modular form $q + 9q^2 + 27q^3 + O(q^4)$ vanishes at the cusp at infinity (because $a_0 = 0$ in the expansion), it is not a cusp form for $\Gamma_0(3)$ because it does not vanish at *all* the cusps of $X_0(3)$ (infinity is not the only cusp!). The command `AllCusps(N)` produces a list of all (representatives of) cusps of $X_0(N)$.

`AllCusps(3)` returns `[(inf),(0)]`.

A.3. L-functions

Let $E/\mathbb{Q}$ be an elliptic curve, and let $L(E, s)$ be the Hasse-Weil L-function associated to E, as in Definition 5.1.1. This L-function is defined in Sage using the command

```
L=E.lseries()
```

or one can use `L=E.lseries().dokchitser()` to use Dokchitser's algorithms to calculate values ([Dok04]). Once we have defined $L = L(E, s)$, we can evaluate L. For example:

`E=EllipticCurve([1,2,3,4,5]);`

`L=E.lseries();`

`L(1)` which returns 0,

`L(1+I) = -0.485502124065793 + 0.627256178203893*I`.

The value $L(E, 1) = 0$ is predicted in this case by the Birch and Swinnerton-Dyer conjecture (Conjecture 5.2.1), since the rank of E is > 0 (in fact, the rank is 1). One can also plot $L(E, x)$ when x takes

real values (because $L(E,x)$ is real valued for $x \in \mathbb{R}$). For instance, the graph in Figure 2 was created with the following lines of code:

```
E0=EllipticCurve([0,-1,1,-10,-20]);
L0=E0.lseries().dokchitser();
P0=plot(lambda x:  L0(x).real(),0, 3);
show(P0,xmin=-0.5, ymin=0, dpi=150).
```

If you want to create a PDF file with your graph, you can use

```
P=plot(lambda x:  real(L0(x)),0, 3).save(
"bsdrank0.pdf",xmin=-0.5, ymin=-0.2, dpi=150).
```

You may also want to calculate the Taylor polynomial of $L(E,s)$ around the point $x=a$ of degree $n-1$ with `L.taylor_series(a,n)`.

A.3.1. Data related to the BSD conjecture. The Shafarevich-Tate group of $E/\mathbb{Q}$ is defined in Sage by `E.sha()` but, in general, it is difficult to calculate its order. The user can calculate a conjectural value of Sha by typing `E.sha().an()`. The conductor N of $E/\mathbb{Q}$ is calculated with `E.conductor()`. The Tamagawa product $\prod_{p|N} c_p$ can be calculated directly with `E.tamagawa_product()` or the individual Tamagawa numbers c_p, for each prime $p|N$, may be calculated with `E.tamagawa_number(p)`. The regulator of $E/\mathbb{Q}$ can be calculated by `E.regulator()`. Finally, the real period Ω_E is calculated as follows:

```
E=EllipticCurve([1,2,3,4,5]);
M=E.period_lattice();
```

Then `M.omega` returns $\Omega_E = 2.78074001376673\ldots$.

The reader should try to use the commands above to calculate all the invariants listed in Examples 5.2.3 and 5.2.4 (see Figure 6 and Figure 7).

A.4. Other Sage commands

- Continued fractions:

`continued_frac_list(N)` returns the continued fraction of N;
`continued_frac_list(N,partial_convergents=True)` or
`convergents(v)` return convergents for the cont. frac. v.

- The Kronecker symbol (defined in Example 1.3.3):

`kronecker(-n,m)` returns the Kronecker symbol $\left(\frac{-n}{m}\right)$.

Appendix B

Complex analysis

In this appendix we review some of the basic notions of complex numbers and the theory of analytic and meromorphic functions on the complex plane. This brief appendix is by no means a replacement for a good course or a good book on complex analysis such as [**Ahl79**].

B.1. Complex numbers

The complex numbers, usually denoted by $\mathbb{C}$, are defined as an extension of the real numbers $\mathbb{R}$. Over the reals, the equation $x^2 + 1 = 0$ has no solutions, so we define a new number i that satisfies $i^2 = -1$. Therefore $x^2 + 1 = 0$ now has two solutions, namely i and $-i$. We define $\mathbb{C}$ by adjoining our new number i to $\mathbb{R}$:

$$\mathbb{C} = \{a + bi : a, b \in \mathbb{R},\ i^2 = -1\}.$$

The real and imaginary parts of a complex number $\alpha = a + bi$ are denoted, respectively, by $\Re(\alpha) = a$ and $\Im(\alpha) = b$. If $\Im(\alpha) = b = 0$ we say that α is a *real number*, and if $\Re(\alpha) = a = 0$ we say that α is *purely imaginary*. We can add and multiply two complex numbers $\alpha = a+bi$ and $\beta = c+di$ to obtain a new complex number, as follows:

$$\begin{aligned} \alpha + \beta &= (a + bi) + (c + di) = (a + c) + (b + d)i; \text{ and} \\ \alpha \cdot \beta &= (a + bi) \cdot (c + di) = (ac - bd) + (ad + bc)i. \end{aligned}$$

The set of all complex numbers together with the operations of addition and multiplication form a field (see Exercise B.7.1).

There are two other operations on complex numbers that occur often: complex conjugation and calculating the modulus, or absolute value. The *complex conjugate* of $\alpha = a+bi$ is $\overline{\alpha} = a-bi$. The *modulus* or *absolute value* of α is

$$|\alpha| = \sqrt{\alpha \cdot \overline{\alpha}} = \sqrt{(a+bi)(a-bi)} = \sqrt{a^2+b^2}.$$

Notice that, for any $\alpha, \beta \in \mathbb{C}$, we have

$$\overline{\alpha+\beta} = \overline{\alpha} + \overline{\beta},\ \overline{\alpha \cdot \beta} = \overline{\alpha} \cdot \overline{\beta},\ |\alpha\beta| = |\alpha||\beta|,\ \text{and}\ |\alpha+\beta| \leq |\alpha| + |\beta|.$$

We constructed the complex numbers by adjoining i to $\mathbb{R}$ so that $i \in \mathbb{C}$ and therefore the equation $x^2+1=0$ has two solutions in $\mathbb{C}$. But something extremely surprising happened in this construction. It turns out that not only x^2+1 has a root in $\mathbb{C}$ but, in fact, *every polynomial* with complex coefficients has a root in $\mathbb{C}$. This is an extremely important result:

Theorem B.1.1 (Fundamental Theorem of Algebra)**.** *Let $p(z)$ be a polynomial*

$$p(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$$

with complex coefficients $a_i \in \mathbb{C}$ and degree ≥ 1. Then there exists a complex number $\alpha \in \mathbb{C}$ such that $p(\alpha) = 0$.

The proof is left to the reader (Exercise B.7.6).

B.2. Analytic functions

Definition B.2.1. Let $\alpha \in \mathbb{C}$ and $\delta \in \mathbb{R}^+$. An *open disc* $D_\delta(\alpha)$ in the complex plane, centered at α and of radius $\delta > 0$, is the set

$$D_\delta(\alpha) = \{z \in \mathbb{C} : |z-\alpha| < \delta\}.$$

Definition B.2.2. We say that a set $S \subseteq \mathbb{C}$ is *open* if for every $\alpha \in S$ there is a real number $\delta > 0$ such that $D_\delta(\alpha) \subseteq S$. We say that a set $T \subseteq \mathbb{C}$ is *closed* if the complement of T in $\mathbb{C}$, i.e., $\mathbb{C} - T$, is open.

Definition B.2.3. A non-empty connected open set in the complex plane is called a *region*.

Let U be a region in the complex plane and let $f(z) : U \to \mathbb{C}$ be a complex-valued function on U. Let $\alpha \in U$. We say that f *has a derivative at* α if the usual limit converges:

$$f'(\alpha) = \lim_{h \to 0} \frac{f(\alpha + h) - f(\alpha)}{h},$$

where h runs over complex numbers inside U that approach 0. Alternatively (or more precisely), we can define $f'(z)$ using ϵ and δ as follows. We say that f has a derivative at α with value $m = f'(\alpha)$ if the following statement holds: for every real $\epsilon > 0$ there exists a real $\delta > 0$ such that, if $h \in D_\delta(\alpha)$, then

$$\left| \frac{f(\alpha + h) - f(\alpha)}{h} - m \right| < \epsilon.$$

Definition B.2.4. Let $U \subseteq \mathbb{C}$ be a region and let $f(z)$ be a complex-valued function $f : U \to \mathbb{C}$ defined for every $z \in U$. We say that $f(z)$ is *analytic* (or *holomorphic*, or *entire*) on U if it has a derivative at each $z \in U$.

Example B.2.5. The function $f(z) = z$ is analytic on the whole complex plane $\mathbb{C}$ (Exercise B.7.3). The function $g(z) = 1/z$ is analytic on $\mathbb{C} - \{0\}$.

It is not hard to show that the sum, product and composition of two analytic functions are also analytic. Thus, all polynomials in one variable with complex coefficients define analytic functions. Similarly, the quotient of two analytic functions is analytic except at the zeros of the denominator. Thus, all rational functions (quotients of polynomials) are analytic in the complex plane except at the zeros of the polynomial in the denominator.

Remark B.2.6. Let U be a region of $\mathbb{C}$ and let $f : U \to \mathbb{C}$ be an analytic function. We write $f(z)$ where $z = x + yi \in U$, with $x, y \in \mathbb{R}$. We may also write

$$f(z) = u(z) + v(z)i,$$

where $u, v : U \to \mathbb{R}$ are real-valued functions. Since f is analytic on U, the functions f, u and v are continuous on U (Exercise B.7.4). Since f is analytic, the limit

$$f'(z) = \lim_{h \to 0} \frac{f(z + h) - f(z)}{h} \tag{B.1}$$

exists for every $z \in U$. The parameter h runs over complex numbers in U approaching zero, but we may restrict h to real values (thus, we are calculating $\partial f/\partial x$). The value of the limit in Eq. (B.1) does not change under this restriction, and this means that the partial derivative of f with respect to x equals $f'(z)$. Hence

$$f'(z) = \frac{\partial f}{\partial x} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} i.$$

Similarly, we may restrict h to purely imaginary values $h = ik$, and then

$$\begin{aligned} f'(z) &= \lim_{h\to 0} \frac{f(z+h)-f(z)}{h} = \lim_{k\to 0} \frac{f(z+ik)-f(z)}{ik} \\ &= \frac{1}{i} \cdot \lim_{k\to 0} \frac{f(z+ik)-f(z)}{k} = (-i)\frac{\partial f}{\partial y}. \end{aligned}$$

It follows that $f'(z) = (-i)\frac{\partial f}{\partial y} = (-i)\frac{\partial u}{\partial y} + \frac{\partial v}{\partial y}$. Therefore,

$$f'(z) = \frac{\partial f}{\partial x} = (-i)\frac{\partial f}{\partial y} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x} i = \frac{\partial v}{\partial y} - \frac{\partial u}{\partial y} i.$$

The last equality implies that the real and imaginary parts of every analytic function must satisfy the following differential equations:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y} \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{B.2}$$

These are called the *Cauchy-Riemann differential equations.*

Differentiability (or being analytic) over $\mathbb{C}$, as in Definition B.2.4, is a much stronger condition than differentiability over $\mathbb{R}$. Indeed, the existence of a complex derivative implies that the function is in fact *infinitely differentiable* and *locally equal to its own Taylor series.* We explain what these terms mean in the following theorem.

Theorem B.2.7. *Let $U \subseteq \mathbb{C}$ be a region and let $f : U \to \mathbb{C}$ be analytic. Then f has derivatives of all orders on U (i.e., the derivatives $f'(z), f''(z), \ldots$ and, more generally, $f^{(n)}(z)$ for all $n \geq 1$ are continuous and differentiable complex-valued functions on U).*

Moreover, for every $\alpha \in U$, the Taylor series of $f(z)$ about $z = \alpha$ converges to $f(z)$ in some neighborhood of α. In other words, for

every $\alpha \in U$, there is a real $\delta > 0$ such that the Taylor series

$$T(z;\alpha) = \sum_{n=0}^{\infty} \frac{f^{(n)}(\alpha)}{n!}(z-\alpha)^n$$

converges for all $z \in D_\delta(\alpha)$, and $T(z;\alpha) = f(z)$.

Conversely, if $S(z) = \sum_{n=0}^{\infty} a_n(z-\alpha)^n$ is a power series with complex coefficients a_n with a radius of convergence R (i.e., $S(z)$ converges for all $z \in \mathbb{C}$ with $|z-\alpha| < R$), then $S(z)$ defines an analytic function on the open disc $D_R(\alpha)$.

Example B.2.8. Let $f(z) = \sum_{n=0}^{\infty} \frac{z^n}{n!}$. The radius of convergence of this series is infinite (over $\mathbb{C}$ as well as over $\mathbb{R}$), so it defines an analytic function in the complex plane. The function $f(z)$ is, of course, the complex exponential function which we discuss below in B.4 in some more detail. Similarly, we define $\sin(z)$ and $\cos(z)$ using the usual Taylor expansions

$$\sin(z) = \sum_{n=0}^{\infty}(-1)^{2n+1}\frac{z^{2n+1}}{(2n+1)!}, \quad \cos(z) = \sum_{n=0}^{\infty}(-1)^{2n}\frac{z^{2n}}{(2n)!}.$$

Since the radius of convergence of these series is infinite, $\sin(z)$ and $\cos(z)$ define analytic functions on $\mathbb{C}$.

B.3. Meromorphic functions

At this juncture, it is useful to extend the complex numbers by introducing a point at infinity ∞. We will write $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ for the *extended complex plane*. We set the convention that every straight line shall pass through the point at infinity. (Note that $\widehat{\mathbb{C}}$ is simply the projective line over $\mathbb{C}$, i.e., $\mathbb{P}^1(\mathbb{C})$. See Appendix C for an introduction to the projective line and projective geometry.)

With this definition of ∞, suppose that $f(z)$ is a complex-valued function not defined at α. The expression

$$\lim_{z\to\alpha} f(z) = \infty$$

means that $|f(z)|$ is unbounded as z approaches α. For instance, $f(z) = 1/z$ is not defined at 0 and $\lim_{z\to 0} 1/z = \infty$. This "$\infty$" is the complex point at infinity, and it should not be confused with the

infinity that we use in real analysis ("very, very far along the positive x-axis"). In fact, in $\mathbb{R}$, the limit $\lim_{x\to 0} 1/x$ is undefined (as the value may be $\pm\infty$ depending on how we approach 0), but in $\widehat{\mathbb{C}}$, the limit $\lim_{z\to 0} 1/z = \infty$ simply means that if z is close to 0, then $1/z$ is far from 0 (in some direction, not necessarily along the x-axis).

Suppose that $f(z)$ is some complex-valued function that is not defined at α but is analytic in a neighborhood of α. How can f fail to be analytic at α? The function $f(z)$ may have a *removable singularity* (e.g., $\sin(z)/z$), an *essential singularity* (e.g., $\sin(1/z)$) or a *pole* (e.g., $1/z$). Here we will only discuss poles in some detail (for a complete discussion, see [**Ahl79**], Ch. 4, §3).

Definition B.3.1. Let f be a complex-valued function, and let $\alpha \in \mathbb{C}$. We say that f has a *pole* (or *isolated pole*) at $z = \alpha$ if:

(1) The function $f(z)$ is analytic on some disc $D_\delta(\alpha)$ centered at α, except at α itself. In other words, f is analytic on the punctured disc

$$\{z \in \mathbb{C} : 0 < |z - \alpha| < \delta\}$$

for some $\delta > 0$; and

(2) The limit of f at α is infinite:

$$\lim_{z\to\alpha} f(z) = \infty.$$

Definition B.3.2. A function $f(z)$ is *meromorphic* in a region U if f is analytic on U except for a set of isolated poles.

Remark B.3.3. Suppose that $f(z)$ is meromorphic in a region U with an isolated pole at $\alpha \in U$. It does not make sense to write $f : U \to \mathbb{C}$, since $\lim_{z\to\alpha} f(\alpha) = \infty$. Instead, we may write $f : U \to \widehat{\mathbb{C}}$.

Example B.3.4. Let $p(z)$ and $q(z)$ be polynomials in $\mathbb{C}[z]$ such that p and q have no common factors. Then the rational function $p(z)/q(z)$ is a meromorphic function with isolated poles at the zeros of $q(z)$.

Example B.3.5. The function $\sin(1/z)$ has infinitely many zeros accumulating near $z = 0$ (there is a zero at each $z = 1/(\pi k)$ for each $k \geq 1$). Therefore, $g(z) = (\sin(1/z))^{-1}$ is not meromorphic because the singularity at 0 is not isolated. In fact, the function g

has infinitely many poles in any open neighborhood of 0. Notice, however, that $(\sin(z))^{-1}$ is a meromorphic function.

Remark B.3.6. Let $f(z)$ be a function that is analytic in a disc $D_R(\alpha)$ except, perhaps, at $\alpha \in \mathbb{C}$. Then $f(z)$ has a Laurent expansion of the form

$$f(z) = \sum_{n=-\infty}^{\infty} c_n(z-\alpha)^n.$$

Then, the function $f(z)$:

(1) is analytic at α if $c_n = 0$ for all $n < 0$ and $f(\alpha) = c_0$ (if $f(\alpha) \neq c_0$, or if $f(\alpha)$ is undefined, then there is a removable singularity at α),

(2) is meromorphic at α if there is some $M > 0$ such that $c_n = 0$ for all $n < -M$; i.e., the expansion of $f(z)$ is of the form

$$f(z) = \sum_{n=-M}^{\infty} c_n(z-\alpha)^n,$$

and

(3) has an essential singularity at α if there are infinitely many $n < 0$ such that $c_n \neq 0$.

B.4. The complex exponential function

The usual real exponential function e^x can be extended to the field of complex numbers as follows. Let $z = x + yi$ with $x, y \in \mathbb{R}$. Then we define e^z by

$$e^z = e^{x+yi} := e^x(\cos(y) + \sin(y)i).$$

Equivalently, e^z can be defined as a Taylor series (which coincides with the Taylor series of the real valued exponential function):

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}.$$

If $z = x + yi$ with $x, y \in \mathbb{R}$:

$$\begin{aligned} |e^z| &= |e^{x+yi}| = |e^x \cos(y) + (e^x \sin(y))i| \\ &= \sqrt{(e^x \cos(y))^2 + (e^x \sin(y))^2} \\ &= \sqrt{e^{2x}(\cos^2(y) + \sin^2(y))} = e^x. \end{aligned}$$

Notice that, if $\theta \in \mathbb{R}$, then $e^{\theta i}$ is a complex number that lies on the unit complex circle $\{z \in \mathbb{C} : |z| = 1\}$. Indeed, by the formula above, $|e^{\theta i}| = |e^{0+\theta i}| = e^0 = 1$.

In the theory of L-functions, we often calculate powers of natural numbers $n \in \mathbb{N}$ with complex exponents $s \in \mathbb{C}$. Next, we define what n^s means precisely. If $n \in \mathbb{N}$ and $s = x + yi \in \mathbb{C}$, we define $n^s = e^{\log(n)s}$, i.e.,

$$\begin{aligned} n^s &= e^{\log(n)s} = e^{\log(n)x + \log(n)yi} \\ &= e^{\log(n)x}(\cos(\log(n)y) + \sin(\log(n)y)i) \\ &= n^x(\cos(\log(n)y) + \sin(\log(n)y)i). \end{aligned}$$

B.5. Theorems in complex analysis

In this section we state some of the most important and useful theorems about analytic functions. We have already stated two fundamental theorems, namely Theorems B.1.1 and B.2.7.

The first two theorems concern line integrals along closed curves. If γ is a closed curve (the starting point is equal to the end point) in $\mathbb{C}$, and $f(z)$ is a function defined at every point of γ, then the symbol $\int_\gamma f(z)dz$ represents the line integral of $f(z)$ along γ. A curve is contractible in a region U if it can be continuously shrunk to a point, always staying inside U. The winding number of a curve γ with respect to a point $\alpha \in \mathbb{C}$, denoted by $n(\gamma, \alpha)$, counts the number of times that the path γ winds around α. The winding number is positive if the curve goes around α in the counterclockwise direction, and negative otherwise. (See [**Ahl79**], Ch. 4.)

Theorem B.5.1 (Cauchy's Theorem)**.** *Let U be a region in $\mathbb{C}$, let $f(z)$ be a complex-valued function that is analytic on U, and let γ be*

any contractible closed curve contained in U. Then

$$\int_\gamma f(z)dz = 0.$$

Theorem B.5.2 (Cauchy's Integral Formula)**.** *Let $f(z)$ be a function that is analytic in a region U, and let γ be a closed curve inside U. For any point α not on γ, we have*

$$n(\gamma, \alpha) \cdot f(\alpha) = \frac{1}{2\pi i} \int_\gamma \frac{f(z)}{z - \alpha} dz,$$

where $n(\gamma, \alpha)$ is the winding number of γ around α.

Cauchy's Theorem B.5.1 has the following converse.

Theorem B.5.3 (Morera's Theorem)**.** *If $f(z)$ is defined and continuous in a region $U \subseteq \mathbb{C}$, and if $\int_\gamma f(z)dz = 0$ for all closed curves γ in U, then $f(z)$ is analytic in U.*

Another important theorem about line integrals is the Residue Theorem (see [**Ahl79**], Ch. 4, §5.1). Before stating the next theorem, we remind the reader that a region is by definition a non-empty connected open set.

Theorem B.5.4 (The Maximum Principle)**.** *If $f(z)$ is analytic and non-constant in a region U, then its absolute value $|f(z)|$ has no maximum in U. Alternatively, if $f(z)$ is an analytic function on a closed bounded set T, then the maximum of $|f(z)|$ occurs on the boundary of T.*

Theorem B.5.5 (Liouville's Theorem)**.** *A function which is analytic and bounded in the whole complex plane must be constant.*

We say that α in a set $S \subseteq \mathbb{C}$ is an *accumulation point* in S if for every $\delta > 0$ there is point $\beta \in S$, $\beta \neq \alpha$ such that $|\beta - \alpha| < \delta$.

Theorem B.5.6. *If $f(z)$ and $g(z)$ are analytic in a region U, and if $f(z) = g(z)$ for every z in a set S which has an accumulation point in U, then $f(z)$ is identically equal to $g(z)$ on all points of U.*

The previous theorem has some remarkable consequences: if $f(z)$ is analytic in U and it is identically zero in a set $S \subseteq \mathbb{C}$ that contains

an accumulation point, then $f(z)$ is identically zero. Also, we deduce that an analytic function is uniquely determined by its values on any set with an accumulation point in the region of analyticity.

Theorem B.5.7 (Conformal Mapping Theorem)**.** *A complex function is analytic if and only if it maps pairs of intersecting curves into pairs that intersect at the same angle.*

B.6. Quotients of the complex plane

In the theory of elliptic curves over $\mathbb{C}$, we often work with a quotient of the complex plane $\mathbb{C}$ modulo some lattice L. See Section 3.1 for the definition of lattice, the definition of the quotient $\mathbb{C}/L$ and the relationship to elliptic curves. In this section, we define what it means for a map $\mathbb{C}/L \to \mathbb{C}$ to be analytic.

Let $L \subset \mathbb{C}$ be a lattice with a basis $L = \langle w_1, w_2 \rangle$. Usually, we fix a fundamental domain for L as follows

$$\mathcal{F}_L = \{\lambda w_1 + \mu w_2 \in \mathbb{C} : 0 \leq \lambda, \mu < 1\}.$$

For our purposes here, we will define a fundamental domain for $\mathbb{C}/L$ for each $\alpha \in \mathbb{C}$ such that α is positioned in the interior of the domain:

$$\mathcal{F}_{L,\alpha} = \{\alpha + \lambda w_1 + \mu w_2 \in \mathbb{C} : -1/2 \leq \lambda, \mu < 1/2\}$$

and we also define the interior of $\mathcal{F}_{L,\alpha}$ by

$$\mathcal{F}^0_{L,\alpha} = \{\alpha + \lambda w_1 + \mu w_2 \in \mathbb{C} : -1/2 < \lambda, \mu < 1/2\}.$$

Notice that $\mathcal{F}^0_{L,\alpha}$ is a region in $\mathbb{C}$ (it is non-empty, connected and open), and α is at the center of the region. Notice that there is a bijection

$$\psi_{L,\alpha} : \mathbb{C}/L \to \mathcal{F}_{L,\alpha}. \tag{B.3}$$

Let $f : \mathbb{C}/L \to \mathbb{C}$ be a complex-valued function that is well-defined for every element of the quotient $\mathbb{C}/L$. Let $\alpha \bmod L$ be such an element. We say that $f : \mathbb{C}/L \to \mathbb{C}$ *is analytic at* α if the map

$$\widehat{f} : \mathcal{F}^0_{L,\alpha} \to \mathbb{C}, \quad \widehat{f}(z) = f(z \bmod L)$$

is analytic at α.

When we discuss maps between elliptic curves (e.g., Proposition 3.1.6), we talk about analytic maps $f : \mathbb{C}/L \to \mathbb{C}/L'$, where L and

L' are lattices. What does "analytic" mean in this context? How do we define analyticity? It is simply a matter of choosing correct charts for each $\mathbb{C}/L$ and $\mathbb{C}/L'$, as we shall see next.

Let $f : \mathbb{C}/L \to \mathbb{C}/L'$ be a continuous map. Let $\alpha \in \mathbb{C}$ and suppose that $f(\alpha \bmod L) = \beta \bmod L'$. Let $\mathcal{F}^0_{L,\alpha}$ be the region about α defined above, and similarly define $\mathcal{F}^0_{L',\beta}$. Let $\epsilon > 0$ be small enough so that the disc $D_\epsilon(\beta)$ is completely contained in $\mathcal{F}^0_{L',\beta}$. Then, by continuity of f, there is a δ such that if $|z - \alpha| < \delta$, then $f(z \bmod L)$ is inside $D_\epsilon(\beta)$. Pick δ small enough so that $D_\delta(\alpha)$ is completely contained in $\mathcal{F}^0_{L,\alpha}$. We are now ready to state our definition: we say that the continuous map $f : \mathbb{C}/L \to \mathbb{C}/L'$ *is analytic at* $\alpha \bmod L$ if the map

$$\widehat{f} : D_\delta(\alpha) \to D_\epsilon(\beta), \quad \widehat{f}(z) = \psi_{L',\beta}(f(z \bmod L)) \in D_\epsilon(\beta) \subseteq \mathcal{F}_{L',\beta}$$

is analytic at α, where $\psi_{L',\beta} : \mathbb{C}/L' \to \mathcal{F}_{L',\beta}$ is the bijection we defined in Eq. (B.3).

B.7. Exercises

Exercise B.7.1. The goal of this exercise is to prove that $\mathbb{C}$ is a field.

(1) Show that any non-zero complex number $\alpha = a + bi$ has a multiplicative inverse which is also a complex number $\alpha^{-1} = c + di$ with $c, d \in \mathbb{R}$.

(2) Convince yourself that $\mathbb{C}$ is a field; i.e., justify why $\mathbb{C}$ satisfies each of the field axioms.

Exercise B.7.2. Let α be a complex number. Show that $\alpha \in \mathbb{R}$ if and only if $\alpha = \overline{\alpha}$.

Exercise B.7.3. Show that $f(z) = z$ is analytic on $\mathbb{C}$. Also, show that $g_n(z) = z^n$ is analytic on $\mathbb{C}$ for every $n \geq 1$ and that the derivative is $g_n'(z) = nz^{n-1}$.

Exercise B.7.4. Let $f(z)$ be a complex-valued function that is analytic in a region $U \subseteq \mathbb{C}$.

(1) Show that f is also continuous at every point of U (i.e., $\lim_{h \to a} f(h) = f(a)$ for every $a \in U$).

(2) Let $f(z) = u(z) + v(z)i$, where $u(z)$ and $v(z)$ are real-valued. Show that u and v are continuous on U.

Exercise B.7.5. Let $f(z)$ be an analytic function on a region U, and write $\Re(f(z)) = u(z)$, $\Im(f(z)) = v(z)$ for the real and imaginary parts of $f(z)$, respectively. We define the *Laplacians* of u and v by

$$\Delta u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad \text{and} \quad \Delta v = \frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2}.$$

Show that $\Delta u = \Delta v = 0$. (Hint: use the Cauchy-Riemann differential equations, i.e., Eq. B.2.)

Exercise B.7.6. Prove the Fundamental Theorem of Algebra B.1.1: if $P(z)$ is a non-constant polynomial, then there is a root of P in $\mathbb{C}$. (Hint: suppose that $P(z)$ has no roots in $\mathbb{C}$. Then $1/P(z)$ would be analytic. Now use Liouville's Theorem B.5.5.)

Appendix C

Projective space

C.1. The projective line

Let us begin with an example. Consider the function $f(x) = \frac{1}{x}$. We know from Calculus that f is continuous (and differentiable) on all of its domain (i.e., $\mathbb{R}$) except at $x = 0$. Would it be possible to extend the real line so that $f(x)$ is continuous everywhere? The answer is yes, it is possible, and the solution is to *glue* the "end" of the real line at ∞ with the other "end" at $-\infty$. We will describe the solution in detail below. Formally, we need the *projective line*, which is a line with points $\mathbb{R} \cup \{\infty\}$, i.e., a real line plus a single point at infinity that ties the line together (into a circle).

The formal definition of the projective line is as follows. It may seem a little confusing at first, but it is fairly easy to work and compute with it. First, we need to define a relation between vectors of real numbers in the plane. Let a, b, x, y be real numbers such that neither (x, y) nor (a, b) is the zero vector. We say that $(x, y) \sim (a, b)$ if the vector (x, y) is a non-zero multiple of the vector (a, b). In other words, if we consider (a, b) and (x, y) as points in the plane, we say that $(a, b) \sim (x, y)$ if they both lie in one line on the plane that passes through the origin. Again:

$(x, y) \sim (a, b)$ if and only if there is $\lambda \in \mathbb{R}$ such that $x = \lambda a$, $y = \lambda b$.

For instance, $(\sqrt{2}, \sqrt{2}) \sim (1, 1)$. We denote by $[x, y]$ the set of all vectors (a, b) such that $(x, y) \sim (a, b)$:

$$[x, y] = \{(a, b) : a, b \in \mathbb{R} \text{ such that } (a, b) \neq (0, 0) \text{ and } (x, y) \sim (a, b)\}.$$

Finally, we define the real projective line by

$$\mathbb{P}^1(\mathbb{R}) = \{[x, y] : x, y \in \mathbb{R} \text{ with } (x, y) \neq (0, 0)\}.$$

If you think about it, $\mathbb{P}^1(\mathbb{R})$ is the set of all lines through the origin (each class $[x, y]$ consists of all points — except the origin — on the line that goes through (x, y) and $(0, 0)$). The important thing to notice is that if $[x, y] \in \mathbb{P}^1(\mathbb{R})$ and $y \neq 0$, then $(x, y) \sim (\frac{x}{y}, 1)$, so the class of $[x, y]$ contains a unique representative of the form $(a, 1)$ for some $a = \frac{x}{y} \in \mathbb{R}$. This allows the following decomposition of $\mathbb{P}^1(\mathbb{R})$:

$$\mathbb{P}^1(\mathbb{R}) = \{[x, 1] : x \in \mathbb{R}\} \cup \{[1, 0]\}.$$

The set of points $\{[x, 1]\}$ are in bijection with $\mathbb{R}$ and, therefore, form a real line. The point $[1, 0]$, which is the only point in $\mathbb{P}^1(\mathbb{R})$ that does not belong to the real line $\{[x, 1]\}$, is called the *point at infinity* (see Figure 1).

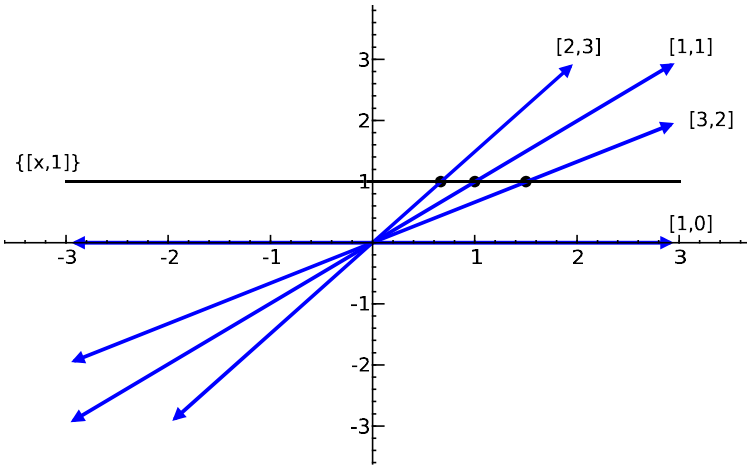


Figure 1. Some points in the projective line, e.g., $[2, 3] \in \mathbb{P}^1(\mathbb{R})$, and their representatives of the form $[x, 1]$, e.g. $[\frac{2}{3}, 1]$, except for $[1, 0]$.

Notice that when $x \in \mathbb{R}$ gets large (i.e., $x \to \infty$ or $x \to -\infty$), the point $[x, 1] \in \mathbb{P}^1(\mathbb{R})$ corresponds to a line in the real plane that is closer and closer to the horizontal line. Since the horizontal line corresponds to the point $[1, 0] \in \mathbb{P}^1(\mathbb{R})$, we see that as x gets large (in either the positive or negative direction!), the points $[x, 1]$ get closer and closer to $[1, 0]$, the point at infinity. This is what we meant at the beginning of this section by "glueing" both ends of the real line, ∞ and $-\infty$, at one point.

Let us see that, with this definition, the function $f : \mathbb{R} \to \mathbb{R}$, $f(x) = 1/x$ is continuous everywhere when extended to $\mathbb{P}^1(\mathbb{R})$. We define instead an extended function $F : \mathbb{P}^1(\mathbb{R}) \to \mathbb{P}^1(\mathbb{R})$ by

$$F([x, y]) = [y, x].$$

Notice that a point on the real line of $\mathbb{P}^1$, i.e., a point of the form $[x, 1]$, is sent to the point $[1, x]$ of $\mathbb{P}^1$, and $(1, x) \sim (\frac{1}{x}, 1)$ as long as $x \neq 0$. So $[x, 1]$ with $x \neq 0$ is sent to $[\frac{1}{x}, 1]$ via F (i.e., the real point x is sent to $\frac{1}{x}$). Hence, F coincides with f on $\mathbb{R} - \{0\}$. But F is perfectly well-defined on $x = 0$, i.e., on the point $[0, 1]$, and $F([0, 1]) = [1, 0]$ so that $[0, 1]$ is sent to the point at infinity. Moreover, both sided limits coincide:

$$\lim_{x \to 0^+} F([x, 1]) = \lim_{x \to 0^-} F([x, 1]) = F([0, 1]) = [1, 0].$$

C.2. The projective plane

We may generalize the construction above of the projective line in order to construct a projective plane that will consist of a real plane plus a number of points at infinity, one for each direction in the plane; i.e., the projective plane will be a real plane plus a projective line of points at infinity.

Let $a, b, c, x, y, z \in \mathbb{R}$ such that neither (a, b, c) nor (x, y, z) are the zero vector:

$(x, y, z) \sim (a, b, c)$ if and only if there is $\lambda \in \mathbb{R}$ such that $x = \lambda a$, $y = \lambda b$, $z = \lambda c$.

We also define classes of similar vectors by

$[x, y, z] = \{(a, b, c) : a, b, c \in \mathbb{R}$ such that $(a, b, c) \neq \vec{0}$ and $(x, y, z) \sim (a, b, c)\}$.

Notice that, as before, the class $[x, y, z]$ contains all the points in the line in $\mathbb{R}^3$ that goes through (x, y, z) and $(0, 0, 0)$ except the origin. We define the projective plane to be the collection of all such lines:

$$\mathbb{P}^2(\mathbb{R}) = \{[x, y, z] : x, y, z \in \mathbb{R} \text{ such that } (x, y, z) \neq (0, 0, 0)\}.$$

If $z \neq 0$, then $(x, y, z) \sim (\frac{x}{z}, \frac{y}{z}, 1)$. Thus,

$$\mathbb{P}^2(\mathbb{R}) = \{[x, y, 1] : x, y \in \mathbb{R}\} \cup \{[a, b, 0] : a, b \in \mathbb{R}\}.$$

The points of the set $\{[x, y, 1] : x, y \in \mathbb{R}\}$ are in 1-to-1 correspondence with the real plane $\mathbb{R}^2$, and the points in $\{[a, b, 0] : a, b \in \mathbb{R}\}$ are called the points at infinity and form a $\mathbb{P}^1(\mathbb{R})$, a projective line.

One interesting consequence of the definitions is that any two parallel lines in the real plane $\{[x, y, 1]\}$ intersect at a point at infinity $[a, b, 0]$. Indeed, let $L : y = mx + b$ and $L' : y = mx + b'$ be distinct parallel lines in the real plane. If points in the real plane $\{[x, y, 1]\}$ correspond to lines in $\mathbb{R}^3$, then lines in the real plane correspond to *planes* in $\mathbb{R}^3$:

$$L = \{[x, y, z] : mx - y + bz = 0\}, \quad L' = \{[x, y, z] : mx - y + b'z = 0\}.$$

What is $L \cap L'$? The intersection points are those $[x, y, z]$ such that $mx - y + bz = mx - y + b'z = 0$, which implies that $(b - b')z = 0$. Since $L \neq L'$, we have $b \neq b'$ and, therefore, we must have $z = 0$. Hence

$$L \cap L' = \{[x, mx, 0] : x \in \mathbb{R}\} = \{[1, m, 0]\},$$

and so the intersection consists of a single point at infinity: $[1, m, 0]$.

C.3. Over an arbitrary field

The projective line and plane can be defined over any field. Let K be a field (e.g. $K = \mathbb{Q}, \mathbb{R}, \mathbb{C}$ or $\mathbb{F}_p$). The usual *affine plane* (or Euclidean plane) is defined by

$$\mathbb{A}^2(K) = \{(x, y) : x, y \in K\}.$$

The projective plane over K is defined by

$$\mathbb{P}^2(K) = \{[x, y, z] : x, y, z \in K \text{ such that } (x, y, z) \neq (0, 0, 0)\}.$$

As before, $(x, y, z) \sim (a, b, c)$ if and only if there is $\lambda \in K$ such that $(x, y, z) = \lambda \cdot (a, b, c)$.

C.4. Curves in the projective plane

Let K be a field and let C be a curve in affine space, given by a polynomial in two variables:

$$C : f(x, y) = 0$$

for some $f(x, y) \in K[x, y]$, e.g. $C : y^2 - x^3 - 1 = 0$. We want to extend C to a curve in the projective plane $\mathbb{P}^2(K)$. In order to do this, we consider the points in the curve (x, y) to be points in the plane $[\frac{x}{z}, \frac{y}{z}, 1]$ of $\mathbb{P}^2(K)$. Thus, we have

$$C : \left(\frac{y}{z}\right)^2 - \left(\frac{x}{z}\right)^3 - 1 = 0$$

or, equivalently, $zy^2 - x^3 - z^3 = 0$. Notice that the polynomial $F(x, y, z) = zy^2 - x^3 - z^3$ is homogeneous in its variables (each monomial has degree 3) and $F(x, y, 1) = f(x, y)$. The curve in $\mathbb{P}^2(K)$, given by

$$\widehat{C} : F(x, y, z) = zy^2 - x^3 - z^3 = 0,$$

is the curve we were looking for, which extends our original curve C in the affine plane. Notice that if the points $(x, y) \in C$, then $[x, y, 1] \in \widehat{C}$. However, there may be some extra points in $\widehat{C}$ which were not present in C, namely those points of $\widehat{C}$ at infinity. Recall that the points at infinity are those with $z = 0$, so $F(x, y, 0) = -x^3 = 0$ implies that $x = 0$ also, and the only point at infinity in $\widehat{C}$ is $[0, 1, 0]$.

In general, if $C \subseteq \mathbb{A}^2(K)$ is given by $f(x, y) = 0$ and d is the highest degree of a monomial in f, then $\widehat{C} \in \mathbb{P}^2(K)$ is given by

$$\widehat{C} : F(x, y, z) = 0,$$

where $F(x, y, z) = z^d \cdot f\left(\frac{x}{z}, \frac{y}{z}\right)$. Conversely, if $\widehat{C} : F(x, y, z) = 0$ is a curve in the projective plane, then $C : F(x, y, 1) = 0$ is a curve in the affine plane. In this case, C is the projection of $\widehat{C}$ onto the chart $z = 1$; we may also look at other charts, e.g., $x = 1$, which would yield a curve $C' : F(1, y, z) = 0$.

Here is another example. Let C be given by

$$C : y - x^2 = 0$$

so that C is a parabola. Then $\widehat{C}$ is given by

$$\widehat{C} : F(x, y, z) = z^2 f\left(\frac{x}{z}, \frac{y}{z}\right) = zy - x^2 = 0.$$

The curve $\widehat{C}$ has a unique point at infinity, namely $[0, 1, 0]$. This means that the two "arms" of the parabola meet at a single point at infinity. Thus, a parabola has the shape of an ellipse in $\mathbb{P}^2(K)$. How about hyperbolas? Let

$$C : x^2 - y^2 = 1.$$

Then $\widehat{C} : x^2 - y^2 = z^2$ and there are two points at infinity, namely $[1, 1, 0]$ and $[1, -1, 0]$. Thus, the four arms of the hyperbola in the affine plane meet in two points, and the hyperbola also has the shape of an ellipse in the projective plane $\mathbb{P}^2(K)$.

C.5. Singular and smooth curves

We say that a projective curve $C : F(x, y, z) = 0$ is singular at a point $P \in C$ if and only if $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$. In other words, C is singular at P if the tangent vector at P vanishes. Otherwise, we say that C is non-singular at P. If C is non-singular at every point, we say that C is a smooth (or non-singular) curve.

For example, $C : zy^2 = x^3$ is singular at $P = [0, 0, 1]$ because $F(x, y, z) = zy^2 - x^3$ and

$$\frac{\partial F}{\partial x} = -x^2, \ \frac{\partial F}{\partial y} = 2yz, \ \frac{\partial F}{\partial z} = y^2.$$

Thus, $\frac{\partial F}{\partial x}(P) = \frac{\partial F}{\partial y}(P) = \frac{\partial F}{\partial z}(P) = 0$ for $P = [0, 0, 1]$.

Here is another example. The curve $D : z^2y^2 = x^4 + z^4$ has partial derivatives

$$\frac{\partial F}{\partial x} = -4x^3, \ \frac{\partial F}{\partial y} = 2yz^2, \ \frac{\partial F}{\partial z} = 2y^2z - 4z^3.$$

Thus, if $P = [x, y, z] \in D(\mathbb{Q})$ is singular, then

$$-4x^3 = 0, \quad 2yz^2 = 0, \quad \text{and} \quad 2y^2z - 4z^3 = 0.$$

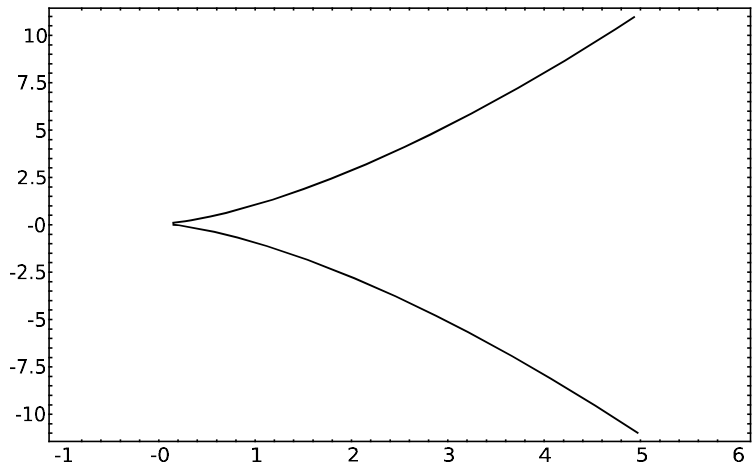


Figure 2. The chart $\{[x, y, 1]\}$ of the curve $zy^2 = x^3$.

The first two equalities imply that $x = 0$ and $yz = 0$ (what would happen if we were working over a field of characteristic 2, such as $\mathbb{F}_2$?). If $y = 0$, then $z = 0$ by the third equation, but $[0, 0, 0]$ is not a well-defined point in $\mathbb{P}^2(\mathbb{Q})$, so this is impossible. However, if $x = z = 0$, then y may take any value. Hence, $P = [0, 1, 0]$ is a singular point. Notice that the affine curve that corresponds to the chart $z = 1$ of D, given by $y^2 = x^4 + 1$, is non-singular at all points in the affine plane but is singular at a point at infinity, namely $P = [0, 1, 0]$.

An elliptic curve of the form $E : y^2 = x^3 + Ax + B$, or in projective coordinates given by $zy^2 = x^3 + Axz^2 + Bz^3$, is non-singular if and only if $4A^3 + 27B^2 \neq 0$. The quantity $\Delta = -16 \cdot (4A^3 + 27B^2)$ is called the discriminant of E.

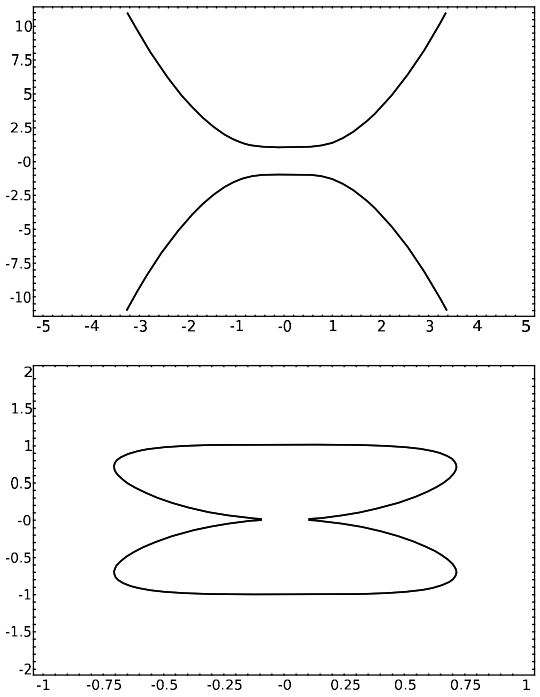


Figure 3. The chart $\{[x, y, 1]\}$ of the curve $z^2y^2 = x^4 + z^4$ (above, non-singular) and the chart $\{[x, 1, z]\}$ (below, the curve is singular).

Appendix D

The p-adic numbers

In this appendix we briefly introduce the p-adic integers $\mathbb{Z}_p$ and the p-adic numbers $\mathbb{Q}_p$. We strongly recommend [**Gou97**] to learn more about the p-adics.

Let $p \geq 2$ be a prime. The p-adic numbers may be thought of as a generalization of $\mathbb{Z}/p\mathbb{Z}$. The main difference is that the p-adic numbers form a ring of characteristic zero, while $\mathbb{Z}/p\mathbb{Z}$ has characteristic p. In $\mathbb{Z}/p\mathbb{Z}$ we only consider congruences modulo p, while in $\mathbb{Z}_p$ we consider congruences modulo p^n for all $n > 0$. The p-adic integers, denoted by $\mathbb{Z}_p$, are defined as follows:

$$\mathbb{Z}_p = \{(a_1, a_2, \ldots) : a_n \in \mathbb{Z}/p^n\mathbb{Z} \text{ such that } a_{n+1} \equiv a_n \bmod p^n\}.$$

In other words, a p-adic integer is an infinite vector $(a_n)_{n=1}^{\infty}$ such that the nth coordinate belongs to $\mathbb{Z}/p^n\mathbb{Z}$ and the sequence is coherent under congruences; i.e., $a_{n+1} \in \mathbb{Z}/p^{n+1}\mathbb{Z}$ reduces to the previous term a_n modulo p^n. For instance,

$$(2, 2, 29, 29, 272, 758, \ldots)$$

are the first few terms of a 3-adic integer; notice that all the coordinates are coherent with the previous terms under congruences modulo powers of 3. The vector $(2, 2, 2, 2, \ldots)$ is another element of $\mathbb{Z}_3$ (which we will denote simply by 2).

The p-adic integers have addition and multiplication operations, defined coordinate-by-coordinate:

$$(a_n)_{n=1}^{\infty} + (b_n)_{n=1}^{\infty} = (a_n + b_n \bmod p^n)_{n=1}^{\infty},$$

and

$$(a_n)_{n=1}^{\infty} \cdot (b_n)_{n=1}^{\infty} = (a_n \cdot b_n \bmod p^n)_{n=1}^{\infty}.$$

The reader should check that the sum and product of two coherent vectors is also coherent under congruences and, therefore, a new element of $\mathbb{Z}_p$. These operations make $\mathbb{Z}_p$ a commutative ring with identity element $1 = (1, 1, 1, 1, \ldots)$ and zero element $0 = (0, 0, 0, 0, \ldots)$.

For any prime $p \geq 2$, the p-adic integers contain a copy of $\mathbb{Z}$, where the integer m is represented by the element

$$m = (m \bmod p,\ m \bmod p^2,\ m \bmod p^3, \ldots).$$

For example, the number 200 in $\mathbb{Z}_3$ is given by

$$200 = (2, 2, 11, 38, 200, 200, 200, 200, 200, 200, \ldots).$$

Thus, we may write $\mathbb{Z} \subseteq \mathbb{Z}_p$ (see Exercise D.2.1). However, there are elements in $\mathbb{Z}_p$ that are not in $\mathbb{Z}$, so $\mathbb{Z} \subsetneq \mathbb{Z}_p$. Here is an example for $p = 7$: we are going to show that $\mathbb{Z}_7$, unlike $\mathbb{Z}$, contains an element whose square is 2 (which we will denote by "$\sqrt{2}$"). Indeed, 2 is a quadratic residue in $\mathbb{Z}/7\mathbb{Z}$, and 2 has two square roots, namely 3 and 4 modulo 7. A standard theorem of number theory shows that, hence, 2 is in fact a quadratic residue modulo 7^n for all $n \geq 1$. Thus, there exist integers a_n such that $a_n^2 \equiv 2 \bmod p^n$ for all $n \geq 1$. Moreover, it can also be shown that, if a_n is chosen, then there is $a_{n+1} \bmod p^{n+1}$ with $a_{n+1}^2 \equiv 2 \bmod p^{n+1}$ and $a_{n+1} \equiv a_n \bmod p^n$ (we say that a_n can be *lifted* to $\mathbb{Z}/p^{n+1}\mathbb{Z}$; see Exercise D.2.2). Indeed, here are the first few coordinates of an element α of $\mathbb{Z}_7$ such that $\alpha^2 = (2, 2, 2, \ldots)$:

$$\alpha = (3, 10, 108, 2166, 4567, \ldots).$$

Thus, α should be regarded as "$\sqrt{2}$ " inside $\mathbb{Z}_7$, and $-\alpha$ is another square root of 2.

The usual integers, $\mathbb{Z}$, are not a field because not every element has a multiplicative inverse (only ± 1 have inverses!). Similarly, the p-adic integers $\mathbb{Z}_p$ do not form a field either; e.g., $p = (p, p, p, \ldots)$ is not invertible in $\mathbb{Z}_p$, but many elements of $\mathbb{Z}_p$ *are* invertible. For

instance, if $p > 2$, then 2 is invertible in $\mathbb{Z}_p$ (in other words, there is a number $\frac{1}{2} \in \mathbb{Z}_p$). Indeed, the inverse of 2 is given by

$$\frac{1}{2} = \left(\frac{1+p}{2}, \frac{1+p^2}{2}, \ldots, \frac{1+p^n}{2}, \ldots \right).$$

For example, in $\mathbb{Z}_5$, the inverse of 2 is given by $(3, 13, 63, 313, \ldots)$. It is easy to see that if $\alpha = (a_n)_{n=1}^{\infty}$ with $a_1 \not\equiv 0 \bmod p$, then α is invertible in $\mathbb{Z}_p$. If $a_1 \equiv 0 \bmod p$, then α is not invertible. Moreover, for any $\alpha \in \mathbb{Z}_p$ there is an $r \geq 0$ such that $\alpha = p^r \beta$, where $\beta \in \mathbb{Z}_p$ is invertible.

Even though $\mathbb{Z}_p$ is not a field, we can embed $\mathbb{Z}_p$ in a field in the same way that $\mathbb{Z}$ sits inside $\mathbb{Q}$. We define the field of p-adic numbers by

$$\mathbb{Q}_p = \left\{ \frac{\alpha}{p^k} : k \geq 0 \text{ and } \alpha \in \mathbb{Z}_p \right\}.$$

Thus, every element of $\alpha \in \mathbb{Q}_p$ can be written as $\alpha = p^r \beta$ with $r \in \mathbb{Z}$ and an invertible $\beta \in \mathbb{Z}_p^{\times}$.

D.1. Hensel's lemma

The following results are used to show the existence of a solution to polynomial equations over *local fields*. Here we will only discuss the application to the p-adics, $\mathbb{Q}_p$ (which is an example of a local field). Notice the similarities with Newton's method.

Theorem D.1.1 (Hensel's Lemma)**.** *Let $p \geq 2$, let $\mathbb{Q}_p$ be the field of p-adic numbers and let $\mathbb{Z}_p$ be the p-adic integers. Let ν_p be the usual p-adic valuation (i.e., $\nu_p(p^e n) = e$ if $n \in \mathbb{Z}$ and $\gcd(n, p) = 1$). Let $f(x)$ be a polynomial with coefficients in $\mathbb{Z}_p$ and suppose there exist $\alpha_0 \in \mathbb{Z}_p$ such that*

$$\nu_p(f(\alpha_0)) > \nu_p(f'(\alpha_0)^2).$$

Then there exists a root $\alpha \in \mathbb{Q}_p$ of $f(x)$. Moreover, the sequence

$$\alpha_{i+1} = \alpha_i - \frac{f(\alpha_i)}{f'(\alpha_i)}$$

converges to α. Furthermore;

$$\nu_p(\alpha - \alpha_0) \geq \nu_p\left(\frac{f(\alpha_i)}{f'(\alpha_i)} \right) > 0.$$

Corollary D.1.2 (Trivial case of Hensel's lemma)**.** *Let $p \geq 2$, and let $\mathbb{Z}_p$ and $\mathbb{Q}_p$ be as before. Let $f(x)$ be a polynomial with coefficients in $\mathbb{Z}_p$ and suppose there exist $\alpha_0 \in \mathbb{Z}_p$ such that*

$$f(\alpha_0) \equiv 0 \ mod\ p, \quad f'(\alpha_0) \neq 0 \ mod\ p.$$

Then there exists a root $\alpha \in \mathbb{Q}_p$ of $f(x)$, i.e., $f(\alpha) = 0$.

Example D.1.3. Let p be a prime number greater than 2. Are there solutions to $x^2 + 7 = 0$ in the field $\mathbb{Q}_p$? If there are, -7 must be a quadratic residue modulo p. Thus, let p be a prime such that

$$\left(\frac{-7}{p}\right) = 1,$$

where $(\frac{\cdot}{p})$ is Legendre's quadratic residue symbol. Hence, there exist $\alpha_0 \in \mathbb{Z}$ such that $\alpha_0^2 \equiv -7 \bmod p$. We claim that $x^2 + 7 = 0$ has a solution in $\mathbb{Q}_p$ if and only if -7 is a quadratic residue modulo p. Indeed, if we let $f(x) = x^2 + 7$ (so $f'(x) = 2x$), the element $\alpha_0 \in \mathbb{Z}_p$ satisfies the conditions of the (trivial case of) Hensel's lemma. Therefore, there exists a root $\alpha \in \mathbb{Q}_p$ of $x^2 + 7 = 0$. ■

Example D.1.4. Let $p = 2$. Are there any solutions to $x^2 + 7 = 0$ in $\mathbb{Q}_2$? Notice that if we let $f(x) = x^2 + 7$, then $f'(x) = 2x$ and, for any $\alpha_0 \in \mathbb{Z}_2$, the number $f'(\alpha_0) = 2\alpha_0$ is congruent to 0 modulo 2. Thus, we cannot use the trivial case of Hensel's lemma (i.e., Corollary D.1.2).

Let $\alpha_0 = 1 \in \mathbb{Z}_2$. Notice that $f(1) = 8$ and $f'(1) = 2$. Thus,

$$3 = \nu_2(8) > \nu_2(2^2) = 2$$

and the general case of Hensel's lemma applies. Hence, there exists a 2-adic solution to $x^2 + 7 = 0$. ■

D.2. Exercises

Exercise D.2.1. Show that if q and t are distinct integers (in $\mathbb{Z}$), then their representatives in $\mathbb{Z}_p$ for any prime $p \geq 2$, given by $q = (q \bmod p^n)_{n=1}^{\infty}$ and $t = (t \bmod p^n)_{n=1}^{\infty}$, are also distinct in $\mathbb{Z}_p$.

Exercise D.2.2. Let $p > 2$ be a prime number.

(1) Let $b \in \mathbb{Z}$ with $\gcd(b, p) = 1$, and let $n \geq 1$. Suppose $a_n \in \mathbb{Z}$ such that $a_n^2 \equiv b \bmod p^n$. Show that there exists $a_{n+1} \in \mathbb{Z}$ such that $a_{n+1}^2 \equiv b \bmod p^{n+1}$ and $a_{n+1} \equiv a_n \bmod p^n$. (Hint: write $a_n^2 = b + kp^n$ and consider $f(x) = a_n + xp^n$. Find x such that $f(x)^2 \equiv b \bmod p^{n+1}$.)

(2) Suppose $a_1^2 \equiv b \bmod p$, where $\gcd(b, p) = 1$. Show that the vector $\alpha = (a_n)_{n=1}^{\infty}$, defined recursively by

$$a_{n+1} = a_n - \frac{a_n^2 - b}{2a_n} \bmod p^{n+1},$$

is a well-defined element of $\mathbb{Z}_p$ and, moreover, $\alpha^2 = b$, i.e.,

$$\alpha^2 = (b \bmod p,\ b \bmod p^2,\ b \bmod p^3, \ldots),$$

so α is a square root of b.

Exercise D.2.3. Find the first 4 coordinates of the 5-adic expansion of $\frac{1}{3}$ in $\mathbb{Z}_5$.

Exercise D.2.4. Find the first 4 coordinates of the 5-adic expansions of $\pm\sqrt{6}$ in $\mathbb{Z}_5$; i.e., find the first 4 coordinates of α and $-\alpha$ such that $\alpha^2 = 6$ in $\mathbb{Z}_5$.

Appendix E

Parametrization of torsion structures

In this appendix we provide one-parameter infinite families of elliptic curves with all the possible torsion subgroups that may occur for elliptic curves over $\mathbb{Q}$. The main reference for this appendix is **[Kub76]**, Table 3, p. 217.

In each table below, Figure 1 and Figure 2, we provide elliptic curves $E_{a,b}$ whose equations depend on two rational parameters $a, b \in \mathbb{Q}$, and such that the torsion subgroup $E_{a,b}(\mathbb{Q})_{\text{tors}}$ has a given subgroup G; i.e., the full torsion subgroup contains G as a subgroup, but may be larger in certain cases (see Example E.1.1 below).

The families that appear in Figure 1 depend on two independent parameters a, b, and the only condition that needs to be satisfied is that the discriminant $\Delta_{a,b}$ of $E_{a,b}$ must be non-zero. This condition on the discriminant is given in the second column of the table.

The families that appear in Figure 2 depend on one single rational parameter $t \in \mathbb{Q}$, and a and b are rational functions in the variable t. The curves $E_{a,b}$ that appear in this table are all of the form

$$E_{a,b} : y^2 + (1-a)xy - by = x^3 - bx^2.$$

The point $(0, 0)$ is a torsion point of the maximal order in the group. The discriminant $\Delta_{a,b}$ of $E_{a,b}$ is always assumed to be non-zero.

$E_{a,b}/\mathbb{Q}$	$\Delta_{a,b} \neq 0$	G
$y^2 = x^3 + ax^2 + bx$	$a^2b^2 - 4b^3 \neq 0$	$\mathbb{Z}/2\mathbb{Z}$
$y^2 + axy + by = x^3$	$a^3b^3 - 27b^4 \neq 0$	$\mathbb{Z}/3\mathbb{Z}$
$y^2 = x(x+a)(x+b)$	$0 \neq a \neq b \neq 0$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2\mathbb{Z}$

Figure 1. Two-parameter families of elliptic curves $E_{a,b}/\mathbb{Q}$ such that $E_{a,b}(\mathbb{Q})_{\text{tors}}$ has a subgroup G.

Curves of the form $E_{a,b} : y^2 + (1-a)xy - by = x^3 - bx^2$		
a	b	G
$a = 0$	$b = t$	$\mathbb{Z}/4\mathbb{Z}$
$a = t$	$b = t$	$\mathbb{Z}/5\mathbb{Z}$
$a = t$	$b = t + t^2$	$\mathbb{Z}/6\mathbb{Z}$
$a = t^2 - t$	$b = t^3 - t^2$	$\mathbb{Z}/7\mathbb{Z}$
$a = \frac{(2t-1)(t-1)}{t}$	$b = (2t-1)(t-1)$	$\mathbb{Z}/8\mathbb{Z}$
$a = t^2(t-1)$	$b = t^2(t-1)(t^2-t+1)$	$\mathbb{Z}/9\mathbb{Z}$
$a = -\frac{t(t-1)(2t-1)}{t^2-3t+1}$	$b = \frac{t^3(t-1)(2t-1)}{(t^2-3t+1)^2}$	$\mathbb{Z}/10\mathbb{Z}$
$a = \frac{t(1-2t)(3t^2-3t+1)}{(t-1)^3}$	$b = -a \cdot \frac{2t^2-2t+1}{t-1}$	$\mathbb{Z}/12\mathbb{Z}$
$a = 0$	$b = t^2 - \frac{1}{16}$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/4\mathbb{Z}$
$a = \frac{10-2t}{t^2-9}$	$b = \frac{-2(t-1)^2(t-5)}{(t^2-9)^2}$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z}$
$a = \frac{(2t+1)(8t^2+4t+1)}{2(4t+1)(8t^2-1)t}$	$b = \frac{(2t+1)(8t^2+4t+1)}{(8t^2-1)^2}$	$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z}$

Figure 2. One-parameter families of elliptic curves $E_{a,b}/\mathbb{Q}$ such that $E_{a,b}(\mathbb{Q})_{\text{tors}}$ has a subgroup G.

Example E.1.1. For each $t \in \mathbb{Q}$, according to Figure 2, the torsion subgroup of the elliptic curve $E_{t,t} : y^2 + (1-t)xy - ty = x^3 - tx^2$ contains $G = \mathbb{Z}/5\mathbb{Z}$ as a subgroup, as long as the discriminant $\Delta_{t,t} = t^5(t^2 - 11t - 1)$ is non-zero (thus, $\Delta_{t,t} = 0$ if and only if $t = 0$). In other words, the point $(0, 0)$ of $E_{t,t}$ is a torsion point of order 5.

Notice, however, that this does not imply that the torsion subgroup of $E_{t,t}(\mathbb{Q})$ is identical to $\mathbb{Z}/5\mathbb{Z}$. For instance, let $t = 12$. The torsion subgroup of the elliptic curve

$$E_{12,12} : y^2 - 11xy - 12y = x^3 - 12x^2$$

is isomorphic to $\mathbb{Z}/10\mathbb{Z}$. The point $(0, 0)$ is a point of order 5, but the point $(-6, -18)$ has exact order 10.

Example E.1.2. According to Figure 2, each curve in the family

$$y^2 + (1 + t - t^2)xy + (t^2 - t^3)y = x^3 + (t^2 - t^3)x^2$$

has a torsion point of exact order 7, namely $P = (0, 0)$, as long as the discriminant $\Delta = t^7(t - 1)^7(t^3 - 8t^2 + 5t + 1)$ is non-zero, which can only happen for the rational values $t = 0$ and $t = 1$. By Mazur's Theorem 2.5.2, the only possible torsion subgroup for an elliptic curve over $\mathbb{Q}$ that contains $\mathbb{Z}/7\mathbb{Z}$ as a subgroup is $\mathbb{Z}/7\mathbb{Z}$ itself. Thus, the torsion subgroup of each elliptic curve in this family is exactly $\mathbb{Z}/7\mathbb{Z}$.

Similarly, if $E_{a,b}$ is an elliptic curve in one of the families in Figure 2 that correspond to G in the list

$$\mathbb{Z}/7\mathbb{Z},\ \mathbb{Z}/9\mathbb{Z},\ \mathbb{Z}/10\mathbb{Z},\ \mathbb{Z}/12\mathbb{Z},\ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/6\mathbb{Z},\ \text{ or } \ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/8\mathbb{Z},$$

then the torsion subgroup of $E_{a,b}(\mathbb{Q})$ must be exactly G.