Elliptic Curves

Dhruv Goel

June 2026

Contents

Preface

These are lecture notes for a course on the arithmetic of elliptic curves aimed at peer mentors and counselors that I taught at Ross/Ohio 2026. The course covers the basic theory of elliptic curves at the level of [1] or [2]. The course assumes only a tiny bit more than familiarity with the material on the Ross first-year sets, and a tiny bit more.

Chapter 1

Lecture Notes

1.1 26/06/15 - Introduction I: Diophantine Equations, Plane Curves, and Bachet's Example

One of the oldest and most natural questions in mathematics goes back at least to the 3rd century CE, and probably much more. This question is the study of solutions to *Diophantine equations*, named after Diophantus of Alexandria, who "formulated and solved many such problems" ([2, p. xv]).

A *Diophantine equation* is an equation of the form

$$f(x_1, \dots, x_n) = 0$$

where $n \in \mathbb{Z}_{\geq 1}$, and $f \in \mathbb{Q}[x_1, \dots, x_n]$ is a polynomial in n variables with rational coefficients. To *solve it* means to find all rational solutions $(x_1, \dots, x_n)$ to the equation, although what it means to "find" all solutions can indeed be quite subjective. What is usually meant and hoped for is some sort of complete description or parametrization of such solutions. Variations of this problem ask for solutions over a ring R to similar equations defined over R instead of $\mathbb{Q}$ for various rings R such as $\mathbb{Z}$, a number field K, or its ring of integers $\mathcal{O}_K$. Of course, there is a rich interplay between solutions over $\mathbb{Z}$ and $\mathbb{Q}$; see 1.1.3 for one example.

There are (at least) two measures of the complexity of a given Diophantine equation: the number n of variables, and the degree $d \in \mathbb{Z}_{\geq 1}$ of the polynomial f. Let's study some examples for small n and d.

Example 1.1.1 (Single-Variable Equations)**.** The case of $n = 1$ corresponds to equations of a single variable $x = x_1$.

(a) When $d = 1$, we are looking at linear equations in one variable, which are of the form $ax + b = 0$ for some $a, b \in \mathbb{Q}$ with $a \neq 0$. I trust the reader knows how to solve such equations.

(b) When $d = 2$, we are looking at solving quadratic equations over $\mathbb{Q}$, which look like $ax^2 + bx + c = 0$ for $a, b, c \in \mathbb{Q}$ with $a \neq 0$. Our insistence on asking for rational solutions means that there are essentially four things that can happen, depending on the discriminant $\Delta = b^2 - 4ac$; these are:

  (i) There is a unique root of "multiplicity two"; this happens iff $\Delta = 0$ (e.g., $x^2 - 4x + 4 = 0$).

  (ii) The two distinct roots are non-real complex conjugates; this happens iff $\Delta < 0$ (e.g., $x^2 - 5x + 5 = 0$).

  (iii) The two distinct roots are irrational but real Galois conjugates; this happens iff $\Delta > 0$ but Δ is not a perfect square (e.g., $x^2 - 5x + 7 = 0$).

  (iv) The two distinct roots are both rational; this happens iff Δ is a perfect square (e.g., $x^2 - 5x + 6 = 0$).

(c) When $d = 3, 4$, more complicated formulae exist for the solutions, but these are not always helpful. Sometimes we can use elementary calculus to estimate the number of real roots (see, e.g., 2.1.1). When $d \geq 5$, a famous theorem in Galois theory says that no such formula in terms of radicals even exists in general [TOCITE]. However, as long as we are only interested in rational solutions, there exists a complete algorithmic procedure to find them all; this is contained in 1.1.2.

Theorem 1.1.2 (Rational Root)**.** Let $d \in \mathbb{Z}_{\geq 1}$, let $a_0, \dots, a_d \in \mathbb{Z}$, and consider the equation

$$a_0 x^d + a_1 x^{d-1} + \cdots + a_d = 0.$$

Suppose that $x = p/q$ is a rational solution to this equation, where $p, q \in \mathbb{Z}$ are coprime and $q \neq 0$. Then $p \mid a_d$ and $q \mid a_0$. In particular, if $a_0 = 1$ (i.e., the equation is monic), then every rational root is an integer.

Remark 1.1.3. Do you see why this gives a complete algorithmic answer (in theory) of finding all rational (and hence integral) roots to a single-variable rational polynomial?

Proof of 1.1.2. Since p/q is a solution to the equation, we have

$$a_0 p^d + a_1 p^{d-1} q + \cdots + a_d q^d = 0.$$

Therefore, $q \mid a_0 p^d$; since $(p, q) = 1$, this forces $q \mid a_0$. The part with p is similar and left to the reader. ∎

Exercise 1.1.4. Finish the proof by showing that $p \mid a_d$. You might have to be slightly more careful than for q because p could be zero; what would that mean?

We didn't really use any special properties of the integers here–other than the fact that they admit unique factorization. Therefore, the same result and proof works also with $\mathbb{Z}$ replaced by any unique factorization domain (UFD) R; see 2.1.2.

Example 1.1.5 (Plane Curves)**.** The next simplest case of $n = 2$ corresponds to *plane curves*. These correspond to a single equation relating two variables $x = x_1$ and $y = x_2$; the corresponding locus can be plotted as a subset of the Euclidean plane $\mathbb{R}^2$.

(a) When $d = 1$, we are talking about linear equations of the form $ax + by + c = 0$, with $a, b, c \in \mathbb{Q}$ and not both a and b zero. The theory of rational solutions here is familiar from high-school coordinate geometry, and the theory of integer solutions is also well-understood, although it might need a little thought (2.1.3).

(b) When $d = 2$, we are looking at *conic sections*, so named because these curves can be obtained by slicing a (doubly infinite) cone by planes at various angles (see this excellent video by 3Blue1Brown illustrating this). Examples include circles (e.g., $x^2 + y^2 = 1$), parabolae (e.g., $y = x^2$), ellipses (e.g., $x^2+2y^2 = 1$), hyperbolae (e.g., $xy = 1$ or $3x^2-2y^2 = 1$), pairs of lines lines (e.g., $y^2-x^2 = 0$), and even "doubled lines" (e.g., $x^2 = 0$) (whatever that means). Plot these on Desmos! Here, the theory of rational solutions is very well-understood. These curves may or may not have any rational points (e.g., $x^2 + y^2 = 1$ and $x^2 + 3y^2 = 2$ respectively; see 2.1.4(a)), but if there is at least one, then we understand all of them very well. We will discuss this more next time.

(c) When $d = 3$, we are looking at *cubic curves*. Again these may or may not have any rational points at all (e.g., $y^2 = x^3 + 1$ and $x^3 + 2y^3 = 4$ respectively; see 2.1.4(b)), but the theory of cubic curves, even those with rational points (these are called *elliptic curves*), is quite subtle. The object of this summer will be to study precisely these objects.

We mention in closing that the cases $d \geq 4$ or $n \geq 3$ are even more complicated; the study of such diophantine equations properly belongs to the field of math known as *arithmetic geometry*. We will not say much about such objects in general, except perhaps for some specific examples.

In 1.1.5, more or less the same discussion can be had with $\mathbb{Q}$ replaced with essentially any field K. (We will systematically make this replacement and work with general fields starting very soon.)

For another Diophantine equation, consider

Example 1.1.6 (Bachet's Example)**.** When does a cube differ from a square by 1? Phrased algebraically, this amounts to solving the equation

$$y^2 = x^3 + 1$$

in pairs of integers (x, y). Five such solutions are $(-1, 0), (0, \pm 1)$, and $(2, \pm 3)$. Are there more? (The answer can be found at [TODO].)

More generally, given a $c \in \mathbb{Z}$, we can ask for integer solutions to the equation

$$y^2 = x^3 + c. \tag{1.1.7}$$

The case $c = -2$ was studied extensively by French mathematicians Pierre de Fermat [TODO] and [TODO] Bachet [TODO] in the early 17^{th} century; see the discussion of the history of this problem in [3, Ch. 17, §10]. Bachet observed in 1621 that if (x_0, y_0) is a solution to (1.1.7) with $y_0 \neq 0$, then so is

$$(x_1, y_1) := \left(\frac{x_0^4 - 8cx_0}{4y_0^2}, \frac{-x_0^6 - 20cx_0^3 + 8c^2}{8y_0^3} \right). \tag{1.1.8}$$

This is quite crazy! It's very easy algebra to verify that it works, but how does one come up with such a thing?

Let's assume this for a moment. Given a "seed solution" (x_0, y_0), we can iterate the map $(x_0, y_0) \mapsto (x_1, y_1)$ to produce an infinite sequence (x_n, y_n) of solutions to 1.1.7. For instance, if $c = 1$, then starting from the seed $(x_0, y_0) = (0, 1)$ returns the sequence

$$(0, 1), \quad (0, 1), \quad (0, 1), \ldots.$$

Boring! If we start from the seed $(2,3)$, then we get the sequence

$$(2,3), \quad (0,-1), \quad (0,-1), \ldots.$$

That's slightly more interesting; it's cool that it stabilizes.

Let's take a different c. If we take $c=3$ and the seed $(x_0,y_0)=(1,2)$ gets us the sequence

$$(1,2), \quad \left(\frac{-23}{4^2}, \frac{11}{4^3}\right), \quad \left(\frac{2540833}{88^2}, \frac{4050085583}{88^3}\right), \ldots.$$

That's very interesting! These numbers seem to get pretty huge pretty quickly.We will prove later (1.17.5) that the number of decimal digits in the numerator and denominator in lowest terms roughly quadruples with each iteration.

When do we get infinitely many distinct solutions in the sequence (x_n,y_n)? It is a theorem of Fueter from 1930 (reproven by Mordell in 1966), which we will prove later [TODO], that if $c \notin \{1,-432\}$, then for any seed (x_0,y_0) with $x_0y_0 \neq 0$, the sequence (x_n,y_n) gives us infinitely many rational solutions.

What about integer solutions? It is a theorem of Thue from 1908 that for any nonzero c, there are only finitely many integer solutions (x,y) to 1.1.7. Therefore, even if there are infinitely many solutions produced by Bachet's formula (as for $c \notin \{1,-432\}$), only finitely many can be integral. We may or may not get to this result by the end of the course; if you're particularly interested in seeing a proof, I can try to incorporate it. [TODO]

All of this still begs the question: where does Bachet's formula (1.1.8) really come from? The beautiful answer is that it really comes from *geometry*.[1]

The equation $y^2=x^3+c$ defines a *cubic curve*, say C, in the plane. This cubic curve has the property that it is *smooth*, i.e., has a well-defined tangent line at every point; equivalently, at every point of the curve, at least one of the partial derivatives $2y$ or $3x^2$ is nonzero.[2] Given a point $P_0=(x_0,y_0)$ on this cubic curve C, we can draw the tangent line L to C at P_0. Then an elementary case of *Bézout's Theorem* tells us that L will intersect C in a third point P_1 on C; this is quite clear if we parametrize L and restrict the equation of C via this parametrization to get a cubic equation in one variable whose roots correspond to the intersection points. "Two" of these intersection points are at P_0 (i.e., P_0 corresponds to a double root of this cubic equation), and since a cubic has three roots, there must be a third point–this is P_1. The coordinates of P_1 are exactly the solution (x_1,y_1) expressed in Bachet's formula. In older texts, this procedure $P_0 \mapsto P_1$ is called the "chord and tangent process".

Let's work this out explicitly. Suppose we are given a point (x_0,y_0) on the curve (1.1.7). An easy exercise in elementary geometry or calculus tells us that the slope of the curve 1.1.7 at (x_0,y_0) when $y_0 \neq 0$ is $(3x_0^2)/(2y_0)$, so the equation to the tangent line is

$$y = \left(\frac{3x_0^2}{2y_0}\right)x + \left(y_0 - \frac{3x_0^3}{2y_0}\right). \tag{1.1.10}$$

Substitute this expression for y into 1.1.7 to get the cubic equation

$$x^3 - \left[\left(\frac{3x_0^2}{2y_0}\right)x + \left(y_0 - \frac{3x_0^3}{2y_0}\right)\right]^2 + c = 0.$$

This cubic equation has a double root at $x=x_0$. Let's call its third root x_1. Then Vieta's formula allow us to compute the the product of the roots as the negative of the constant term, which gives us

$$x_0^2x_1 = \left(y_0 - \frac{3x_0^3}{2y_0}\right)^2 - c,$$

giving an expression for x_1 in terms of x_0, y_0, and c. Plugging this into 1.1.10 gives us our y_1. A little bit of simplification using that (x_0,y_0) lies on 1.1.7–left to the reader (2.1.5)–then expresses (x_1,y_1) in the promised form 1.1.8.

[1] I say *really*, because this is probably not how Bachet found it.

[2] This requires us to work over a field of characteristic other than 2 or 3, so $\mathbb{Q}$ is certainly OK.

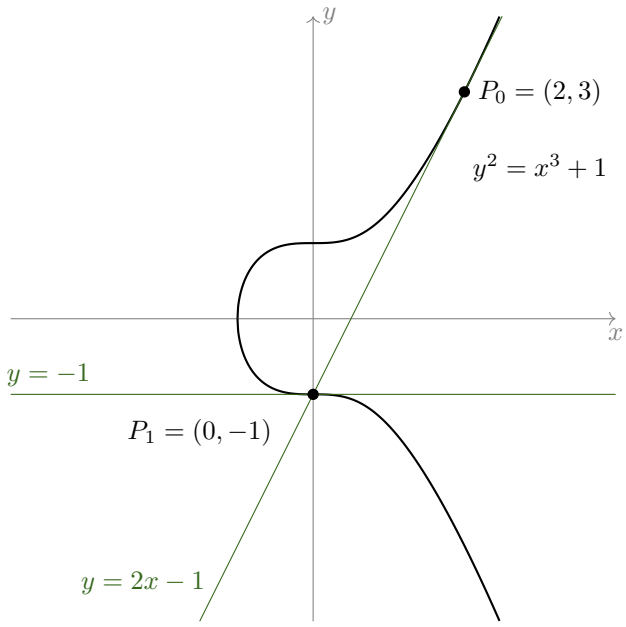


Figure 1.1.9: The "chord and tangent process" which yields Bachet's formula, illustrated for $c = 1$ and seed $P_0 = (x_0, y_0) = (2, 3)$.

The above procedure also has an analog when you are given two points, say P, Q on C: you take the line L joining P and Q, and consider the third intersection points R of P and Q with C (this is the "chord" part of the "chord and tangent process"). It is a fascinating conjecture of Poincaré from around 1901, proven by Mordell 1922 (and then generalized by Weil to number fields in 1928), which says that given any such smooth cubic C over $\mathbb{Q}$, there are finitely many "seed points" $P_i \in C(\mathbb{Q})$ such that *every* other rational point on C can be obtained by iteratively applying the chord and tangent processes to the points P_i. This is now known as the Mordell Theorem or the Mordell-Weil Theorem, and one of our goals for this summer will be to prove this. This cool interplay between algebra, number theory, and geometry is the tip of the iceberg which is the theory of elliptic curves, and what we shall study this summer.

1.2 26/06/16 - Introduction II: Pythagorean Triples and Fermat for Exponent 3

Let us now study some special cases of the *Fermat problem*, the statement of which is colloquially know as *Fermat's Last Theorem*. Recall that this says that if $n \in \mathbb{Z}_{\geq 3}$ and $X, Y, Z \in \mathbb{Z}$ are such that $X^n + Y^n = Z^n$, then the solution (X, Y, Z) is trivial, i.e., $XYZ = 0$. We will discuss and prove the cases $n = 3$ and $n = 4$.

1.2.A Pythagorean Triples

Before we do that however, it is helpful to review the case $n = 2$ of Pythagorean triples. Here we are trying to solve $X^2 + Y^2 = Z^2$ in the integers. The first thing we note is that this is a *homogeneous equation*; this means that if (X, Y, Z) is a solution, then so is $(\lambda X, \lambda Y, \lambda Z)$ for any λ. This is very helpful, and allows some geometric and arithmetic simplifications. We will develop the language of projective curves in the next lecture to deal with the geometric situation more generally, and we will deal with the arithmetic implications below, but for now it suffices to note that if $Z \neq 0$ (the solutions with $Z = 0$ are easy to understand!), then we can divide throughout by Z to get that $(x, y) := (X/Z, Y/Z)$ satisfies

$$x^2 + y^2 = 1.$$

This curve is familiar to us as the circle C. Thus, Pythagorean triples give rise to points on the circle. (This is a fundamental insight going back even to Descartes!) What is more, if X, Y, Z are integers, then x and y are rational numbers; therefore, if we have a good understanding of rational points on the circle $x^2 + y^2 = 1$, then we will have a way to find all Pythagorean triples. For instance, if we could parametrize all points on the circle, then we can hope to parametrize all Pythagorean triples.

Now the circle is a curve of degree two, i.e., a conic. This has the important consequence that a general line meets it in *two points* (some are tangent, and some might need you to allow complex numbers). This leads us to the following simple strategy: if we pick any rational point P on C and project from it to a general line L, then we will get (essentially) a bijective correspondence between points on this line and on the circle. (Make sure you believe this!) Since it is very easy to parametrize points on a line, this gives us a way of parametrizing points on the conic curve. Let's see this work out in practice.

Take the point P to be $(-1, 0)$ and the line L to be the x-axis. Given a point (x_0, y_0) on the circle C, the line joining P and (x_0, y_0) meets the line L in the point $(0, t)$ where $t = y_0/(x_0 + 1)$. Note that this does not work when $x_0 = -1$ corresponding to P itself, when the the limiting line is the tangent line $x = -1$, which does not meet L–in the affine plane. We'll figure out how to fix this next time.

Going the other way, we can parametrize points on L as $(0, t)$ for varying t. The line joining $(-1, 0)$ and $(0, t)$ has equation

$$y = tx + t,$$

which we would like to intersect with the circle. Plugging this in to the equation for the cirlce gives us the quadratic equation

$$x^2 + (tx + t)^2 = 1,$$

which factors as

$$(x + 1)((t^2 + 1)x + (t^2 - 1)) = 0.$$

One root is the expected $x = -1$; it's the other one which we are after, which has $x = (1 - t^2)/(1 + t^2)$. Using $y = tx + t$, this gives us the other point of intersection

$$(x, y) = \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right).$$

(Those who have seen a little trigonometry might recognize this is as the formula for $(\cos\theta, \sin\theta)$ in terms of $\tan(\theta/2)$; why is that?). These operations are visibly inverses to each other and the resulting maps

$$(x, y) \mapsto y/(x + 1) \text{ and } t \mapsto \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right)$$

are almost inverses to each other (check!). The only point that gives us some trouble is the point $P = (-1, 0)$ itself, which should somehow correspond to "$t = \infty$". Now, the above discussion works over any field (but see 1.2.1). In particular, if t is a rational number, then so is (x, y), and conversely. In this way, we have completely figured out all rational points on C.

Remark 1.2.1. So far we haven't really used many properties of the rational numbers. The same algebra can be carried out over more or less any field K (even if it's not easy visualize what that might mean), with some important caveats. Here are two things to think about:

(a) What (if anything) goes wrong when you work with a field of characteristic two such as $K = \mathbb{F}_2$ instead?
(b) If we work over the field $K = \mathbb{C}$, what happens to the values $t = \pm i$ of the parameter t?

Remark 1.2.2. Note that we could have "dehomogenized with respect to other variables" too. For instance, if $Y \neq 0$, then we may set $(s, t) := (X/Y, Z/Y)$ instead to get instead the equation $s^2 + 1 = t^2$, or equivalently $t^2 - s^2 = 1$. This would have led to a parametrization similar to the one above. This additional symmetry–while not useful here–can be extremely useful in other contexts; we will study this systematically next time.

Let us see what consequence this has for the arithmetic. Suppose we are interested in finding integer solutions X, Y, Z to $X^2 + Y^2 = Z^2$ with $X, Y, Z \neq 0$. We may assume by flipping signs if needed that $X, Y, Z > 0$. We may also divide throughout by the greatest common divisor of all X, Y, Z to assume that X, Y, Z have no common factors as a *triple*. This has the consequence that X, Y, Z have no common factors *pairwise*: if a prime p divides two of X, Y, Z, then the equation $X^2 + Y^2 = Z^2$ forces it to divide the third. (Therefore, things like $(6, 10, 15)$ will never work.) With these simplifications, we can show

Theorem 1.2.3. Let X, Y, Z be pairwise coprime positive integers such that $X^2 + Y^2 = Z^2$. Then there are unique coprime $m, n \in \mathbb{Z}$ of different parity with $m > n > 0$ such that exactly one of (X, Y, Z) or (Y, X, Z) is

$$(m^2 - n^2, 2mn, m^2 + n^2).$$

Proof. The point $(x, y) = (X/Z, Y/Z)$ lies on the circle $C : x^2 + y^2 = 1$ and is different from $(-1, 0)$, so the above discussion tells us that there is a unique $t \in \mathbb{Q}$ such that

$$\left(\frac{X}{Z}, \frac{Y}{Z}\right) = \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right).$$

Since x and y are positive, we conclude that $0 < t < 1$ (check!). Write $t = n/m$ for coprime positive $m, n \in \mathbb{Z}$ such that $m > n > 0$, so that

$$\left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2}\right) = \left(\frac{m^2 - n^2}{m^2 + n^2}, \frac{2mn}{m^2 + n^2}\right).$$

We claim that the fractions on the right are in their lowest terms (up to small factors of two; see below). Why is this?

Let's analyze the parities of m and n. Since they are coprime, they can't both be even. If they have different parity, then $m^2 \pm n^2$ are odd, and this forces

$$\gcd(m^2 + n^2, m^2 - n^2) = \gcd(2mn, m^2 + n^2) = 1.$$

i.e., these fractions are in lowest form (check!). This proves that (X, Y, Z) is the of the form desired.

Finally, can they both be odd? If this is the case, then the integers $m_1 = (m + n)/2$ and $n_1 := (m - n)/2$ are coprime, satisfy $m_1 > n_1 > 0$, have different parity (check!), and further we have that

$$\left(\frac{m^2 - n^2}{m^2 + n^2}, \frac{2mn}{m^2 + n^2}\right) = \left(\frac{2m_1 n_1}{m_1^2 + n_1^2}, \frac{m_1^2 - n_1^2}{m_1^2 + n_1^2}\right),$$

and so the argument from the previous paragraph implies $(Y, X, Z) = (m_1^2 - n_1^2, 2m_1 n_1, m_1^2 + n_1^2)$, finishing the proof. ∎

Exercise 1.2.4. Fill in the details in the above proof: show that if $m, n \in \mathbb{Z}$ are coprime and of different parity, then $\gcd(m^2+n^2, m^2-n^2) = \gcd(2mn, m^2+n^2) = 1$, and show that if $m > n > 0$ are both odd, then $m_1 = (m+n)/2$ and $n_1 := (m-n)/2$ are coprime and have different parity.

In the above study, we didn't really use any properties of the curve C except that it is of degree two–and that it had a rational point P to get our method started. Therefore, the same argument applies to any conic C with a rational point $P \in C(\mathbb{Q})$. This allows us to give complete parametrizations of all rational points of such objects; see, for instance, 2.1.6(a). (Get a hold on the integer points this way might be a little harder; see 2.1.6(b).)

Clearly, then we would have a complete understanding of rational solutions to all conic sections C if we could just prove that they all contained some rational point P. Sadly, this is not the case (2.1.4(a)). How can we figure out when a given conic has a rational point, and when it doesn't?

Example 1.2.5. Let's look at a few examples.

(a) Does the curve
$$4x^2 + 12xy + 9y^2 + 1 = 0$$
have any rational points? No. This is because it can be written as $(2x+3y)^2+1 = 0$, which doesn't have any real solutions (x, y), let alone rational ones. We say such curves are *locally obstructed at* ∞ or *locally obstructed over* $\mathbb{R}$.

(b) Does the curve
$$x^2 + xy + y^2 = 2$$
have any rational points? No, but this is not because it doesn't have any real points (e.g., $(\sqrt{2}, 0)$ works). This is more subtle. Suppose there is a rational solution (x, y). Write $x := X/Z$ and $y := Y/Z'$ for integers $X, Y, Z, Z' \in \mathbb{Z}$ with $Z, Z' > 0$ and $\gcd(X, Z) = \gcd(Y, Z') = 1$. Plugging this into the equation and multiplying through by $Z^2(Z')^2$ gives us
$$X^2(Z')^2 + XYZZ' + Y^2Z^2 = 2Z^2(Z')^2.$$
This equation implies that $Z \mid (Z')^2$ and so by symmetry that $Z' \mid Z^2$. This is enough to conclude that Z and Z' have the same prime factors, but this is not enough to conclude that $Z = Z'$ (consider $2^2 3^3$ vs. $2^3 3^2$). We need to be a little more careful. For this, let p be a prime dividing both Z and Z', and let $\alpha := v_p(Z)$ (resp. $\alpha' := v_p(Z')$), so $\alpha, \alpha' \geq 1$. We have that $v_p(X) = v_p(Y) = 0$, that $\alpha \leq 2\alpha'$ and $\alpha' \leq 2\alpha$. We have to show that $\alpha = \alpha'$. Suppose to the contrary that $\alpha \neq \alpha'$; we may then assume without loss of generality that $\alpha' > \alpha$ (the other case being symmetrical). In this case, the p-adic valuation of the terms on the left is $2\alpha', \alpha + \alpha'$, and 2α, which forces the sum to have p-adic valuation 2α. The right hand side has v_p equal to either $2\alpha + 2\alpha$ if $p \neq 2$, and $2\alpha + 2\alpha' + 1$ if $p = 2$, and we easily see that none of these options work. Thus, we must have $\alpha = \alpha'$. Since this is true for all p and $Z, Z' > 0$, we conclude that $Z' = Z$, and so we can remove it to get the homogeneous equation
$$X^2 + XY + Y^2 = 2Z^2 \tag{1.2.6}$$
we were hoping for. Now we are almost done. Let's consider the parities of X and Y. Well, it cannot be that both of them are odd or that precisely one is odd, because then the left side of 1.2.6 is odd. Therefore, both of them must be even, say $X = 2X_1$ and $Y = 2Y_1$. This then gives us the equation $2(X_1^2 + X_1Y_1 + Y_1^2) = Z^2$, which forces Z to be even as well–but $\gcd(X, Z) = 1$, and this is a contradiction. This proves that there couldn't have been any solutions to begin with.
We say that the curve $x^2 + xy + y^2 = 2$ is *locally obstructed at 2*. Note that this does not mean it has no solutions mod 2 (i.e., that $C(\mathbb{F}_2) = \emptyset$, which is false); rather, it means that it has no *2-adic solutions*, i.e., $C(\mathbb{Q}_2) = \emptyset$. (Don't worry if you don't know yet what that means.)

Similarly, curves can be locally obstructed at other primes; see 2.1.7.

For right now, you can take local obstruction at a prime $p > 0$ to mean that "the arithmetic doesn't work out at p". Clearly, if there is a local obstruction at any prime (including the "prime" $p = \infty$ corresponding to $\mathbb{Q}_\infty = \mathbb{R}$), then there is no rational solution. But the converse is not that clear. Can we have a curve that has solutions *everywhere locally*, but none globally?

It turns out that this "failure of the local-to-global principle" cannot happen for conics, but it can happen for cubics. The fact that it cannot happen to conics is known as the Hasse-Minkowski

Theorem. This holds more generally not just for conic curves, but for quadratic forms in any number of variables. Proving this would take us a little far afield, so I will leave the simplest case of quadratic equations as an exercise (2.1.10), and refer the reader to [4, Ch. IV, Thm. 8] for the general proof.

The really interesting thing I want to draw your attention to, however, is that this result fails in general for curves of degree ≥ 3: there are cubic curves which have solutions in $\mathbb{Q}_p$ for all $p \leq \infty$, but no solutions in $\mathbb{Q}$. One of the simplest examples of this sort is the curve due to Selmer with homogeneous equation

$$3X^3 + 4Y^3 + 5Z^3 = 0. \tag{1.2.7}$$

(This is the homogeneous equation; the inhomogeneous equation is $3x^3 + 4y^3 = -5$.) We will revisit this curve later in the course, and develop tools to study such "obstructions to the Hasse-Minkowski principle". For now, let us turn to

1.2.B Fermat for Exponent 3

The next case of Fermat's problem is $n = 3$, where we are solving $X^3 + Y^3 = Z^3$. Again, this is a homogeneous equation. It is now believed that Fermat did not have a proof for this case, and that the first proof was given by Euler in 1770 (although see the discussion on the Wikipedia page). Euler's proof used the arithmetic of $\mathbb{Z}[\omega]$, where $\omega = (-1 + \sqrt{-3})/2$ is a primitive cube root of unity.

We'll pursue a totally different idea. The idea is that the curve $X^3 + Y^3 = Z^3$ is a cubic plane curve (again, given the understanding of (de)homogenization), and has a rational inflection point; this allows us to put it in *Weierstrass form*. More generally, we can consider for any nonzero α the equation

$$X^3 + Y^3 = \alpha Z^3.$$

This equation has rational point $[1 : -1 : 0]$. Let's move this a little bit, specifically to $[0 : 1 : 0]$: consider the "change of coordinates" given by

$$(X_1, Y_1, Z_1) = (12\alpha Z, 36\alpha(X - Y), X + Y). \tag{1.2.8}$$

Exercise 1.2.9. Figure out how to invert this change of coordinates: express (X, Y, Z) in terms of (X_1, Y_1, Z_1).

We will see below why this change of coordinates is a good one; we will explain later where it comes from. Given this "inverse change of coordinates", we note that the curve above transforms to the one of the form

$$Y_1^2 Z_1 = X_1^3 - 432\alpha^2 Z_1^3,$$

which after dehomogenezing (dividing throughout by Z_1^3) is the curve

$$y^2 = x^3 - 432\alpha^2.$$

This is nothing but our example from 1.1.6, with $c = -432\alpha^2$, and we are back to finding rational points on this cubic curve. The Fermat case corresponds to $\alpha = 1$, and hence $c = -432$. (This might explain whey $c = -432$ is special in 1.1.6.) This is again what we call an *elliptic curve*. We will show later using techniques for elliptic curves that the only points on here are the points $(12, \pm 36)$.

Exercise 1.2.10. Which points on $X^3 + Y^3 = Z^3$ do the points $(12, \pm 36)$ on $y^2 = x^3 - 432$ correspond to?

This finishes the proof of Fermat's last theorem modulo the claim made about the curve $y^2 = x^3 - 432$, which we will prove later. I also still need to explain where this change of coordinates comes from; that will be done sooner. (Hopefully this motivates you to keep coming to this course!) We'll deal with the case $n = 4$ next time.

1.3 26/06/17 - Introduction III: Fermat for Exponent 4 and Infinite Descent; Review of Projective Geometry I: Basics

Let's first finish up our discussion of the Fermat problem for the exponent $n = 4$.

1.3.A Fermat for Exponent 4, Proof 1

Let's try to solve the equation

$$X^4 + Y^4 = Z^4.$$

This is the only case for which Fermat is believed to have a proof, and Fermat's proof uses a very important strategy, known now as "Fermat's method of infinite descent". Let us now review (in modern terminology) Fermat's proof. In fact, we will show something stronger.

Theorem 1.3.1 (Fermat). Let $X, Y, W \in \mathbb{Z}_{\geq 0}$ be such that $X^4 + Y^4 = W^2$. Then $XYW = 0$.

Proof. Suppose to the contrary there is a solution (X, Y, W) with $X, Y, W > 0$, and pick such a solution with the smallest possible W. This implies in particular that X^2, Y^2, W are pairwise coprime (why?) and hence using 1.2.3 that (possibly up to permuting X and Y, which we do), there are unique coprime $m, n \in \mathbb{Z}$ of different parity with $m > n > 0$ such that

$$(X^2, Y^2, W) = (m^2 - n^2, 2mn, m^2 + n^2);$$

in particular, X is odd. Now $X^2 = m^2 - n^2$, so that $X^2 + n^2 = m^2$. Again X, n, m are pairwise coprime positive integers with X odd, so 1.2.3 applied once again gives us unique coprime $p, q \in \mathbb{Z}$ of different parity with $p > q > 0$ and such that

$$(X, n, m) = (p^2 - q^2, 2pq, p^2 + q^2).$$

In particular, n is even and hence m is odd. Finally, $m = p^2 + q^2$ implies that m, p, and q are pairwise coprime. Therefore

$$Y^2 = 2mn = 4mpq$$

implies that p, q, m are all squares. Write $(p, q, m) = (X_1^2, Y_1^2, W_1^2)$ for some integers $X_1, Y_1, W_1 > 0$. Then

$$X_1^4 + Y_1^4 = p^2 + q^2 = m = W_1^2,$$

so that (X_1, Y_1, W_1) is also a solution. Finally,

$$W_1 \leq W_1^2 = m \leq m^2 < m^2 + n^2 = W$$

tells us that this is a "smaller" solution, giving us the required contradiction. ∎

Can you see why this is called "infinite descent"?

The machinery of elliptic curves (specifically descent on them) is a vast modern generalization of this technique due to Fermat; this is where the name "descent" in the theory of elliptic curves comes from. Now Fermat's equation for $n = 4$ does not lend itself directly to the technology of elliptic cuvres, because it not one: $X^4 + Y^4 = Z^4$ defines a *quartic* curve in the plane which is a *canonical curve of genus two* (whatever that means). However, there is a clever way to massage it to give points on an elliptic curve; it is this curve that we do "descent" on. Since it's fun to figure out how to do this massaging, I will leave this as an exercise (2.1.12). We will figure out how to do descent on that curve–and hence prove Fermat's result in a second way–in a more general context much later.

In closing, let me remark that the case $n \geq 5$ of Fermat's theorem is genuinely very difficult, and it is beyond my abilities to go into the details here (see the history on the Wikipedia page, for instance). However, Andrew Wiles's proof from 1994 of it still nontrivially uses the theory of elliptic curves, following an idea going back to Frey from 1984. If we have time and interest, I am happy to explain the outline of the proof towards the end of the course; let me know if this is something you would like to see.

To fruitfully study plane curves, we need some basic understanding of the projective geometry of the plane. Let's take a detour to rapidly review what we need; a much fuller account can be found at [5].

1.3.B Basics of Projective Geometry

Let's start with an example.

Example 1.3.2. Fix a field K, and for $s \in K$ consider the intersection of the two lines

$$sy = x \text{ and } x = 1.$$

When $s \neq 0$, there is a unique intersection point, namely $(1, 1/s)$, but when $s = 0$, there is no intersection at all, and the lines are parallel. What has happened is that as $s \to 0$, the intersection point has "escaped to infinity".

Such behavior can be extremely troublesome while doing geometry, and it is helpful to develop basic vocabulary to deal with "points at infinity". The modern approach to doing this involves working with *homogeneous coordinates on projective space.*

Definition 1.3.3. Let K be a field and $n \in \mathbb{Z}_{\geq 0}$. We define the *set of K-points of projective n-space over K* to be

$$\mathbb{P}^n(K) := \{[X_0 : X_1 : \cdots : X_n] : X_0, \ldots, X_n \in K \text{ not all zero}\} / \sim$$

where $\sim$ means that we identify points with the same coordinates up to scaling by a nonzero number, i.e., for all $X_0, \ldots, X_n \in K$, not all zero, and $\lambda \in K^\times$ we have

$$[X_0 : \cdots : X_n] \sim [\lambda X_0 : \cdots : \lambda X_n].$$

The coordinates $X_0, \ldots, X_n$ are called *homogeneous coordinates*, and are well-defined only up to simultaneous scaling.

Unimportant Remark 1.3.4. For those who have seen linear algebra, the set $\mathbb{P}^n(K)$ can be identified with the set of one-dimensional K-linear subspaces of K^{n+1} (how?). For those who have seen a little topology, when $K = \mathbb{R}$ or $K = \mathbb{C}$ (or more generally any topological field such as $K = \mathbb{Q}_p$), then $\mathbb{P}^n(K)$ can be given a natural topology, which for $K = \mathbb{R}, \mathbb{C}, \mathbb{Q}_p$ even makes it a(n) (analytic) manifold.

Example 1.3.5.

(a) When $n = 0$, the space $\mathbb{P}^0(K)$ is just one point.

(b) When $n = 1$, the points of $\mathbb{P}^1(K)$ look like $[S : T]$, where $S, T \in K$ are not both zero, and we can simultaneously scale S and T by the same nonzero element λ of K. In particular, when $S \neq 0$, then scaling by $\lambda = S^{-1}$ gives the unique representative $[1 : t]$ where $t := T/S \in K$. In this way, $\mathbb{A}^1(K)$ sits naturally in $\mathbb{P}^1(K)$. The only point not contained in it is the point $[0 : 1]$ corresponding to $S = 0$; this can be thought of as corresponding to "$t = 1/0 = \infty$" and will often be called the *point at infinity* and denoted ∞. Therefore, $\mathbb{P}^1(K) = \mathbb{A}^1(K) \amalg \{\infty\}$ is the "one-point compactification" of $\mathbb{A}^1(K)$. (Warning: This is not true for higher n.) This allows us to handle infinity easily–no actual division by zero needed.

(c) When $n = 2$, we are looking at triples $[X : Y : Z]$ of $X, Y, Z \in K$, not all zero, up to simultaneous scaling. Again, if $Z \neq 0$, then scaling by $\lambda = Z^{-1}$ gives us the unique representative $[x : y : 1]$ where $(x, y) = (X/Z, Y/Z)$, and again we get an $\mathbb{A}^2(K) \subset \mathbb{P}^2(K)$. The complement corresponds to $Z = 0$, which is the set of points $[X : Y : 0]$, in natural bijection with $\mathbb{P}^1(K)$ itself. This locus is called the *line at infinity*; the point $[X : Y : 0]$ should be thought of as the point infinitely far away in the direction of the vector $[X : Y]$. In this way, $\mathbb{P}^2(K) = \mathbb{A}^2(K) \amalg \mathbb{P}^1(K)$. Said yet another way, we have

$$\big\{[X : Y : Z] \in \mathbb{P}^2(K) : Z \neq 0\big\} = \big\{[X/Z : Y/Z : 1] \in \mathbb{P}^2(K)\big\} = \{(x, y) \in \mathbb{A}^2(K)\},$$

where again $(x, y) = (X/Z, Y/Z)$.

Unimportant Remark 1.3.6.

(a) Continuing in the above fashion, we can describe a decomposition of $\mathbb{P}^n(K)$ of the form

$$\mathbb{P}^n(K) = \mathbb{A}^n(K) \amalg \mathbb{P}^{n-1}(K) = \mathbb{A}^n(K) \amalg \mathbb{A}^{n-1}(K) \amalg \cdots \amalg \mathbb{A}^1(K) \amalg \mathbb{A}^0(K).$$

(b) We will probably not seriously need $\mathbb{P}^n(K)$ for $n \geq 3$ in this course, but it may be helpful to note that much of what follows will apply with minor modifications to the general case.

(c) If you have seen barycentric coordinates before, then homogeneous coordinates on projective space $\mathbb{P}^2$ can be thought of as barycentric coordinates with the coordinate (or pure weight) points at $[1:0:0], [0:1:0]$, and $[0:0:1]$.

All faults of affine geometry are rectified in projective geometry, and this boils down to basic linear algebra/matrix theory; as an example, let's see how to fix 1.3.2.

Example 1.3.7. The "projectivization" of the lines of 1.3.2 are the lines

$$sY = X \text{ and } X = Z$$

in $\mathbb{P}^2(K)$. These two *always* have a unique intersection point, namely $[s:1:s]$, in $\mathbb{P}^2(K)$. Note that for $s = 0$, this is the point $[0:1:0]$ at infinity along the vertical direction, as expected.

Example 1.3.8. The "projectivization" of the parametrization in §1.2.A is the map $\mathbb{P}^1 \to \mathbb{P}^2$ given by

$$[S:T] \mapsto [S^2 - T^2 : 2ST : S^2 + T^2],$$

with image $X^2 + Y^2 = Z^2$. (Why is this map well-defined? For this to be true, we have to check that $S^2 - T^2, 2ST, S^2 + T^2$ are never simultaneously zero. This requires our base field to have characteristic other than 2.) When $\operatorname{ch} K \neq 2$, this gives us a *genuine* bijection (even isomorphism) between $\mathbb{P}^1(K)$ and $C(K)$, where $C = \mathbb{V}(X^2 + Y^2 - Z^2) \subset \mathbb{P}^2_K$. This way of looking at it solves various mysteries: for instance, the unique point mapping to $P = (-1, 0) = [-1:0:1]$ is precisely $[S:T] = [0:1]$, also known as "$t = \infty$". Similarly, over $K = \mathbb{C}$ the points $[1:\pm i]$ map to the points $[1:\pm i:0]$ in $\mathbb{P}^2(\mathbb{C})$, known as the "circular points at infinity"; these are not visible over $\mathbb{Q}$ or $\mathbb{R}$.

Let us study solutions of polynomial equations on projective spaces; for simplicity, we will stick to $n = 3$ and use homogeneous coordinates X, Y, Z. On projective space, the vanishing locus of an arbitrary $F \in k[X, Y, Z]$ does not make sense in general: if $F = X^2 - Y$, for instance, then $(1, 1, 0)$ lies on F but $(2, 2, 0)$ does not. However, the one case in which this problem does *not* arise is when F is *homogeneous*, i.e., when all monomials appearing in F have the same degree, say $d \in \mathbb{Z}_{\geq 0}$ (c.f. 2.1.11). In this case we have

$$F(\lambda X, \lambda Y, \lambda Z) = \lambda^d F(X, Y, Z)$$

for $\lambda \in K^\times$, so it makes sense to speak of the *curve* $C = \mathbb{V}(F)$ cut out by F in projective space. In other words, C consists of K-points

$$C(K) := \{[X:Y:Z] \in \mathbb{P}^2(K) : F(X, Y, Z) = 0\}.$$

Example 1.3.9. If $F = Y^2Z - X^3 - Z^3$ and $C = \mathbb{V}(F) \subset \mathbb{P}^2_K$, then $\mathscr{O} := [0:1:0] \in C(K)$ but $[1:0:0] \notin C(K)$.

Clearly, replacing F by a nonzero scalar multiple μF for $\mu \in K^\times$ is not going to change the vanishing locus, and in this way we get an identification between curves and nonzero homogeneous polynomials in $K[X, Y, Z]$ up to scaling. We take this to be our definition.

Definition 1.3.10. A *projective plane curve* C over a field K is an equivalence class of nonzero homogeneous polynomials $F \in K[X, Y, Z]$ under the equivalence relation of scaling: F and μF for $\mu \in K^\times$ define the same curve $C = \mathbb{V}(F) = \mathbb{V}(\mu F)$. A representative F of this equivalence class is said to be a *defining equation* of the curve. The degree of any defining equation is then called the *degree* of the curve.

We emphasize that this is an *algebraic* definition; the curve is determined by its equation and not its set of K-points, of which there may not be any (e.g., $C = \mathbb{V}(X^2 + Y^2 + Z^2)$ over $K = \mathbb{R}$). A curve

of degree one (resp. two, three, four, etc.) is called a *line* (resp. *conic*, *cubic*, *quartic*, etc.). However, it is still natural to think of the curve C–to a first approximation–in terms of its K-points $C(K)$.

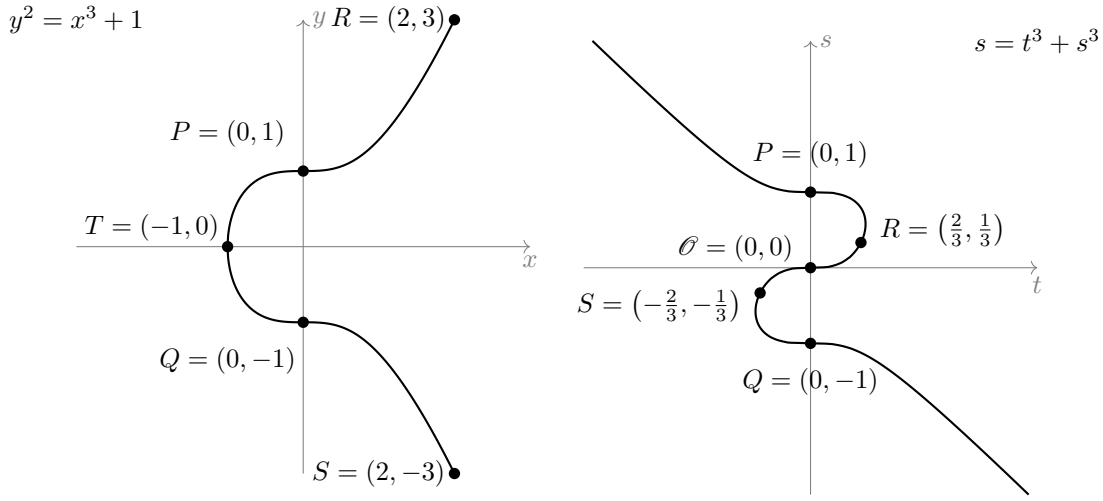
How do we visualize a projective curve C? Let us now briefly discuss something we have seen in practice already. Given a projective curve $C = \mathbb{V}(F) \subset \mathbb{P}^2_K$ over some field K and of degree $d \in \mathbb{Z}_{\geq 1}$, we can dehomogenize it with respect to any of the variables X, Y, or Z to get an affine curve, where dehomogenization, say, with respect to the variable Z amounts to either dividing throughout by Z^d and using coordinates $x := X/Z$ and $y := Y/Z$, or equivalently looking at points in the affine patch $Z \neq 0$, where we can normalize to $Z = 1$. Algebraically, this amounts to replacing $F \in K[X, Y, Z]$ by $f(x, y) := F(x, y, 1) \in K[x, y]$. In other words,

$$\begin{aligned} C(K) \cap \{Z \neq 0\} &= \{[X : Y : Z] \in \mathbb{P}^2(K) : F(X, Y, Z) = 0 \text{ and } Z \neq 0\} \\ &= \{(x, y) \in \mathbb{A}^2(K) : F(x, y, 1) = 0\} \\ &= \{(x, y) \in \mathbb{A}^2(K) : f(x, y) = 0\}. \end{aligned}$$

There is nothing special about Z here, and we can equally well dehomogenize with respect to say Y, by taking coordinates $t := X/Y$ and $s := Z/Y$ instead. These different "local patches" of the curve can then be glued together to get a fuller picture of the projective curve.

Example 1.3.11. Consider the projective curve $C := \mathbb{V}(Y^2Z - X^3 - Z^3) \subset \mathbb{P}^2_K$ (for any field K). The affine patch $Z \neq 0$ looks like the familiar curve $y^2 = x^3 + 1$, and the affine patch $Y \neq 0$ looks like $s = t^3 + s^3$. The first patch has all points on C except for $\mathscr{O} = [0 : 1 : 0]$; the second patch has all points except for $[-\omega^j : 0 : 1]$ for $j = 0, 1, 2$, where $\omega = \zeta_3$ is a primitive cube root of unity. (Some of these point may not exist in $C(K)$. If K has characteristic three, there is only one point $[-1 : 0 : 1]$.) These two patches are related by the coordinate change

$$(t, s) = \left(\frac{x}{y}, \frac{1}{y}\right) \text{ and } (x, y) = \left(\frac{t}{s}, \frac{1}{s}\right).$$



(a) The affine chart $Z \neq 0$.

(b) The affine chart $Y \neq 0$.

Figure 1.3.12: Two affine patches of the curve $Y^2Z = X^3 + Z^3$ with a few marked points. Note that the point $\mathscr{O}$ is not visible in the first patch, and the point T is not visible in the second patch, but together they cover everything. Note also that the line at infinity in the first patch becomes the t-axis $s = 0$ in the second patch, and that this line has a "triple" intersection with the curve at $\mathscr{O}$, i.e., is an inflectional tangent to C at $\mathscr{O}$.

I would recommend playing with Desmos3D and plotting $Y^2Z = X^3 + Z^3$ and seeing how it intersects the loci $Z = 1$, $Y = 1$, and $X^2 + Y^2 + Z^2 = 1$ to see how these pictures are related to each other.

Going in the other direction, given an affine curve $C \subset \mathbb{A}^2_K$ given by the vanishing of say $f(x, y) \in K[x, y]$, we can look at its *projective closure*. This corresponds to replacing f by its *homogenization* $F \in K[X, Y, Z]$ obtained by replacing x by X/Z and y by Y/Z and clearing denominators, or

(equivalently) by replacing x by X and y by Y, and inserting suitable powers of Z in the monomials to get a homogeneous polynomial of lowest possible degree. This projective closure is usually denoted by $\overline{C} \subset \mathbb{P}^2_K$. We can recover C from $\overline{C}$ by dehomogenizing with respect to Z once again.

Example 1.3.13. The line $2x - 5y + 3 = 0$ in $\mathbb{A}^2_K$ has projective closure $2X - 5Y + 3Z = 0$ in $\mathbb{P}^2_K$. To obtain the points at infinity on this projective closure, we set $Z = 0$ to get the unique point $[5 : 2 : 0]$ in the direction of this line. Make sure this makes sense to you!

Example 1.3.14. The rectangular hyperbola $x^2 - y^2 = 1$ in $\mathbb{A}^2_K$ has projective closure $X^2 - Y^2 = Z^2$ in $\mathbb{P}^2_K$. To obtain the points at infinity of this projective closure, we set $Z = 0$ to get the two points $[1 : \pm 1 : 0]$. Of course, this is as expected: these correspond to the asymptotes of the hyperbola.

Exercise 1.3.15. Are the operations of homogenization and dehomogenization inverses?

1.4 25/06/18: Review of Projective Geometry II: Projective Changes of Coordinates

Perhaps I didn't do a great job of explaining projective space last time, so let me try again.

1.4.A More on Curves and Homogenization

Let us review some things from last time. Given a field K, we met the *affine plane* $\mathbb{A}^2(K)$ over K, which is just the good old plane: it consists of all points (x, y) where $x, y \in K$:

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

Note that for any field extension L/K, we have a natural inclusion $\mathbb{A}^2(K) \subset \mathbb{A}^2(L)$.

Given a nonzero $f \in K[x, y]$, we can look at its *vanishing locus* in $\mathbb{A}^2(K)$, which is just

$$\{(x, y) \in \mathbb{A}^2(K) : f(x, y) = 0\}.$$

This is what we want to call a curve. Note that for any $\lambda \in K^\times$, the polynomials f and λf have the same vanishing loci; they define the same *curve*. This motivates the algebraic

Definition 1.4.1.

- An *affine curve* is an equivalence class of nonzero $f \in K[x, y]$ under the equivalence relation $f \sim \lambda f$ for nonzero $f \in K[x, y]$ and $\lambda \in K^\times$.

  The curve defined by a nonzero f will be denoted $C := \mathbb{V}(f)$, so that $\mathbb{V}(f) = \mathbb{V}(\lambda f)$ for all $\lambda \in K^\times$.
- Such an f as above is called *a defining equation* for the curve C, and the *degree of the curve* C is defined to be the degree of any defining equation for C.

A(n) (affine) curve of degree one (resp. two, three, four, etc.) is called a(n) (affine) *line* (resp. *conic*, *cubic*, *quartic*, etc.).

Given such an f and the associated curve $C = \mathbb{V}(f) \subset \mathbb{A}^2_K$, we can then look at its K-points

$$C(K) := \{(x, y) \in \mathbb{A}^2(K) : f(x, y) = 0\}.$$

The reason for keeping this notation separate is because given a polynomial $f \in K[x, y]$, we would like to study its solutions over any field extension L/K. For instance, K could be $\mathbb{Q}$, and L could be $\mathbb{R}$ or $\mathbb{C}$ or $\mathbb{Q}_p$ or some number field like $\mathbb{Q}[i]$ or $\mathbb{Q}[\sqrt{5}]$. The above notation easily lets us do that, so we get a natural inclusion $\mathbb{A}^2(K) \subset \mathbb{A}^2(L)$ and $C(K) \subset C(L)$.

Example 1.4.2. Let $K = \mathbb{Q}$, let $f(x, y) = x^2 + y^2 - 1 \in \mathbb{Q}[x, y]$, and let $C := \mathbb{V}(f)$. Then $C(\mathbb{Q}) \subset C(\mathbb{R}) \subset C(\mathbb{C})$. We have that $(1, 0) \in C(\mathbb{Q})$, that $(1/\sqrt{2}, 1/\sqrt{2}) \in C(\mathbb{R}) \smallsetminus C(\mathbb{Q})$, and $(i, \sqrt{2}) \in C(\mathbb{C}) \smallsetminus C(\mathbb{R})$.

This algebraic definition also allows us to work with curves that have no points over the base field.

Example 1.4.3. Let $K = \mathbb{Q}$, let $f(x, y) := x^2 + y^2 + 1 \in \mathbb{Q}[x, y]$, and let $C := \mathbb{V}(f)$. Then $C(\mathbb{Q}) = C(\mathbb{R}) = \emptyset$ but $C(\mathbb{C}) \neq \emptyset$, because for instance $(0, i) \in C(\mathbb{C})$. In particular, C is not "empty"; it's just that $\mathbb{R}$ is not a good enough field to see any points on it.

Exercise 1.4.4. Let K be an algebraically closed field. Then for any nonzero $f \in K[x, y]$ of positive degree, if we set $C := \mathbb{V}(f)$, then $C(K) \subset \mathbb{A}^2(K)$ is nonempty. (Note that of course this is not true if K is not algebraically closed, as we have seen above in 1.4.3.)

The abstract curve $C = \mathbb{V}(f) \subset \mathbb{A}^2_K$ arising from this algebraic definition encodes this data of points in every field extension L/K.

Unimportant Remark 1.4.5. The objects $\mathbb{A}^2_K$ and $\mathbb{A}^2(K)$ are slightly different; the latter is the set of K-points of the former (and is probably more properly written $\mathbb{A}^2_K(K)$). The first is a "formal object" that records the data of all $\mathbb{A}^2(L)$ for all extensions L/K. The same goes for the curve C and its L-points $C(L)$ for any L/K. The subscript K in the formal object reminds us what field we are doing the algebra; for instance, if $K = \mathbb{F}_3$ and $C = \mathbb{V}((1 \mod 3)y + (2 \mod 3)x + (1 \mod 3)) \subset \mathbb{A}^2_{\mathbb{F}_3}$, then the subscript reminds us that it makes no sense to talk about $C(\mathbb{Q})$, although it makes perfect sense to talk about $C(\mathbb{F}_3)$ or more generally $C(L)$ for any $L/\mathbb{F}_3$. (For the cognoscenti: $\mathbb{A}^2_K := \operatorname{Spec} K[x,y]$ and $C := \operatorname{Spec} K[x,y]/(f)$.)

Remark 1.4.6. One strange consequence of the above definition is that sometimes you can have two curves $C, C' \subset \mathbb{A}^2_K$ such that for all fields L/K we have $C(L) = C'(L)$ even though $C \neq C'$. For instance, C and C' might even have different degrees, e.g., if $C = \mathbb{V}(x)$ and $C' = \mathbb{V}(x^2)$. Why do you think this is a good thing? Is it?

More or less the same things as above can be done in the projective plane. As above,

$$\mathbb{P}^2(K) = \{[X : Y : Z] : X, Y, Z \in K \text{ not all zero}\}/\sim$$

where the $\sim$ means that we identify points $[X : Y : Z]$ and $[\lambda X : \lambda Y : \lambda Z]$ for all $\lambda \in K^\times$. As before, for each L/K, we have $\mathbb{P}^2(K) \subset \mathbb{P}^2(L)$. Finally, for any K, the affine space $\mathbb{A}^2(K)$ sits inside $\mathbb{P}^2(K)$ by sending (x, y) to $[x : y : 1]$, with the complement being the "line at infinity" given by $Z = 0$.

Given a nonzero *homogeneous* $F \in K[X, Y, Z]$, we can look at its *vanishing locus* in $\mathbb{P}^2(K)$, which is just

$$\{[X : Y : Z] \in \mathbb{P}^2(K) : F(X, Y, Z) = 0\}.$$

As before, we have

Definition 1.4.7.

- A *projective curve* is an equivalence class of nonzero homogeneous $F \in K[X, Y, Z]$ under the equivalence relation $F \sim \lambda F$ for nonzero homogeneous $F \in K[X, Y, Z]$ and $\lambda \in K^\times$.

The curve defined by a nonzero F will be denoted $D := \mathbb{V}(F)$, so that $\mathbb{V}(F) = \mathbb{V}(\lambda F)$ for all $\lambda \in K^\times$.

- Such an F as above is called *a defining equation* for the curve D, and the *degree of the curve* D is defined to be the degree of any defining equation for D.[a]

[a] For the cognoscenti: D stands for "divisor".

As before, we have projective lines, conics, cubics, quartics, etc., and as before for any $D = \mathbb{V}(F) \subset \mathbb{P}^2_K$ it makes sense to talk about $D(L)$ for any L/K, with $D(K) \subset D(L)$.

Example 1.4.9. A projective line has the form $D = \mathbb{V}(aX + bY + cZ)$ for $a, b, c \in K$ not all zero, and where we may simultaneously scale a, b, and c. In particular, the set of all lines in $\mathbb{P}^2_K$ it itself a $\mathbb{P}^2(K)$, called the *dual projective plane*. (This duality has some lots of interesting consequences.) In general, this simultaneous scaling business feels awfully similar to what we have done above in the definition of projective spaces itself. This is no coincidence. The "space" of all curves of a given degree $d \in \mathbb{Z}_{\geq 0}$ is itself a projective space $\mathbb{P}^N(K)$ for some N (which?).

Remark 1.4.10. Doing the same procedure as above for arbitrary n gives us a *hypersurface* $D = \mathbb{V}(F) \subset \mathbb{P}^n_K$ associated to any nonzero homogenous $F \in K[X_0, \ldots, X_n]$. When $n = 1$, for any field extension L/K, the set $D(L)$ is then finite and counts the number of roots–including those at infinity–of a (homogenized) single-variable equation in L (do you see why?). For example, if $K = \mathbb{Q}$ and $F = S^2 - 5ST + 5T^2$, then $D(\mathbb{Q}) = D(\mathbb{R}) = \emptyset$ but $D(\mathbb{C})$ consists of two points (c.f. 1.1.1(b)(i)).

Finally, as before we have the notions of homogenization and dehomogenization. We have that the locus $Z \neq 0$ (i.e., the complement of the line $\mathbb{V}(Z)$ at infinity) in $\mathbb{P}^2_K$ is naturally an $\mathbb{A}^2_K$:

$$\mathbb{A}^2_K = \{Z \neq 0\} \subset \mathbb{P}^2_K.$$

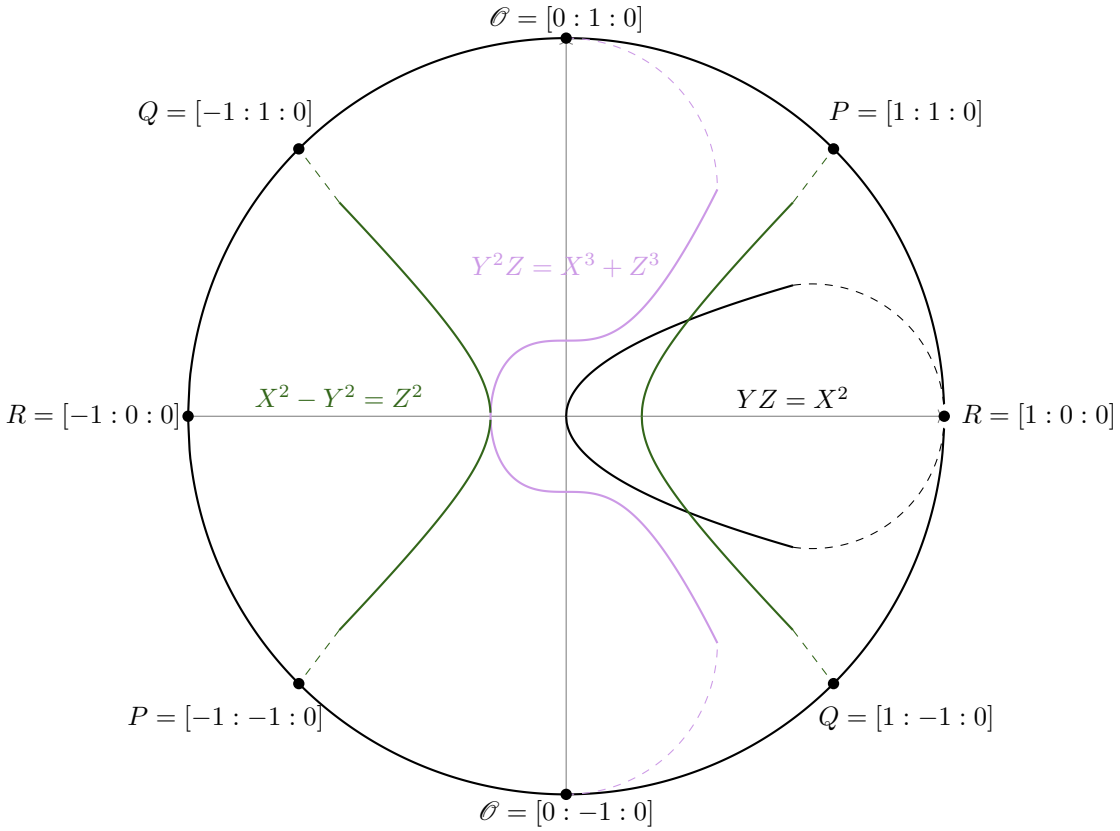


Figure 1.4.8: Some curves in the projective plane and their points at infinity.

For a projective curve $D = \mathbb{V}(F) \subset \mathbb{P}^2_K$, the intersection

$$C := D^\circ := D \cap \mathbb{A}^2_K = \mathbb{V}(f) \subset \mathbb{A}^2_K,$$

where $f(x,y) := F(x,y,1) \in K[x,y]$. This is the dehomogenization with respect to Z. Conversely, given an $f(x,y) \in K[x,y]$ of degree $d \in \mathbb{Z}_{\geq 1}$ and corresponding affine curve $C = \mathbb{V}(f)$, its *projective completion* is the projective curve

$$D := \overline{C} = \mathbb{V}(F) \subset \mathbb{P}^2_K,$$

where $F := Z^d f(X/Z, Y/Z) \in K[X,Y,Z]$. These two operations are basically inverses, but this is not literally true. For instance, we certainly have $(\overline{C})^\circ = C$ and $D \subset \overline{D^\circ}$, but equality need not hold in the latter–we may lose components corresponding to the line at infinity.

Exercise 1.4.11. What on earth does this last line mean? Are the operations of homogenization and dehomogenization inverses to each other?

However, in projective space the coordinates X, Y, Z are completely symmetrical. This flexibility of homogenization with respect to one variable, applying a linear transformation of coordinates, and then dehomogenization with respect to another yields lots of birational transformations of curves.

Example 1.4.12. Consider the affine curve $x^3 + y^3 = \alpha$ from 1.2.B. The transformation of coordinates given by

$$(x_1, y_1) := \left(\frac{12\alpha}{x+y}, \frac{36\alpha(x-y)}{x+y} \right)$$

is a birational transformation onto the curve $y_1^2 = x_1^3 - 432\alpha^2$. Of course, this arises from homogenization first to $X^3 + Y^3 = \alpha Z^3$, applying the transformation 1.2.8, and then dehomogenization with respect to Z_1. (Make sure to convince yourself that this is a birational transformation!)

Here we used a linear change of coordinates in the middle. Let's study these a little more now.

Unimportant Remark 1.4.13. Sometimes there are equations "over $\mathbb{Z}$", which can then be reduced mod p for various primes p. For instance, we could look at the curve C given by $Y^2Z = X^3 + Z^3$, or equivalently $y^2 = x^3 + 1$, and count points over $\mathbb{F}_p$ for various p. One would like to say that the original curve is defined over something like $\mathbb{Z}$ (i.e., $C \subset \mathbb{P}^2_{\mathbb{Z}}$) and these $\mathbb{F}_p$-counts are the usual counts $C(\mathbb{F}_p)$ for the "extensions $\mathbb{F}_p/\mathbb{Z}$". This is subtle, but can be made to work using the modern language of schemes. Indeed, figuring out the right definition of $\mathbb{P}^2_{\mathbb{Z}}$ (it may not be what you think!) was an impetus for the development of modern scheme theory.

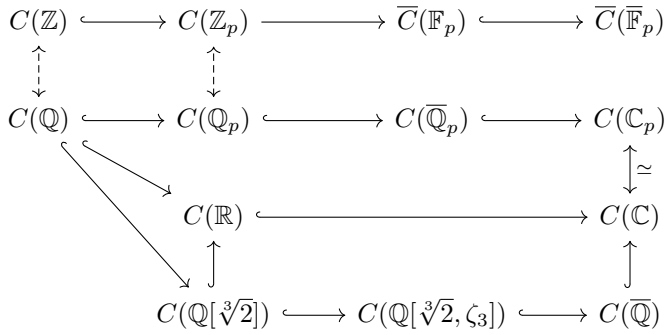


Figure 1.4.14: A figure explaining the relationship between the K-points of a projective curve C defined over $\mathbb{Q}$ (and lifted to $\mathbb{Z}$) for various K. Some of the terminology and relationships will be explained later.

1.4.B Projective Changes of Coordinates

One advantage of working with projective spaces is the presence of lots of extra symmetries not visible in the affine plane. These come from linear changes of coordinates and look like

$$(X_1, Y_1, Z_1) = (aX + bY + cZ, dX + eY + fZ, gX + hY + iZ) \tag{1.4.15}$$

for $a, b, c, d, e, f, g, h, i \in K$. Clearly, not all choices of these "transformation coordinates" will work: you certainly can't take all of them to be zero. But, in fact the situation is a little stricter: you cannot take $(a, b, c, d, e, f, g, h, i) = (1, 0, 0, 1, 0, 0, 1, 0, 0)$ either. Similarly, but perhaps more subtly, you can't take $(a, b, c, d, e, f, g, h, i) = (2, 3, 5, -1, 2, -2, 0, 7, -1)$ either (why?). To see exactly what this constraint is, it is helpful to write this change of coordinates in the form

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{1.4.16}$$

The idea is that if this transformation is to work, then there better not be a triple of numbers X, Y, Z, not all zero, such that $X_1 = Y_1 = Z_1 = 0$. This amounts to saying that the matrix A as above "representing this transformation" has no kernel, or equivalently (by a little linear algebra) has nonzero determinant. Conversely, if this matrix has nonzero determinant, then we get a way to invert this transformation by applying A^{-1} on the left. This is therefore the correct condition for invertibility: $\det A \neq 0$.

This matrix representation is also very helpful when we compose changes of coordinates: this just amounts to multiplying the corresponding change-of-coordinate matrices. This would lead us to believe that the group (not abelian!) of changes of coordinates in the group of invertible matrices, $\mathrm{GL}_3(K)$. But it's a little more subtle than that. For instance, the change of coordinates $(X_1, Y_1, Z_1) = (\lambda X, \lambda Y, \lambda Z)$ for any $\lambda \in K^\times$ isn't *really* a change of coordinates: it's just the identity operation of $\mathbb{P}^2_K$. More generally, two matrices A and λA for $A \in \mathrm{GL}_3(K)$ and $\lambda \in K^\times$ represent the same change of coordinates. This tells us that the group of automorphisms is actually the quotient of $\mathrm{GL}_3(K)$ by its normal subgroup $K^\times$, which is called the *projective general linear group*

$$\mathrm{PGL}_3(K) := \mathrm{GL}_3(K)/K^\times.$$

Therefore, $\mathrm{PGL}_3(K)$ acts on $\mathbb{P}^2_K$ by linear changes of coordinates. (Similarly, $\mathrm{PGL}_{n+1}(K)$ acts on $\mathbb{P}^n_K$ for any $n \geq 1$. The action of $\mathrm{PGL}_2(K)$ on $\mathbb{P}^1_K$ is by *Möbius transformations*.)

Unimportant Remark 1.4.17. It can be shown (using heavy machinery that I will not get into here) that *all* (algebraic) automorphisms of $\mathbb{P}^2_K$ (over K) come from $\mathrm{PGL}_3(K)$: these are all the possible automorphisms you can have. Note that this is not true of $\mathbb{A}^2_K$: the map $(x, y) \mapsto (x, y - x^3)$ is a non-linear change of coordinates (i.e., automorphism) of $\mathbb{A}^2_K$.

Exercise 1.4.18. When does a change of coordinates preserve that line at infinity $Z = 0$? You will have to figure out what this means. For more along these lines, see 2.2.2.

The flexibility of these changes of coordinates, therefore, comes down to linear algebra. For instance, we can take any point to any other point by a linear change of coordinates. The same holds for any pair of points. The same is *not* true of any triple of points, since linear changes of coordinates preserve incidence relations: so you can't take a collinear triple of points to a non-collinear triple of points. But this is the only constraint: any collinear triple of points maps to any other, and any non-collinear triple maps to any other, by a linear change of coordinates. Finally, because projective changes of coordinates preserve incidence relations, this tells us that we can use them to map any line to any other. Similarly, any pair of lines goes to any other. Finally, not all triples of lines go to each other for the same reason as above: there is a difference between lines that "form a triangle" and coincident lines. Proving all of this amounts to a little linear algebra. You are invited to explore things of this sort in 2.1.13.

We have seen one example of change of coordinates in 1.2.B, and will see many examples of linear transformations throughout the course, so I don't feel terribly guilty giving too many examples now. Changes of coordinates are extremely useful, but lest you think those are all possible transformations, let me give you a warning:

Remark 1.4.19. Not every "change of coordinates" we will be interested in arises via a linear change of coordinates (on a projective closure). There are definitely other (birational) changes of coordinates we will be interested in. For a concrete example, see 2.1.12.

1.5 26/06/22 - Review of Projective Geometry III: Smoothness

The next thing about curves we will need is the notion of smoothness. Before we give a formal definition, let's consider some examples.

Example 1.5.1. Which of the following curves is "smooth"? (a) $3x+5y=1$ (b) $x^2+y^2=1$ (c) $x^2-y^2=1$ (d) $x^2-y^2=0$ (e) $y^2=x^3+1$ (f) $y^2=x^3+x$ (g) $y^2=x^3+x^2$ (h) $y^2=x^3$.

Hopefully, the following definition feels reasonably motivated:

Definition 1.5.2. Let K be a field, and $f(x,y)\in K[x,y]$ be nonzero. Let $C=\mathbb{V}(f)\subset\mathbb{A}^2_K$ be the affine curve defined by f.

(a) Given a field extension L/K and a point $P=(x_0,y_0)\in C(L)$, we say that P is a *singular point* of C iff it is simultaneously a solution of

$$f=\frac{\partial f}{\partial x}=\frac{\partial f}{\partial y}=0$$

in $\mathbb{A}^2(L)$, and a *smooth point* otherwise. When P is a smooth point, we define the *tangent line* T_PC to C at P to be the line

$$\mathrm{T}_PC:=\mathbb{V}\left(\left.\frac{\partial f}{\partial x}\right|_P(x-x_0)+\left.\frac{\partial f}{\partial y}\right|_P(y-y_0)\right)\subset\mathbb{A}^2_L.$$

(b) The curve C is said to be *smooth*[a] iff for every field L/K, every point of $C(L)$ is smooth.

[a] Also known as *geometrically regular*.

Make sure this agrees with your intuitive definition of the tangent line to a curve at a point!

Exercise 1.5.3. Let $\ell\subset\mathbb{A}^2_K$ be an affine line. Show that for any $P\in\ell(K)$, we have that $\mathrm{T}_P\ell=\ell$.

Exercise 1.5.4. Figure out the "singular locus" of each of the curves from 1.5.1. Does the characteristic of the base field matter? What happens for the curve $x^2+y^2=0$ over $K=\mathbb{R}$?

Unimportant Remark 1.5.5. Why do we need to insist on checking all extensions L/K? We don't need to; a version of Hilbert's Nullstellensatz tells us that it suffices to check one: any algebraically closed field Ω (e.g., $\Omega=\overline{K}$) containing K (e.g., if $K=\mathbb{Q}$, it suffices to look at $\Omega=\mathbb{C}$, i.e., $\mathbb{C}$-valued points). In other words, if every point of $C(\Omega)$ is a smooth point, then C is smooth. (This needs an argument, which I will not give here. Ask me if you're interested.) Here's an example to explain why we can't just check K-points: take $K=\mathbb{F}_3(t)$ and $f(x,y):=y^2-x^3+t\in K[x,y]$. The partial derivatives here are 0 and $2y$ respectively. We see that the "singular locus" is cut out by the equations $y=x^3-t=0$, which has no K-point, but is nonetheless nonempty: it contains the L-point $(t^{1/3},0)$ where $L:=\mathbb{F}_3(t^{1/3})$. Such examples are pathologies we wish to avoid by taking our definition above, which is consistent with that of modern algebraic geometry. This example will return later when we discuss singularities of plane cubics.

When dealing with projective curves, we could either work with affine patches, or give a global definition. If we take the former approach, we would then have to justify that the choice of affine patch doesn't matter. This can be done, but the global definition is probably easier:

Definition 1.5.6 (Smoothness)**.** Let K be a field, $F\in K[X,Y,Z]$ a nonzero homogeneous polynomial, and $D=\mathbb{V}(F)\subset\mathbb{P}^2_K$ the plane curve with defining equation F.

(a) Given a field extension L/K and a point $P\in D(L)$, we say that P is a *singular point* of D iff it is simultaneously a solution of

$$F=\frac{\partial F}{\partial X}=\frac{\partial F}{\partial Y}=\frac{\partial F}{\partial Z}=0,$$

in $\mathbb{P}^2(L)$, and a *smooth point* otherwise. When P is a smooth point, we define the *(projective) tangent line* $\mathbb{T}_P D$ *to* D *at* P to be the line

$$\mathbb{T}_P D := \mathbb{V}\left(\left.\frac{\partial F}{\partial X}\right|_P X + \left.\frac{\partial F}{\partial Y}\right|_P Y + \left.\frac{\partial F}{\partial Z}\right|_P Z\right) \subset \mathbb{P}^2_L.$$

Here the notation means: pick a representative $(X, Y, Z) \in L^3 \smallsetminus \{(0,0,0)\}$ for $P = [X : Y : Z]$, and evaluate the partials simultaneously on that representative, noting that the triple of partials only changes up to simultaneous scaling by a different choice of representative, from which it follows that the corresponding line is well-defined.

(b) The curve D is said to be *smooth*[a] iff for every field L/K, every point of $D(L)$ is smooth.

[a] Also known as *geometrically regular*.

At this point, we need to insert an exercise to make sure that these two definitions really are the same. In particular, the projective definition behaves well under taking affine patches. Because this is important enough, I have isolated this and added this to Exercise Sheet 2 (2.2.6(a)). The advantage of the projective definition is that it is highly symmetrical, and in particular stable under projective changes of coordinates (2.2.6(b)).

Unimportant Remark 1.5.7. If $F \in K[X, Y, Z]$ is homogeneous of degree $d \in \mathbb{Z}_{\geq 0}$, then *Euler's formula* tells us that

$$X\frac{\partial F}{\partial X} + Y\frac{\partial F}{\partial Y} + Z\frac{\partial F}{\partial Z} = d \cdot F.$$

(Prove this! See 2.1.11.) In particular, if $\operatorname{ch} K \nmid d$, then the locus given by the vanishing of partials is automatically contained in the vanishing locus of F. The example of 1.5.5 shows that this is not true if $\operatorname{ch} K \mid d$ in general.

Example 1.5.8. If we take $K = \mathbb{Q}$ and $F = Y^2Z - X^3$, then for any field extension $L/\mathbb{Q}$, there is a unique singular point of $D(L)$, namely $[0 : 0 : 1]$. The tangent line to D at the $\mathbb{Q}$-point $P = [1 : 1 : 1]$ is $\mathbb{V}(-3X + 2Y + Z) \subset \mathbb{P}^2_{\mathbb{Q}}$. In affine coordinates in the usual patch, this is saying that the curve $y^2 = x^3$ is singular precisely at $(0, 0)$ and has tangent line $-3x + 2y + 1 = 0$ at $(1, 1)$.

Next time, we will see how to apply these ideas in practice, and work through a few examples.

1.6 26/06/23 - Review of Projective Geometry IV: Wrapping up Smoothness, and Bézout's Theorem

Let me say a few more things about smoothness first, and let's go over a few examples. Then we'll talk about Bézout's Theorem. After this, starting next time, we'll start talking about cubic curves and group laws on them.

1.6.A Wrapping up Smoothness

Let's do a few examples.

Example 1.6.1. Let's study the example of the curve $C = \mathbb{V}(Y^2Z - X^3 - Z^3) \subset \mathbb{P}^2_K$ we saw in 1.3.12. For simplicity, we will assume that $\operatorname{ch} K \neq 2, 3$; we will figure out what happens in these characteristics later.

First, let's study what happens in the two affine patches.

(a) In the patch $Z \neq 0$ with coordinates $(x, y) = (X/Z, Y/Z)$, we have the curve $\mathbb{V}(y^2 - x^3 - 1)$. This equation has partials $-3x^2, 2y$. If $\operatorname{ch} K \neq 2, 3$, then any common solution has $x = y = 0$, but the point $(0, 0)$ doesn't lie on the curve. This proves that this patch is smooth. The tangent line at say $R = (2, 3)$ looks like $-2x + y + 1 = 0$.

(b) In the patch $Y = 0$ with coordinates $(t, s) = (X/Y, Z/Y)$, we have the curve $\mathbb{V}(s - t^3 - s^3)$. This equation has partials $-3t^2, 1 - 3s^2$. Again, there are no common solutions, and this shows that this patch is smooth as well. The tangent line at $R = (2/3, 1/3)$ then is $-2t + 1 + s = 0$.

Therefore, we have shown that the curve C is smooth, at least when $\operatorname{ch} K \neq 2, 3$.

We could do this "globally" instead of looking at affine patches. To study its singular points, for $F = Y^2Z - X^3 - Z^3$, we have partial derivatives

$$-3X^2, 2YZ, Y^2 + 3Z^2.$$

Again, since we are assuming $\operatorname{ch} K \neq 2, 3$, this forces $X = 0$, and either $Y = 0$ or $Z = 0$, and either forces the other by $Y^2 + 3Z^2 = 0$. Therefore, there are no solutions, i.e., the curve is smooth. Finally, the tangent line at $R = [2 : 3 : 1]$ is $\mathbb{T}_R(C) = \mathbb{V}(-2X + Y + Z)$.

Do you see how the "projective" and "affine computations" ((a) and (b)) relate to each other? You should now be in a good position to attempt Exercise 2.2.6.

Let's see what happens with the same curve in characteristic 3; I'll leave characteristic 2 to you below.

Example 1.6.2. Let K be a field with $\operatorname{ch} K = 3$, and $C = \mathbb{V}(Y^2Z - X^3 - Z^3) \subset \mathbb{P}^2_K$. As before, in the affine patches, we get:

(a) In the patch $Z \neq 0$, we get the singularity $(x, y) = (-1, 0)$.
(b) In the patch $Y \neq 0$, we see that there are no singularities, i.e., the patch $Y \neq 0$ is smooth.

Globally, we see that a singularity would have $Y = 0$, which forces it to be the point $[-1 : 0 : 1]$, which is indeed a singularity. Note that this is not in the affine patch $Y \neq 0$. In particular, this projective curve has a unique singular point, namely $[-1 : 0 : 1] \in C(K)$.

Exercise 1.6.3. Figure out what happens when $\operatorname{ch} K = 2$. How many singular points do we have?

As one more example, let's classify all smooth conics in the plane.

Proposition 1.6.4. Let K be a field, and $C \subset \mathbb{P}^2_K$ be a smooth conic with a K-rational point $P \in C(K)$. Then there is a projective change of coordinates taking C to $\mathbb{V}(Y^2 - XZ) \subset \mathbb{P}^2_K$. In particular, C is the (isomorphic) image of a quadratic map $\mathbb{P}^1_K \to \mathbb{P}^2_K$, i.e., can be parametrized via map of the form $[S : T] \mapsto [A(S, T) : B(S, T) : C(S, T)]$ for some homogeneous quadratic polynomials $A, B, C \in K[S, T]$ with no common roots (over an algebraic closure).

Remark 1.6.5. Note that the result of this example is true in any characteristic (including $\operatorname{ch} K = 2$!), and not true if we assume only that $C \subset \mathbb{P}^2_K$ is a smooth conic, since C might not have any K-points at all! A simple example is given by $\mathbb{V}(X^2 + Y^2 + Z^2) \subset \mathbb{P}^2_{\mathbb{R}}$, which is isomorphic to $\mathbb{V}(Y^2 - XZ)$ over $\mathbb{C}$ but not over $\mathbb{R}$.

Proof. The second statement follows from the first, since the conic $\mathbb{V}(Y^2 - XZ)$ can be parametrized as $[S : T] \mapsto [S^2 : ST : T^2]$. Therefore, we just have to figure out the right change of coordinates.

Here's the proof idea: take the point P and the tangent line L to P. Take some *other* line M through P. This M will intersect C in another K-point $Q \in C(K)$ by a baby version of Bézout's Theorem. Now take the tangent line N to C through Q. This is different than L (it doesn't contain P), and so intersects L in a third K-point R. Now take the projective change of coordinates sending (P, Q, R) to $([0 : 0 : 1], [1 : 0 : 0], [0 : 1 : 0])$, and note that this works, up to a minor rescaling of one coordinate.

Let's work this out explicitly in practice. We can certainly take a change of coordinates sending P to $[0 : 0 : 1]$ and the tangent line $\mathbb{T}_P C$ (which exists since we assumed C is smooth!) to $L = \mathbb{V}(X)$. In these coordinates, C looks like

$$aX^2 + bXY + cY^2 + dXZ + eYZ + fZ^2 = 0.$$

Since $P = [0 : 0 : 1]$ lies on C, we see that $f = 0$. Since $\mathbb{T}_P(C) = \mathbb{V}(X)$, we see that $d \neq 0$ and $e = 0$. Therefore, the equation really has the form

$$aX^2 + bXY + cY^2 + dXZ = 0. \tag{1.6.6}$$

Now take the other line M to be $M = \mathbb{V}(Y)$. To figure out its intersection with C, we set $Y = 0$ in 1.6.6 to get $X(aX + dZ) = 0$, so the other point of intersection Q is $Q = [d : 0 : -a]$. This fulfils our promise of finding a different point Q. Now you can check that $P \notin \mathbb{T}_Q C$ (do this!), and this allows us to do what we promised above: take R to be the intersection of L and $\mathbb{T}_Q(C)$, and now note that P, Q, R are non-collinear (check!), which allows us to send them to $[0 : 0 : 1], [1 : 0 : 0], [0 : 1 : 0]$. When we do this, we still get an equation of the form (1.6.6), but now since Q lies on it, we have $a = 0$. Finally, the fact that $\mathbb{T}_Q(C) = \mathbb{V}(Z)$ implies that $b = 0$ as well. This leaves us with an equation of the form $cY^2 + dXZ = 0$. Since C is smooth, neither c nor d can be zero, and hence up to rescaling Z we get an equation of the desired form. ∎

Exercise 1.6.7. Fill in the gaps in the above proof: show that $P \notin \mathbb{T}_Q C$, and that P, Q, and R are not collinear.

Note that 1.6.4 gives us another way to parametrize smooth conics with a rational point in the plane: figure out the change of coordinates explicitly, and then see how the parametrization $[S^2 : ST : T^2]$ changes under it.

Exercise 1.6.8. Apply the above technique to give a second solution to Exercise 2.1.6(a).

That's all we will say about smoothness in general for now, but it's an idea we'll keep revisiting.

1.6.B Bézout's Theorem

Let's first ask a very natural question: given a field and two curves $C, D \subset \mathbb{A}^2_K$, how many points can $C(K) \cap D(K)$ have?

Are there always finitely many points? No. A silly example is when $C = D$. For a less silly example, consider $C = \mathbb{V}(x(x^2 + y^2 - 1))$ and $D = \mathbb{V}(x(y - x^2))$. The problem, of course, is that C and D contain a common "component" of the y-axis $\mathbb{V}(x)$. We'll define components formally soon, but let's pretend we understand what they are. It is then at least believable that there can only be finitely many points of intersection.

How many? Let's consider a few examples. If we intersect $C = \mathbb{V}(x^2 + y^2 - 1)$ with the line $D = \mathbb{V}(x - 2)$. Well, there are no solutions over $\mathbb{Q}$ or $\mathbb{R}$, but we know this is a fault of these fields

(c.f. 1.4.3). This suggests we work over an algebraically closed field such as $K = \mathbb{C}$, where we would expect 2 solutions. Over such a field, this example certainly works, and so does an example of the form $C = \mathbb{V}(y - f(x))$ for some polynomial $f(x) \in K[x]$ and the x-axis $D = \mathbb{V}(y)$, where we expect $\deg f$ intersection points. Similarly, we would expect a cubic and a conic to intersect in $2 \cdot 3 = 6$ points. More generally, we would expect curves of degrees d and e to intersect in de points. However, even over such a field, we could take $D = \mathbb{V}(x - 1)$, which gives us just one point $(1, 0)$. A similar example comes from taking $C = \mathbb{V}(y - f(x))$ for some polynomial $f(x) \in K[x]$ and $D = \mathbb{V}(y)$, the x-axis, when $f(x)$ has a repeated root. This suggests we should count things with multiplicites.

Even when you try to do that, sometimes you run into problems. The simplest example of this is when you take $C = \mathbb{V}(x)$ and $D = \mathbb{V}(x - 1)$. In this case, there are no points because the lines are parallel. A more subtle example is when we take $C = \mathbb{V}(y - x^2)$ and $D = \mathbb{V}(sy - x)$ for various s. Again for $s \neq 0$, there are two points of intersection $(0, 0)$ and $(1/s, 1/s^2)$; what happens as $s \to 0$ is that the second point "escapes to infinity", as in 1.3.2 . This suggests that we should look at the projective plane as before. It is a miracle that nothing else goes wrong:

Theorem 1.6.9 (Bézout). Let K be a field, $C, D \subset \mathbb{P}^2_K$ be curves with no common component. Then for any algebraically closed field $\Omega \supset K$, the intersection $C(\Omega) \cap D(\Omega)$ consists of precisely $(\deg C) \cdot (\deg D)$ points, when counted with multiplicites.

This is named after the same guy, the 18^{th}-century French mathematician Étienne Bézout, after whom the identity involving greatest common divisors is named. The history of the proof of this result is quite intriguing, and I will refer you to the corresponding Wikipedia page to read more if you're interested. The result itself (1.6.9) is somewhat nontrivial, and belongs properly to a course on algebraic geometry. Part of the difficulty lies in coming up with a good definition of the intersection multiplicities at points, which is a subtle notion in general. We won't prove it; if you are interested in seeing a full proof, I can refer you to [5, 1.14.1] or [6, §5.3].

Since we won't prove this result, we won't use it either. The only case we will really need is when one of C and D, say D, is a line; that is the case we will prove next time. In the course of proving this result, we will see how to make each of the terms in the statement of the theorem precise.

1.7 26/06/24 - Review of Projective Geometry V: Proof of Bézout's Theorem For Intersection with a Line; The Geometry of Cubic Curves I: The Group Law

Today, we'll wrap up our discussion of projective geometry, and start talking about the geometry of cubic curves in particular.

1.7.A Proof of Bézout's Theorem for Intersection with a Line

As promised, let us now prove Bézout's Theorem for intersection with a line.

Theorem 1.7.1 (Special Case of 1.6.9)**.** Let K be a field, $C \subset \mathbb{P}^2_K$, and $L \subset \mathbb{P}^2_K$ be a line such that C and L don't have a common component, i.e., that $L \not\subset C$. If Ω is an algebraically closed field containing K, then C and L intersect in exactly $\deg C$ points over Ω when counted with multiplicities.

The definitions of the intersection multiplicities will turn up naturally in the proof, which you will see is very easy and natural.

Proof. Without loss of generality, we may assume that L is the X-axis $L = \mathbb{V}(Y)$. In this case, intersecting with L amounts to setting $Y = 0$ in a defining equation for C. Let $F \in K[X, Y, Z]$ be a definiting equation for C; the intersection points then correspond to homogeneous solutions $[X : 0 : Z]$ to $F(X, 0, Z) = 0$.

Since $L \not\subset C$, the polynomial $F(X, 0, Z) \in K[X, Z]$ is not identically zero. Therefore, it is a homogeneous polynomial of degree $d := \deg F = \deg C \in \mathbb{Z}_{\geq 1}$. Since Ω is algebraically closed, $F(X, 0, Z)$ splits completely into linear factors in $\Omega[X, Z]$ as

$$F(X, 0, Z) = \prod_{i=1}^{k} (\mu_i X - \lambda_i Z)^{m_i}$$

corresponding to the say $k \in \mathbb{Z}_{\geq 1}$ points of intersection $P_i = [\lambda_i : 0 : \mu_i] \in C(\Omega)$, where P_i has multiplicity $m_{P_i}(C, L) := m_i \in \mathbb{Z}_{\geq 1}$. Therefore, the number of intersection points over Ω counted with multiplicity is precisely

$$\sum_{i=1}^{k} m_i = d.$$

■

Example 1.7.2. If $K = \mathbb{Q}$ and $F(X, 0, Z) = X^6Z - 2X^5Z^2 - X^4Z^3 + 4X^3Z^4 - 2X^2Z^5$, then this factors as $F(X, 0, Z) = X^2Z(X - Z)^2(X - \sqrt{2}Z)(X + \sqrt{2}Z)$ over an algebraically closed field, say $\Omega = \mathbb{C}$, and the corresponding points $[0 : 0 : 1], [1 : 0 : 0], [1 : 0 : 1], [\sqrt{2} : 0 : 1], [-\sqrt{2} : 0 : 1]$ have multiplicities $2, 1, 2, 1, 1$ respectively, adding up to $7 = \deg F(X, 0, Z)$.

That was straightforward! Actually, the above proof shows something a little stronger:

Corollary 1.7.3 (of the Proof of 1.7.1.)**.** In the set-up of 1.7.1, suppose that $\deg C - 1$ of the points in $C \cap L$ have coordinates in K. Then so does the last one.

Proof. This is just the projective version of the fact that a polynomial $h(x) \in K[x]$ of degree d that has $d - 1$ roots in K (counted with multiplicity) must have its last root in K as well. ■

One last thing I want to point out about the above discussion is that the nonvanishing of $F(X, 0, Z)$ is equivalent to the statement that the monomial Y does not divide F. This would be automatic if F is *irreducible* as a polynomial, of degree $\deg F \geq 2$. A plane curve with irreducible defining polynomial is said to be an *integral* curve. (This applies to both affine and projective plane curves.)

Example 1.7.4. For any field K, the curves $\mathbb{V}(X^2 - Y^2) \subset \mathbb{P}^2_K$ and $\mathbb{V}(X^2) \subset \mathbb{P}^2_K$ are not integral.

Remark 1.7.5. This notion of irreducibility of polynomials (and hence integrality of curves) depends crucially on the base field. For instance, the polynomial $X^2 + Y^2$ is irreducible in $\mathbb{R}[X, Y, Z]$ (check!), but the same polynomial is not irreducible in $\mathbb{C}[X, Y, Z]$, since it factors as $(X + iY)(X - iY)$. A polynomial $F \in K[X, Y, Z]$ is said to be *geometrically irreducible* if it remains irreducible even after replacing K by any field extension L/K. (Again, by arguments similar to those in 1.5.5, it suffices to check one algebraically closed field $L = \Omega$ containing K.) Therefore, the polynomial $X^2+Y^2 \in \mathbb{R}[X, Y, Z]$ is irreducible but not geometrically irreducible. A curve cut out by a geometrically irreducible defining polynomial is said to be *geometrically integral*. Therefore, the curve $C = \mathbb{V}(X^2 + Y^2) \subset \mathbb{P}^2_{\mathbb{R}}$ is integral but not geometrically integral.

As a fun exercise, you could try

Exercise 1.7.6. Show that the "circle" $\mathbb{V}(X^2 + Y^2 - Z^2) \subset \mathbb{P}^2_{\mathbb{R}}$ is geometrically integral.

We will show later that for any field K, the curve $\mathbb{V}(Y^2Z - X^3 - Z^3) \subset \mathbb{P}^2_K$ is geometrically integral. Let's now actually begin talking about the main subject matter for the summer: cubic curves.

1.7.B The Group Law on Points of a Cubic Curve

Let's start by talking about cubic curves. We'll focus on the smooth case for now, and leave the singular case for later.

So suppose that K is field, $C \subset \mathbb{P}^2_K$ is a smooth cubic curve. Suppose that we are given two points $P, Q \in C(K)$. We can then draw the line $L = L_{PQ}$ through them. By 1.6.9, this will intersect C in a third point, say R. (This R could be P, e.g., if L_{PQ} is tangent to C at P. It could be both P and Q if $P = Q$ is a flex!) In fact, since P and Q have coordinates in K, 1.7.3 tells us that $R \in C(K)$ as well. This gives us a way of producing a new K-point from two K-points on C. Following [2], let us temporarily denote the R obtained by such a procedure from P and Q by $P * Q$. What should we do when $P = Q$? Of course, we should take L to be the tangent line $L = \mathbb{T}_P C$, and take $R = P * P$ to be the third point of intersection.

What properties does this $*$ operation have? We would like it to be a group law, but a little experimentation shows that this doesn't work. For instance, could it have an identity element? Suppose it did, and call it $\mathscr{O} \in C(K)$. Then for $\mathscr{O} = \mathscr{O} * \mathscr{O}$ to be true, we would need $\mathscr{O}$ to be a flex point. (That's not horrible yet.) But now for every $P \in C(K)$, we must have $P = \mathscr{O} * P$, which is to say that the line joining $\mathscr{O}$ and P is tangent to C at P. This would mean that *every line through* $\mathscr{O}$ is tangent to C. It is geometrically very believable that it would be very hard to find such a point! We can show with a little effort that such a point does not exist for any smooth cubic curve C over any field K, but let's not bother with this.[3] We have concluded that we need to modify this procedure slightly. Perhaps it is believable from the above discussion, that the procedure is missing a "reflection in $\mathscr{O}$" step. Therefore, let us consider a slightly different operation given as follows.

Fix a smooth cubic $C \subset \mathbb{P}^2_K$ and a point $\mathscr{O} \in C(K)$. (The point $\mathscr{O}$ doesn't have to be a flex point to begin with, but we'll see that taking it to be a flex point does simplify some operations below. We will also see that the choice of $\mathscr{O}$ doesn't matter–at least when considering only the abstract isomorphism type of the group. See below.) Given two points $P, Q \in C(K)$, we can take the line $L = L_{PQ}$ joining them, and then take the third point of intersection $R \in C(K)$ of L with C (where again R is a K-point thanks to 1.7.3). Next, take the line $L_{R\mathscr{O}}$ joining R and $\mathscr{O}$, and take *its* third intersection point with C. Call this intersection point $P \oplus Q$. As above, if $P = Q$, then we start by taking $L = \mathbb{T}_P C$ instead, with the rest being the same. Symbolically, we have in all cases that $P \oplus Q := \mathscr{O} * (P * Q)$.

The fundamental result in the theory is then

[3] Here's a quick proof for the cognoscenti: If such a point exists, then $\pi_P : C \to \mathbb{P}^1_K$ is a purely inseparable morphism of degree 2. This forces $\operatorname{ch} K = 2$ and $g(C) = g(\mathbb{P}^1_K) = 0$ ([7, Proposition IV.2.5]), but this is not possible because $g(C) = 1$. Such "strange" points which lie on every tangent to a given curve can only exist for (lines in any characteristic and) conics in characteristic two; see [7, Theorem IV.3.9].

Theorem 1.7.7 (Fundamental Theorem of Elliptic Curve Theory)**.** For any field K, smooth cubic $C \subset \mathbb{P}^2_K$, and $\mathscr{O} \in C(K)$. The operation $P, Q \mapsto P \oplus Q$ described above makes $C(K)$ an abelian group with identity element $\mathscr{O}$. Further, for any field extension L/K, the natural inclusion $C(K) \subset C(L)$ makes $C(K)$ a subgroup of $C(L)$.

Before, we prove this, let's remark that this is really cool. We started with the set of points on a curve, and gave it an algebraic structure similar to that found on $\mathbb{Z}$ or $\mathbb{Z}/3$. Finally, we can use this additional structure to say something about the set of points on this curve. It is probably an understatement to say that this fundamental theorem is precisely the result that makes the theory of elliptic curves so rich and vast. No such phenomenon exists for (smooth plane) curves of higher degree than 3.

Unimportant Remark 1.7.8. In fact, much more is true than 1.7.7. The group operations on C really come from *algebraic operations*, and this makes C into a *group variety* (or even *abelian variety*). We're not going to use this result (or even this terminology). If you're really interested, you can find a proof of this fact in various places such as [7, IV.4.8], [8, Theorem 3.6], or [9, 19.10.3-4].

Let's try to prove 1.7.7.

Proof Sketch of 1.7.7. . Let's go through the abelian group axioms one at a time. Commutativity is clear, as is the fact that $\mathscr{O}$ is the identity. To create inverses, one could try $\ominus P := \mathscr{O} * P$, but it is easy to see this doesn't work in general. In fact, a little experimentation shows that the right thing to do is to let $\mathscr{O}' := \mathscr{O} * \mathscr{O}$, and then define $\ominus P := \mathscr{O}' * P$. (This is where we see it would be nice to take $\mathscr{O}$ to be an inflection point, because then $\mathscr{O}' = \mathscr{O}$, and the negation is simply $\ominus P = \mathscr{O} * P$. However, you should remember that this really needs $\mathscr{O}$ to be an inflection point.)

The only thing that remains to be checked is associativity. This is the hardest. Let's see what that involves. A simple geometric argument tells us that associativity is the same as the statement that if $P, Q, R \in C(K)$, then $(P \oplus Q) * R = P * (Q \oplus R)$ (draw a picture!). This step genuinely requires argument. There are three proofs of this result I know, each somewhat nontrivial. One pedestrian approach is to write everything out in coordinates and check that the algebra works. However, this is a very daunting task. The classical geometers had a better proof of this result using something known as Chasles's Theorem. This really amounts to either using some dimensional estimates on the set of cubic curves or some local properties of plane curve intersections (an exposition of this proof I gave two years ago at Ross can be found at [5, 1.15.13-14]; you can also read about in [1, §7]). Finally, the third and modern approach uses divisors on curves and the Riemann-Roch Theorem; such a proof can be found in modern books such as [8, III.3.4(e)] or [9, 19.9.3, 19.9.10]. ■

By popular demand, I will give a brief outline of the third proof (using the Riemann-Roch Theorem) next time.

1.8 26/06/25 - The Geometry of Cubic Curves II: Proof Sketch of Associativity

Today, our goal will be to finish the proof sketch of 1.7.7. To do this, I will outline the proof of the associativity of the group law on the cubic using the circle of ideas involving divisors, genus, and the Riemann-Roch Theorem (although I will not use this language).

But before that, I want to take a second to restate the Mordell-Weil theorem in this language as

Theorem 1.8.1 (Mordell-Weil Theorem)**.** Let K be a number field (e.g., $K = \mathbb{Q}$). Then for any smooth cubic curve $C \subset \mathbb{P}^2_K$ with an $\mathscr{O} \in C(K)$, the abelian group $C(K)$ described in 1.7.7 is finitely generated.

Make sure you understand that this is saying the same thing as: there are finitely many "seed" points on $C(K)$ such that every other point can be obtained from these by iteration of the "chord and tangent processes". One of our goals for the summer will be to prove (a simple case of) 1.8.1.

Unimportant Remark 1.8.2. Note that the theorem 1.8.1 is specific to number fields (or more generally global fields). It is not true if we take something like $K = \mathbb{C}$ (an uncountable group has a hard time being finitely generated!), and it is obviously true over a finite field $K = \mathbb{F}_q$, since then $C(\mathbb{F}_q)$ is just finite! This theorem is less obvious but true over fields of the form $K = \mathbb{F}_q(t)$, which are also global fields.

Given this discussion, let's move on to

1.8.A Proof Sketch of Associativity

To prove associativity of the group law, we need to study what it really means to say something like $S = P \oplus Q$. Specifically, we will give an equivalent definition of the group law $\oplus$ on $C(K)$ (1.8.8)–modulo two blackboxes (1.8.11 and 1.8.14), which you can fill in later–and see that this definition is visibly associative. (The following argument is an expanded version of one of the two found in [1, §7].)

Throughout this section, we will work with the following example to see ideas in practice: let K be a field of characteristic other than 2 or 3, and let $C := \mathbb{V}(Y^2Z - X(X+Z)(X+4Z))$ with $\mathscr{O} := [0:1:0]$.

Exercise 1.8.3. Check that when $\operatorname{ch} K \neq 2, 3$, the cubic curve $C = \mathbb{V}(Y^2Z - X(X+Z)(X+4Z)) \subset \mathbb{P}^2_K$ is smooth.

We'll take the points $P = [-4:0:1]$ and $Q = [-2:2:1]$, so that $U := P * Q = [2:6:1]$ and $S := P \oplus Q = [2:-6:1]$. The affine patch $Z \neq 0$ on this curve is pictured in 1.8.4.

Recall that if A is a homogeneous polynomial in $K[X,Y,Z]$, then it doesn't quite make sense to "evaluate" A at points of $\mathbb{P}^2(K)$. However, if A and B are both homogeneous polynomials of the same degree, then it at least makes sense to evaluate A/B at points of $\mathbb{P}^2(K)$ where B doesn't vanish. We've done this before already when we've taken $A = X, Y$ and $B = Z$; this is what gives us the coordinates $x = X/Z$ and $y = Y/Z$.

For another example, in the curve of 1.8.4, we could take A to be a linear form defining the line L_{PQ}, say $A := Y - X - 4Z$, and similarly B to be a linear form defining the vertical line $L_{U\mathscr{O}}$, say $B := X - 2Z$. Then the quotient would be

$$f := \frac{A}{B} = \frac{Y - X - 4Z}{X - 2Z} = \frac{y - x - 4}{x - 2} = \frac{1 - t - 4s}{t - 2s}, \tag{1.8.5}$$

where as before $(x, y) = (X/Z, Y/Z)$, and $(t, s) = (X/Y, Z/Y)$. It now does make sense to evaluate this fraction 1.8.5 away from the locus where $B = 0$. In particular, we can evaluate this function on our curve C, away from the three intersection points $\mathscr{O}, S, U$. For instance, what is its value at $(x, y) = (0, -1)$? It is $5/2$.

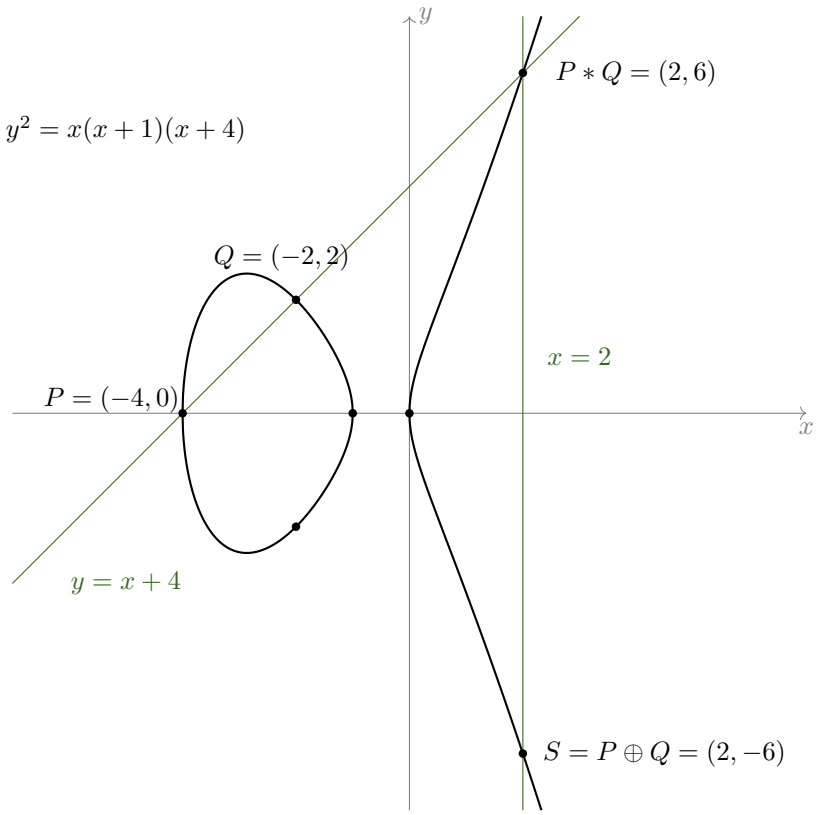


Figure 1.8.4: The $\{Z \neq 0\}$ points of the projective curve $\mathbb{V}(Y^2Z - X(X+Z)(X+4Z)) \subset \mathbb{P}^2_K$.

Where does this function vanish on our curve? Of course at the points P, Q, and U–the locus of intersection of $\mathbb{V}(A)$ and C. Where is it not defined, or where does it blow-up, or where is it infinite? Of course, at the points $\mathscr{O}, S, U$–the locus of intersection of $\mathbb{V}(B)$ and C. But wait: U belongs to both of these. What is happening there?

Exercise 1.8.6. (This exercise is for those who know a little calculus.) Show that if $K = \mathbb{R}$, then

$$\lim_{x \to 2} \frac{\sqrt{x(x+1)(x+4)} - x - 4}{x - 2}$$

exists and is nonzero by calculating it.

In fact, for any field K, the "zero" coming from A and the "pole" coming from B exactly "cancel out" at U, leaving us with no zeroes or poles at this point. Therefore, we have produced an *algebraic function* $f = A/B$ on C with the property that it has zeroes at P and Q, and poles at $\mathscr{O}$ and S.

Exercise 1.8.7. If you know what these words mean, show that the morphism $C \smallsetminus \{U\} \to \mathbb{P}^1_K$ defined by $[A : B]$ extends to give a morphism $C \to \mathbb{P}^1_K$ which maps U to a point different than 0 and ∞. (This is an instance of the "curve-to-projective" extension theorem.) If you don't know what these words mean, you may either try to interpret them or ignore this exercise.

We could take this to be a *definition* of $S = P \oplus Q$:

Definition 1.8.8. In the above set-up, given $P, Q, S \in C(K)$, we define $P \oplus Q = S$ iff there is a nonzero algebraic function f on C with zeroes at P and Q and poles at $\mathscr{O}$ and S (and no zeroes and poles elsewhere).

Of course, for this to be a definition of $P \oplus Q$, we need to do three things:

(a) we need to define what it means to be an algebraic function on the curve C, and what zeroes and poles of an algebraic function mean,

(b) we need to show that given any P and Q, such an S exists, and
(c) we need to show that given any P and Q, such an S is unique.

Let's deal with these issues one at a time.

Given an integral curve $C \subset \mathbb{P}^2_K$, we define the *field of algebraic functions* on C, denoted $K(C)$, to be the field consisting of functions of the form 1.8.5. To say more, recall what it means to say for C to be integral: it means that any defining equation $F \in K[X,Y,Z]$ for C is irreducible. Since $K[X,Y,Z]$ is a UFD, this is the same as saying that F is prime, or that $K[X,Y,Z]/(F)$ is an integral domain. Therefore, we may consider its fraction field, denoted

$$K(\mathrm{Cone}(C)) := \mathrm{Frac}\, K[X,Y,Z]/(F).$$

Now this field is too big and not what we're looking for: it contains things like X, which don't make sense as functions on C. (The notation is suggesting what this is: it is the field of algebraic functions on the cone over C, which is $\mathrm{Cone}(C) := \mathbb{V}(F) \subset \mathbb{A}^3_K$.) What we need is a smaller subfield $K(C) \subset K(\mathrm{Cone}(C))$, which consists of all fractions of the form $[A]/[B]$, where A and B can be chosen to be homogeneous of the same degree.

Definition 1.8.9. Let K be a field and $C \subset \mathbb{P}^2_K$ be an integral projective curve with defining equation $F \in K[X,Y,Z]$. Then we define the *algebraic function field* $K(C)$ of C to be

$$K(C) := \left\{ \frac{[A]}{[B]} : \begin{array}{l} [A],[B] \in K[X,Y,Z]/(F), \text{ such that } [B] \neq 0 \text{ and} \\ A \text{ and } B \text{ can be chosen to be homogeneous of the same degree} \end{array} \right\} \subset K(\mathrm{Cone}(C)).$$

The objects of $K(C)$ can then be thought of as genuine K-valued functions on $C(K)$, except possibly for finitely many points where they have the value "∞".

Example 1.8.10. Let $K = \mathbb{C}$ and $C = \mathbb{V}(Y^2Z - X(X+Z)(X+4Z)) \subset \mathbb{P}^2_{\mathbb{C}}$. Then $\mathbb{C}(C)$ contains things like $x := [X]/[Z]$, $y := [Y]/[Z]$, $t := [X]/[Y]$, and $s := [Z]/[Y]$. The class $[X]/[1]$ belongs to $\mathbb{C}(\mathrm{Cone}(C))$ but not to $\mathbb{C}(C)$. The fraction

$$\frac{[X^3 - Z^3]}{[Y^2Z - X^3 - Z^3]} = \frac{x^3 - 1}{y^2 - x^3 - 1} = \frac{t^3 - s^3}{s - t^3 - s^3}$$

defines an element of $\mathbb{C}(C)$, but the fraction

$$\frac{[X^3 - Z^3]}{[Y^2Z - X(X+Z)(X+4Z)]}$$

does not, since the denominator is zero. As elements of $\mathbb{C}(C)$, we have that

$$y^2 = x(x+1)(x+4) \text{ and } s = t(t+s)(t+4s).$$

In fact,

$$\mathbb{C}(C) \cong \mathrm{Frac}\, \mathbb{C}[x,y]/(y^2 - x(x+1)(x+4)) \cong \mathrm{Frac}\, \mathbb{C}[t,s]/(s - t(t+s)(t+4s)).$$

More generally, the function fields of a curve is the same as the "function field" of any of its nonempty affine patches (2.3.7).

Now we're going to need one important fact:

Fact 1.8.11. Let $C \subset \mathbb{P}^2_K$ be a smooth integral[a] curve with algebraic function field $K(C)$.

(a) For each $P \in C(K)$, we have a (discrete) valuation

$$v_P : K(C)^* \to \mathbb{Z}$$

which measures the order of vanishing of an algebraic function at P.
(b) Given an $f \in K(C)^*$ and a $P \in C(K)$, we say that P is a *zero* (resp. *pole*) of f of order $v_P(f)$ (resp. $-v_P(f)$) if $v_P(f) > 0$ (resp. $v_P(f) < 0$). A zero (resp. pole) of order one is said to be a simple zero (resp. simple pole), and similarly for double/triple zeroes/poles, etc.

(c) If we have a function f of the form $f = [A]/[B]$ where $A, B \in K[X, Y, Z]$ are homogeneous of the same degree and $[B] \neq 0$, then for any $P \in C(K)$, if $A(P) = 0$ but $B(P) \neq 0$, then P is a zero of f. Similarly, if $A(P) \neq 0$ but $B(P) = 0$, then P is a pole of f. (More care is needed if $A(P) = B(P) = 0$; c.f. 1.8.6).

(d) Moreover, if A and B are linear forms, $A(P) = 0$ but $B(P) \neq 0$, and the line $\mathbb{V}(A)$ is *not* tangent to C at P, then $v_P(f) = 1$, i.e., f has a simple zero at P. Similarly, if $A(P) \neq 0$ and $B(P) = 0$, and the line $\mathbb{V}(B)$ is not tangent to C at P, then $v_P(f) = -1$, i.e., f has a simple pole at P. More generally, if $A(P) = 0$ but $B(P) \neq 0$, and $L = \mathbb{V}(A)$, then $v_P(f) = m_P(C, L)$ is the intersection multiplicity of C and L at P as in the proof of 1.7.1.

[a] See also 2.3.8.

For a proof of the first part of 1.8.11, see [8, §II.1] and the references cited there, where v_P is denoted by ord_P. For a proof of the second part (unfortunately not using exactly the same language), see [6, Theorem 1, §3.2].

Remark 1.8.12. The operation v_P on the field $K(C)$ behaves very similarly to the p-adic valuation v_p on the field $\mathbb{Q}$ of rational numbers. This suggests that maybe we can interpret the p-adic valuation in this language. For instance, where does the "function" $12/5$ have zeroes and poles? Of course, it has a "double zero" at 2, a "simple zero" at 3, and a "simple pole" at 5. On what geometric object are rational numbers rational functions?

Unimportant Remark 1.8.13. The same as in 1.8.11 can be done for other closed points $P \in C$, possibly with bigger residue fields. (E.g., this might correspond to something like the $(t^2 + 1)$-adic valuation on the field $\mathbb{F}_3(t)$.) If you have no idea what this means, you can safely ignore this bit.

At this point, we can at least make sense of the individual components of 1.8.8: given a field K, a smooth cubic $C \subset \mathbb{P}^2_K$ with $\mathscr{O} \in C(K)$, and given points $P, Q, S \in C(K)$, we say that $P \oplus Q = S$ iff there is an $f \in K(C)^\times$ such that $v_P(f) = v_Q(f) = 1$ and $v_{\mathscr{O}}(f) = v_S(f) = -1$ (when $P, Q, \mathscr{O}, S$ are all distinct; otherwise, say, if $P = Q$, then we require that $v_P(f) = 2$, etc.) and such that f has no zeroes or poles elsewhere (interpreted correctly, see 1.8.15). To show that this is a definition of $P \oplus Q$, we still need to show that such an S exists and is unique. The existence part, however, is OK: we have seen above that the point $S = P \oplus Q$ defined by our "chord and tangent process" works. (The multiplicities are automatically taken care of by various parts of 1.8.11; details omitted for now.) The uniqueness, however, needs another important fact, namely

Fact 1.8.14. Let K be a field, $C \subset \mathbb{P}^2_K$ a smooth integral curve with algebraic function field $K(C)$. If $S_1, S_2 \in C(K)$ are two distinct points, then there does not exist an $f \in K(C)^\times$ with a simple zero at S_1, a simple pole at S_2, and no zeroes or poles elsewhere, i.e., with $v_{S_1}(f) = -v_{S_2}(f) = 1$, and $v_S(f) = 0$ for all $S \neq S_1, S_2$.

Unimportant Remark 1.8.15. Here again, "elsewhere" needs to be interpreted correctly as: at no other closed points (c.f. 1.8.13). This is not a problem if we assume K is algebraically closed, which we can certainly do for the purposes of proving 1.7.7: simply pass to an algebraic closure $\overline{K}$ and prove the result there, and then note that $C(K) \subset C(\overline{K})$ forms a subgroup. The reason I wanted to not avoid this approach from the outset is that I wanted to emphasize that the function field $K(C)$ can be defined even when K is not algebraically closed; some authors (such as Silverman in [8]) decide to work almost exclusively with $\overline{K}(C)$ instead, where $\overline{K}$ denotes an algebraic closure of K.

This fact (1.8.14) is something special to plane cubic curves (or actually smooth plane curves of degree ≥ 3):

Exercise 1.8.16. Let K be a field, $C \subset \mathbb{P}^2_K$ be a line or a smooth conic. Show that for any two distinct points $S_1, S_2 \in C(K)$, there is an $f \in K(C)^\times$ with $v_{S_1}(f) = -v_{S_2}(f) = 1$, and $v_S(f) = 0$ for all $S \neq S_1, S_2$. (Hint: Use a nice coordinate system. See also 1.6.4.)

Unimportant Remark 1.8.17. For the cognoscenti, the rough idea behind the proof of 1.8.14 is as follows. If f is such a function, then f induces a morphism $C \to \mathbb{P}^1_K$ (c.f. 1.8.7). If f has the mentioned properties, then f has degree 1, and is hence necessarily an isomorphism. But C and $\mathbb{P}^1_K$ are *not* isomorphic: they have different *genera*, namely $g(C) = 1$ but $g(\mathbb{P}^1_K) = 0$. The notion of genus is easiest to understand

when $K = \mathbb{C}$, where it just means "the number of holes". However, it also makes sense in positive characteristic, with one definition being: it is the integer that makes the Riemann-Roch Theorem true. For the notion of genus, see for instance [6, §8.3]. For the Riemann-Roch Theorem, see [6, §8.6]. To see how 1.8.14 follows from the Riemann-Roch Theorem, see [8, Lemma III.3.3].

At this point, uniqueness is disposed of by

Lemma 1.8.18. Let K be a field, $C \subset \mathbb{P}^2_K$ a smooth cubic curve with $\mathscr{O} \in C(K)$. Suppose we are given points $P, Q, S_1, S_2 \in C(K)$, and assume for simplicity that $P \neq Q$ and $S_1, S_2 \neq \mathscr{O}$. Suppose for $i = 1, 2$ that there are $f_i \in K(C)^\times$ such that $v_P(f_i) = v_Q(f_i) = 1$ and $v_{\mathscr{O}}(f_i) = v_{S_i}(f_i) = -1$, with both f_1 and f_2 having no zeroes or poles elswhere. Then $S_1 = S_2$.

Proof. If $S_1 \neq S_2$, then $f := f_2/f_1$ has the property that $v_{S_1}(f) = 1$ and $v_{S_2}(f) = -1$, and that $v_S(f) \neq 0$ elsewhere (1.8.15); this contradicts 1.8.14. ∎

The cases where $P = Q$ or one of the S_i coincide with $\mathscr{O}$ is similar and left to the reader:

Exercise 1.8.19. Show that the result of 1.8.18 is also true if $P = Q$ or if one of the S_i is $\mathscr{O}$. For instance, show that if $P = Q$ but $S_1, S_2 \neq \mathscr{O}$, and for $i = 1, 2$ there are $f_i \in K(C)^\times$ such that $v_P(f_i) = 2$ and $v_{\mathscr{O}}(f_i) = v_{S_i}(f_i) = -1$, with both f_1 and f_2 having no zeroes or poles elsewhere, then $S_1 = S_2$. State and prove a similar result when (a) $P \neq Q$, but one of the S_i coincides with $\mathscr{O}$, and (b) $P = Q$ and one of the S_i coincides with $\mathscr{O}$. (Hint: The same proof works.)

Unimportant Remark 1.8.20. The above exercise 1.8.19 shows that it would be very convenient to have the language of adding and subtracting zeroes and poles as abstract objects; this would allow us to give a uniform proof of 1.8.18 and the result in 1.8.19, which clearly contain the same ideas. This is indeed possible, and is achieved by introducing the notion of a *divisor* on a curve, which is basically an integer-linear combination of points on it. If you're interested in this notion, come and talk to me, or see [6, §8.1] or [8, §II.3].

Finally, at this stage, 1.8.8 makes sense: we've defined what the terms mean, and shown that such an S exists and is unique. The associativity statement now follows from

Lemma 1.8.21. Let K be a field, $C \subset \mathbb{P}^2_K$ be an (integral) smooth cubic curve, and let $\mathscr{O} \in C(K)$. Suppose we are given points $P, Q, R, T \in C(K)$, and let $S = P \oplus Q$. Suppose for simplicity that $\mathscr{O}, P, Q, R, S, T$ are all distinct. Then we have that

$$T = (P \oplus Q) \oplus R$$

if and only if there is an $h \in K(C)^\times$ which has simple zeroes at P, Q, R, a double pole at $\mathscr{O}$, and a double pole at T, and no other zeroes or poles, i.e.,

$$v_P(h) = v_Q(h) = v_R(h) = 1 \text{ and } v_{\mathscr{O}}(h) = -2 \text{ and } v_T(h) = -1,$$

and such that $v_U(h) = 0$ for all $U \neq \mathscr{O}, P, Q, R, T$.

Proof. We need to show two implications. In both cases, from the equality $S = P \oplus Q$ and 1.8.8, there is an $f \in K(C)^\times$ such that $v_P(f) = v_Q(f) = -v_{\mathscr{O}}(f) = -v_S(f) = 1$, and which has no other poles or zeroes.

(a) Suppose first that $T = (P \oplus Q) \oplus R$, and we have to produce an h with the required properties. From $T = S \oplus R$, there is a $g \in K(C)^\times$ such that $v_S(g) = v_R(g) = -v_{\mathscr{O}}(g) = -v_T(g) = 1$, and such that g has no other poles or zeroes. Then taking $h = f \cdot g$ works (check!).

(b) Conversely, suppose that there is an h with the required properties, and we have to show that $T = S \oplus R$. Well, by 1.8.8, it suffices to produce a $g \in K(C)^\times$ such that $v_S(g) = v_R(g) = -v_{\mathscr{O}}(g) = -v_T(g) = 1$, and such that g has no other poles or zeroes. But the choice $g := h/f$ works (check!).

∎

Again, edge cases are left as an

Exercise 1.8.22.

(a) Show that an appropriate version of the result of 1.8.21 also holds when some of the $\mathscr{O}, P, Q, R, S, T$ coincide.
(b) Show that associativity of $\oplus$ follows from 1.8.21 and the edge cases covered in (a).

This completes our proof sketch of the associativity of the group law on a smooth cubic with a rational point.

1.9 26/06/29 - Weierstrass Equations I

Today, we will roughly follow §1.3-1.4 of [2].

Last time, we showed that if K is a field and $C \subset \mathbb{P}^2_K$ a smooth cubic curve, then for any $\mathscr{O} \in C(K)$, the operation $\oplus$ given by $C(K) \ni P, Q \mapsto P \oplus Q := \mathscr{O} * (P * Q)$ makes $C(K)$ into an abelian group. To finish up that discussion, I only want to remark one more thing: a priori, this group structure depends on the choice of $\mathscr{O}$, so it's not clear what happens if you take two different points $\mathscr{O}_1, \mathscr{O}_2 \in C(K)$. Are the resulting group structures on $C(K)$ the same? They can't *literally* be the same if $\mathscr{O}_1 \neq \mathscr{O}_2$, since they have different identity elements. Is there a different way they are the same? This I leave you to explore in Exercise 2.3.9.

At this point, we are finally in a position to be able to define the eponymous *elliptic curves*.

1.9.A Elliptic Curves and Weierstrass Equations

Let me start by giving two different definitions of elliptic curves.

Definition 1.9.1. Let K be a field. An *elliptic curve* over K is a pair $(E, \mathscr{O})$, where E is a smooth projective geometrically integral curve of genus one, and $\mathscr{O} \in E(K)$.

Definition 1.9.2. Let K be a field. An *elliptic curve* over K is a curve $E \subset \mathbb{P}^2_K$ with equation of the form

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3, \tag{1.9.3}$$

where $a_1, a_2, a_3, a_4, a_6 \in K$ are such that $\Delta \neq 0$, where Δ is given by the formula

$$\begin{aligned}\Delta = &-a_1^6a_6 + a_1^5a_3a_4 - a_1^4a_2a_3^2 - 12a_1^4a_2a_6 + a_1^4a_4^2 + 8a_1^3a_2a_3a_4 + a_1^3a_3^3\\ &+ 36a_1^3a_3a_6 - 8a_1^2a_2^2a_3^2 - 48a_1^2a_2^2a_6 + 8a_1^2a_2a_4^2 - 30a_1^2a_3^2a_4 + 72a_1^2a_4a_6 + 16a_1a_2^2a_3a_4\\ &+ 36a_1a_2a_3^3 + 144a_1a_2a_3a_6 - 96a_1a_3a_4^2 - 16a_2^3a_3^2 - 64a_2^3a_6 + 16a_2^2a_4^2 + 72a_2a_3^2a_4\\ &+ 288a_2a_4a_6 - 27a_3^4 - 216a_3^2a_6 - 64a_4^3 - 432a_6^2,\end{aligned}$$

equipped with the point $\mathscr{O} = [0 : 1 : 0] \in E(K)$.

Note that this definition (1.9.2) is correct as stated in any characteristic for K, including $\operatorname{ch} K = 2, 3$. However, this leaves many questions unanswered. For instance, where does this monstrous formula for Δ come from, and how does one come up with it? More on this soon. The form of the curve in 1.9.3 is called the (projective) Weierstrass form, named after the 19^{th} century German mathematician Karl Weierstrass (also spelled "Weierstraß", notably in German), who laid the foundation for the modern theory of elliptic curves.

Given the tools we have so far, the first definition (1.9.1) doesn't even make sense, since we haven't defined the genus of a curve (but see 1.8.17). I am also lying a little bit when I say that we defined everything else in the first definition, since I have only given you the definition of smooth projective *plane* curves; but there are curves more general than these, and we will meet some of these below (1.9.5(c)). Finally, I need to explain where the convention for naming the coefficients of the defining equation of an elliptic curve comes from; this I will explain below in 1.9.15(b). But before all of this, let me mention the

Fact 1.9.4. These two definitions of elliptic curves (1.9.1 and 1.9.2) are equivalent.

We will not need this fact, and so I will not prove it here. We will sketch below (1.9.9) the proof that every object of the kind mentioned in 1.9.2 satisfies the conditions of 1.9.1 (except for the genus one bit), but going the other direction really needs the Riemann-Roch Theorem; the interested reader can find the proof in [8, Proposition III.3.1]. We will adopt 1.9.2 as our official definition. This might raise the question of why bother with the first definition (1.9.1) at all; let me try to explain this now.

The basic idea is that smooth projective geometrically integral curves do show up in lots of other contexts. Let me give you a few examples (without proof).

Example 1.9.5.

(a) If $C \subset \mathbb{P}^2_K$ is a smooth curve and there is an $\mathscr{O} \in C(K)$, then $(C, \mathscr{O})$ is an elliptic curve according to 1.9.1. Indeed, smoothness and projectivity comes from the definition, geometric integrality come from a good salvage to 2.3.8(b), and the fact that C has genus one comes from, say, the degree-genus formula for smooth plane curves (see, e.g., [9, 18.6.7.1]). The transformation from C into Weierstrass form (1.9.3) is easy to achieve when $\mathscr{O} \in C(K)$ is an inflection point (1.9.6), but it harder to achieve when $\mathscr{O}$ is not an inflection point (you can find one way to do this, due to Nagell, in [1, §8]). This allows us to put lots of curves in Weierstrass form, and then use special techniques to study curves of this sort. This is exactly what happened when we were studying the (generalized) Fermat problem for $n = 3$ in §1.2.B; we found a change of coordinates that put $x^3+y^3 = \alpha$ (or more precisely its projective closure) into Weierstrass form to rephrase the problem. (We are yet to finish this proof.) We will also use 1.7.7 combined with our implication that 1.9.2 implies 1.9.1 to define the group law on an elliptic curve in Weierstrass form.

(b) Let K be a field. If a curve C is a smooth projective curve with a degree two cover of $\mathbb{P}^1_K$ branched at four points, then C has genus one, and is hence an elliptic curve in the sense of 1.9.1 if we can equip it with a K-rational point.[4] This is what enabled us to transform a curve of the form $v^2 = g(u)$ for $g(u) \in K[u]$ a quartic into Weierstrass form as in 2.1.12, giving us another way of approaching the Fermat problem for $n = 4$.

(c) Suppose we are interested in four-term sequences of integers $p < q < r < s$ such that p^2, q^2, r^2, s^2 are in arithmetic progression. (We already have enough tools to study such sequences of length three; see 2.3.4.) I hope that you agree that this is a reasonably natural question. To do this, we might look at the homogeneous equations

$$p^2 + r^2 = 2q^2 \text{ and } q^2 + s^2 = 2r^2.$$

These are homogeneous equations, and hence each cut out surfaces, say S_1 and S_2 in $\mathbb{P}^3_K$ when we equip it with homogeneous coordiantes $[p : q : r : s]$. It them turns out that the intersection $C := S_1 \cap S_2 \subset \mathbb{P}^3_{\mathbb{Q}}$ (at least in sufficiently large characteristic; I think ≥ 5 works) is a smooth projective geometrically integral curve of genus one ([9, 18.6.R]). Since it clearly has a rational point, say $[1 : -1 : -1 : 1] \in C(\mathbb{Q})$, this tells us that the curve C is an elliptic curve in the sense of 1.9.1, and so the abstract theory guarantees that we can put it in Weierstrass form 1.9.2. If we can make this transformation explicit, then this give us a way to understand four-term sequences of squares in arithmetic progression. I invite you to do this in an elementary fashion in 2.3.5.

One general cultural remark (really due to Dan Shapiro): if you're studying a new field of mathematics, then you should think about questions or theorems, the statements of which do not involve any concepts from your theory, but the "most natural" answers or proofs of which use the theory in an essential way. Further, if you are having a really hard time coming up with examples of such things, then you should carefully reevaluate why you might be interested in studying that particular theory. I hope that the Fermat examples and the example of 1.9.5(c) can provide a few such natural questions which might motivate you to study the theory of elliptic curves.

In lieu of proving 1.9.4 in full generality, let me at least tell you how to transform a smooth cubic $C \subset \mathbb{P}^2_K$ over some field K with an inflection point $\mathscr{O} \in C(K)$ into Weierstrass form by a change of coordinates.

Lemma 1.9.6. Let K be a field, $C \subset \mathbb{P}^2_K$ be a smooth cubic, and let $\mathscr{O} \in C(K)$ be an *inflection point*. Then there is a projective change of coordinates that brings C into the projective Weierstrass form (1.9.2), with $\mathscr{O}$ mapping to $[0 : 1 : 0]$.

Recall that $\mathscr{O} \in C(K)$ is said to be an *inflection point* or a *flex point* iff $\mathbb{T}_{\mathscr{O}}C$ intersects C with multiplicity 3 at $\mathscr{O}$, or equivalently iff $\mathscr{O} * \mathscr{O} = \mathscr{O}$. The following proof is very similar to that of 1.6.4, which might be helpful to review now.

[4] To be more precise, I should say: the morphism $C \to \mathbb{P}^1_K$ should be finite separable of degree two, where the ramification locus $R \subset \mathbb{P}^1_K$ is a closed subscheme of length 4 with at least one K-rational point in $R(K)$. If you have no idea what this means, you can safely ignore this remark.

Proof. Take a projective change of coordinates which sends $\mathscr{O}$ to $[0:1:0]$, and sends $\mathbb{T}_{\mathscr{O}}C$ to the line $L_\infty = \mathbb{V}(Z)$ at infinity. In this coordinates system, pick a defining equation F for C, and express it in the form

$$F = A_0Y^3 + A_1Y^2 + A_2Y + A_3,$$

where the $A_i \in K[X,Z]$ is homogeneous of degree i for $i = 0,\dots,3$. The fact that $\mathscr{O} = [0:1:0] \in C(K)$ implies that $A_0 = 0$. The fact that $\mathbb{T}_{\mathscr{O}}C = \mathbb{V}(Z)$ implies (check!) that A_1 is a nonzero scalar multiple of Z. Therefore, by rescaling F appropriately, we may assume that $A_1 = Z$. Finally, the fact that $\mathscr{O} = [0:1:0]$ is an inflection point on C implies that the tangent line $L_\infty = \mathbb{V}(Z)$ intersects C with multiplicity three at $\mathscr{O}$, which forces that $Z \mid A_2$ (check!). This means that F has the form

$$Y^2Z + (a_1X + a_3Z)YZ - (a_0X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3)$$

for some $a_0, a_1, a_2, a_3, a_4, a_6 \in K$. Can a_0 be zero? If this is the case, then $Z \mid F$, and so C is not integral; but this can't happen if C is smooth, basically by Bézout's Theorem (see 1.7.1 and 2.3.8(b)). Therefore, $a_0 \neq 0$, and we may multiply F throughout by a_0^2 and let $(X', Y') := (a_0X, a_0Y)$ to get an equation in Weierstrass form (1.9.2), up to minor relabeling of the coefficients. It remains to justify that the smoothness of C implies that $\Delta \neq 0$; this we do in a more general context below (1.9.9). ■

Exercise 1.9.7. Fill in the gaps in the above proof: check that $A_0 = 0$ follows from $\mathscr{O} \in C(K)$, check that $A_1 = Z$ (up to scaling) follows from $\mathbb{T}_{\mathscr{O}}C = \mathbb{V}(Z)$, and check that $Z \mid A_2$ follows from the fact that $\mathscr{O} \in C(K)$ is an inflection point. It may be helpful to use local coordinates $(t,s) = (X/Y, Z/Y)$ around the point $(0,0)$.

Exercise 1.9.8. Let K be a field of characteristic other than 3, and let $\alpha \in K^\times$. Show that the curve $E_\alpha\mathbb{V}(X^3 + Y^3 - \alpha Z^3) \subset \mathbb{P}^2_K$ of §1.2.B and 2.2.4 has an inflection point at $[1:-1:0] \in E_\alpha(K)$, and that the linear change of coordinates 1.2.8 is obtained exactly by the procedure described in the proof of 1.9.6, up to minor modifications (which you should figure out).

Let us now sketch the proof of the result I promised which implies that a curve in Weierstrass form (1.9.2) has the properties of 1.9.1, except for the part involving the genus, which we will not need (and a proof of which can be found at [9, 18.6.7.1]).

Proposition 1.9.9. Let K be a field (of any characteristic!) and $E \subset \mathbb{P}^2_K$ be a curve in Weierstrass form, as described in 1.9.2. Then E is smooth iff $\Delta \neq 0$. In particular, if this is the case, then E is a smooth projective geometrically integral curve of genus one, with $\mathscr{O} = [0:1:0] \in E(K)$, i.e., an elliptic curve in the sense of 1.9.1.

A few remarks are in order. Firstly, as I mentioned above, I will not prove the part involving the genus. Secondly, as before, I will invoke 2.3.8(b) to reduce to showing only that E is smooth iff $\Delta \neq 0$. I don't feel too bad about this, because we might later show that any curve in Weierstrass form (smooth or not!) is automatically geometrically integral. Thirdly, we will also show later a simpler version of this statement in characteristic other than 2 or 3. Finally, I will only give the key ideas behind the proof of 1.9.9, and leave some details to the reader. A full proof, using slightly different terminology, can be found at [8, Proposition III.1.4(a)].

Proof Sketch of 1.9.9. Let's first begin by study the points at infinity. Setting $Z = 0$ in 1.9.3 gives us $X^3 = 0$, which shows that $\mathscr{O} = [0:1:0] \in E(K)$ is the unique point at infinity on a curve in Weierstrass form. Either by using the coordinates $(t,s) = (X/Y, Z/Y)$, or by taking $\partial F/\partial Z$, it is easy to see (check!) that $\mathscr{O}$ is a smooth point with tangent line $L_\infty = \mathbb{V}(Z)$. (In fact, the fact that setting $Z = 0$ gives us $X^3 = 0$ tells us that $\mathscr{O}$ is an inflection point on $E(K)$, but we won't need this.)

Having checked the unique point at infinity for smoothness (which is true irrespective of whether $\Delta \neq 0$), we can now restrict to studying the affine equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \tag{1.9.10}$$

which corresponds to the vanishing of $f := y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6 \in K[x,y]$. To check smoothness, we evaluate the partial derivatives

$$\frac{\partial f}{\partial x} = a_1y - 3x^2 - 2a_2x - a_4 \text{ and } \frac{\partial f}{\partial y} = 2y + a_1x + a_3. \tag{1.9.11}$$

For simplicity, first assume that $\operatorname{ch} K \neq 2, 3$ and that $a_1 \neq 0$. Then if $\partial f/\partial y = 0$, the we can express y in terms of x as

$$y = -\left(\frac{a_1 x + a_3}{2}\right). \tag{1.9.12}$$

Plugging this into the expression for $\partial f/\partial x$ in 1.9.11 yields a quadratic equation for x, which we can solve for using the quadratic formula. Having solved for x, we can then use 1.9.12 to express y in terms x and plug both into 1.9.10 to get a huge equation involving just one huge square root. We can then move this square root to one side, take squares, and rearrange along with some very minor modifications to arrive precisely at Δ. This tells us that there is a singular point iff $\Delta = 0$. (This procedure might explain the huge size of Δ.)

This proof doesn't quite cover all cases. We still need to worry about small characteristics and whether $a_1 = 0$. For the purposes of illustration, let me do one more case: that of $\operatorname{ch} K = 2$ and $a_1 = 0$. Note that in characteristic two, the expression for Δ simplifies tremendously to

$$\Delta = a_1^6 a_6 + a_1^5 a_3 a_4 + a_1^4 a_2 a_3^2 + a_1^4 a_4^2 + a_1^3 a_3^3 + a_3^4. \tag{1.9.13}$$

In particular, if further $a_1 = 0$, then $\Delta = a_3^4$ simply. If there is a singular point, then $\partial f/\partial y = 0$ implies $a_3 = 0$ and hence $\Delta = 0$. Conversely, if $\Delta = 0$, then $\partial f/\partial y = 0$ identically. The expression for $\partial f/\partial x$ simplifies to $x^2 + a_4$, so that we may solve to get $x = a_4^{1/2}$. Finally, plugging $x = a_4^{1/2}$ into 1.9.10 yields the quadratic equation (check!)

$$y^2 + (a_1 a_4^{1/2} + a_3)y + a_2 a_4 + a_6 = 0,$$

which can certainly be solved for y (over an algebraically closed field $\Omega \supset K$!), yielding a singular point $(x, y) \in E(\Omega)$, and showing that E is not smooth.

The other cases are similar and left to the reader as an extended exercise. ■

Exercise 1.9.14. Do a few more cases to convince yourself that the result is true as claimed in all cases. Then, if you feel the need to, look up the general proof in [8, Proposition III.1.4(a)].

Remark 1.9.15.

(a) The affine equation 1.9.10 is what is often called the *Weierstrass form* of the elliptic curve. When we say "an elliptic curve with Weierstrass equation 1.9.10", what is really meant is the projective closure 1.9.3, which is an elliptic curve iff $\Delta \neq 0$. In particular, even though we'll work with the affine equation 1.9.10, it should be kept in mind that the curve $E \subset \mathbb{P}^2_K$ we are considering is a projective curve with the point $\mathscr{O} = [0 : 1 : 0] \in E(K)$ at infinity.

(b) The affine equation 1.9.10 explains the notation for the coefficients a_i: if we assign the weight 2 to x and 3 to y, then 1.9.10 becomes a homogeneous equation iff we assing weight i to a_i for $i = 1, 2, 3, 4, 6$. This also explains why there is no a_5. There are many reasons why we might want to do this; for now, it suffices to note that this will make every constant we define come with a natural weight. For instance, you can check that although Δ (given in 1.9.2) is not homogeneous in $a_1, \ldots, a_6$ if we assign the weight 1 to each a_i, it *is* homogeneous of degree 12 if we assign weight i to a_i. This might be helpful to keep in mind when we define the constants c_4 and j later.

Unimportant Remark 1.9.16. The above discussion may not fully convincing about the real origins of Δ, even though it is easier to check that this statement (1.9.9) is correct. Indeed, the formula for Δ in 1.9.2 really comes from *elimination theory*, which answers the following question: what is the "best" way to eliminate the variables x and y from the system of equations (1.9.10) and (1.9.11)? This amounts to finding generators for the ideal J given as the intersection

$$J := I \cap \mathbb{Z}[a_1, a_2, a_3, a_4, a_6] \subset \mathbb{Z}[a_1, a_2, a_3, a_4, a_6], \tag{1.9.17}$$

where I is the ideal

$$I = \left(f, \frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \in \mathbb{Z}[a_1, \ldots, a_6, x, y].$$

The ideal J of 1.9.17 is known as the *elimination ideal*, and it can be shown, say using the theory of Gröbner bases, that J is principal, generated precisely by Δ. (If you don't trust me–and indeed, you shouldn't be willing to take things on faith–then you can easily check this fact in software such as, say,

Macaulay2. There is a good geometric reason why the ideal J should be principal; I'm not going to include this here, but if you are interested, you should ask me.) This is where Δ really comes from–it is the generator of the elimination ideal. If you want to read more about elimination theory, you can find it in [10] when working over fields and [11, Chapter 4] over more general rings such as $\mathbb{Z}$ (which is basically what Macaulay2 is actually doing).

In closing today, I wish to remark that when the characteristic of K is not small, then we may come up with changes of coordinates to simply 1.9.10 even further. For instance, if $\operatorname{ch} K \neq 2$, then we may complete the square on the left hand side of 1.9.10 by taking $x_1 = x$ and

$$y_1 = y + \frac{a_1 x + a_3}{2}$$

to get a shorter Weierstrass equation of the form

$$y_1^2 = x_1^3 + a x_1^2 + b x_1 + c. \tag{1.9.18}$$

Note that you *cannot* do something along these lines $\operatorname{ch} K = 2$, since the curve 1.9.18 is necessarily singular in characteristic two–check!. Finally, if $\operatorname{ch} K \neq 2, 3$, then we may simplify 1.9.18 even further by taking $(x_2, y_2) = (x_1 + (a/3), y_1)$ to depress the cubic, yielding an even shorter Weierstrass equation of the form

$$y_2^2 = x_2^3 + b' x_2 + c'. \tag{1.9.19}$$

Again, you may not be able to do this if $\operatorname{ch} K = 3$. We will study these shorter Weierstrass forms in more detail next time.

Remark 1.9.20. One can check that Δ remains invariant under the operations of the form that allowed to us bring curves into short Weierstrass form. This is almost miraculous. For an example of what I mean by this: if we use that substitution $(x_1, y_1) = (x, y + (a_1 x + a_3)/2)$ when $\operatorname{ch} K \neq 2$, then you can check that 1.9.10 transforms into 1.9.18, where

$$(a, b, c) = \left(a_2 + \frac{1}{4} a_1^2, a_4 + \frac{1}{2} a_1 a_3, a_6 + \frac{1}{4} a_3^2 \right).$$

(Check this!) What I mean by the invariance of Δ is then that

$$\Delta(a_1, a_2, a_3, a_4, a_6) = \Delta(0, a, 0, b, c) := \Delta\left(0, a_2 + \frac{1}{4} a_1^2, 0, a_4 + \frac{1}{2} a_1 a_3, a_6 + \frac{1}{4} a_3^2\right),$$

both sides being equal to

$$\Delta = -16(4a^3 c - a^2 b^2 - 18abc + 4b^3 + 27c^2). \tag{1.9.21}$$

A similar thing happens with the substitution that depressed the cubic to be of the form 1.9.19. (Check this!) This is something special to the mysterious polynomial Δ, and obviously not true of all polynomials in the a_i. This mysterious invariance property of Δ is incredibly helpful when working with shorter Weierstrass forms 1.9.18 or 1.9.19, because then we can just use the simplified expression 1.9.21.

Unimportant Remark 1.9.22.

(a) For analogs of these shorter Weierstrass forms when $\operatorname{ch} K$ is 2 or 3, you may consult [8, Appendix A] or [12, Chapter 3].

(b) One might ask why I have bothered at all to define the longer Weierstrass form 1.9.10 at all, when we could work with the shorter forms 1.9.18 or 1.9.19, as is done, for instance, in [2]. There are a few reasons for this. One of them is that the longest Weierstrass form 1.9.10 works in *any* characteristic; in particular, it works in characteristics 2 and 3, in which the shorter ones might not. Secondly, software such as Sage is used to taking input into the form of a five-tuple $(a_1, a_2, a_3, a_4, a_6)$ produce elliptic curves, and I wanted you to understand what Sage is doing. Finally, even when working exclusively with elliptic curves over $\mathbb{Q}$, we sometimes need the long Weierstrass form 1.9.10, for instance, when finding convenient forms in which to express 2-isogenies (more on this soon!), or trying to find minimal models to study primes of bad reduction, or when giving short forms in which to express the equation of the curve when the coefficients become huge, as is needed for the curves involved in state-of-the-art records on elliptic curve ranks over $\mathbb{Q}$ (more on this later!). Even though it might be easiest to start with the shortest Weierstrass form, I think that it is not much more difficult pedagogically to start with the longer Weierstrass form 1.9.10, which allows an easier transition into topics mentioned in this remark.

1.10 26/06/30 - Weierstrass Equations II and Sage

In the first part of class today, I want to talk a little more about Weierstrass equations and changes of coordinates, and then focus on using Sage for the rest of the lecture.

1.10.A One More Thing About the Elliptic Discriminant

For simplicity now, let us focus on the case of a field K of characteristic not two. We saw last time that any elliptic curve E over K can be put into the shorter Weierstrass form

$$y^2 = x^3 + ax^2 + bx + c$$

for some a, b, c. For this to define a nonsingular curve, we need the discriminant Δ, given by the formula

$$\Delta = 16(-4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2),$$

to be nonzero. In this setting of $\operatorname{ch} K \neq 2$, let us review where this formula comes from. Let $f(x) := x^3 + ax^2 + bx + c \in K[x]$, so E is defined by $y^2 = f(x)$. The singular points of E are cut out by the vanishing locus of $y^2 - f(x)$, and its partial derivatives $2y$ and $-f'(x)$. Since we are assuming $\operatorname{ch} K \neq 2$, if $2y = 0$, then $y = 0$. Therefore, having verified above that the point at infinity is not a singularity for curves in Weierstrass form, we see that E has a singular point iff the polynomials $f(x)$ and $f'(x)$ share a common root. By what you know from the Ross sets, this happens iff a polynomial $\operatorname{disc}(f)$ in the coefficients a, b, c of f vanishes, and indeed we have that

$$\Delta = 16 \cdot \operatorname{disc}(f).$$

This is essentially where Δ comes from, modulo some more care to accommodate all characteristics at once. The 16 is a reminder that the above discussion only works when $\operatorname{ch} K \neq 2$!

1.10.B Admissible Changes of Coordinates

Example 1.10.1. For simplicity, let's work over $K = \mathbb{Q}$.

(a) Consider the two elliptic curves

$$E : y^2 = x(x+1)(x+4) \text{ and } E' : (y')^2 = (x'-1)((x')^2 - 4).$$

Get Desmos to draw plots of both of these curves. What do you notice? Indeed, these curves are related by the change of coordinates $(x, y) = (x' - 2, y')$.

(b) Consider the two elliptic curves

$$E : y^2 + y = x^3 - x \text{ and } E' : (y')^2 = (x')^3 - 16x' + 16.$$

Repeating the above procedure, we see that E and E' are related by the change of coordinates

$$(x, y) = \left(\frac{1}{4}x', \frac{1}{8}y' - \frac{1}{2}\right).$$

At this point, I want to go back to working over general K (of any characteristic). We need a general procedure to understand when two elliptic curves are related by a simple change of coordinates. Suppose that K is a field, and that $E, E' \subset \mathbb{P}^2_K$ are elliptic curves in Weierstrass form with coefficients a_i and a_i' respectively. Suppose that $\varphi : \mathbb{P}^2_K \to \mathbb{P}^2_K$ is a linear change of coordinates (where we think of the first $\mathbb{P}^2_K$ has having coordinates X', Y', Z', and the second one as having coordinates X, Y, Z), taking E' to E and $\mathscr{O}' \in E(K)$ to $\mathscr{O} \in E'(K)$. Then by 2.2.6(b), φ takes $\mathbb{T}_{\mathscr{O}'}E' = \mathbb{V}(Z')$ to $\mathbb{T}_{\mathscr{O}}(E) = \mathbb{V}(Z)$, and fixes the point $[0:1:0]$. It follows from 2.2.2 that φ corresponds to a matrix in $\mathrm{PGL}_3(K)$ of the form

$$\begin{bmatrix} p & 0 & r \\ n & q & t \\ 0 & 0 & 1 \end{bmatrix}$$

for some $n, p, q, r, t \in K$ with $pq \neq 0$. In other words, the change of coordinates is given by $(X, Y, Z) = (pX' + rZ', nX' + qY' + tZ', Z')$, or in affine coordinates $(x, y) = (px' + r, nx' + qy' + t)$. Using this coordinate change to transform the equation of E into that of E', we get immediately by comparing the leading terms in a Weierstrass equation that $q^2 = p^3$. In particular, $(p, q) = (u^2, u^3)$ for some $u \in K^\times$, namely $u := q/p$. Finally, if we let $s := nu^{-2}$, then we end up with the formula

$$(x, y) = (u^2x' + r, u^3y' + u^2sx' + t) \tag{1.10.2}$$

for the change of coordinates. A change of coordinates as in 1.10.2 is said to be an *admissible change of coordinates*. Such a change of coordinates related the coefficients a_i and a_i' in a very particular way; for instance, $ua_1' = a_1 + 2s$.

Exercise 1.10.3. Check this: check that under 1.10.2, the coefficients a_1 and a_1' are related by $ua_1' = a_1 + 2s$. In particular, if $\operatorname{ch} K = 2$, then whether or not $a_1 = 0$ for a particular elliptic curve E/K is a (boolean) invariant of the curve E, invariant under admissible changes of coordinates (and hence by 1.10.6 under isomorphisms). We will revisit this idea later.

Similarly, you can work out for yourself how the rest of the coefficients a_i and a_i' are related (in terms of (u, r, s, t)) (or you can find it at [8, Table 3.1]). I want to highlight two things: that if $r = s = t = 0$, the transformation is simply $u^i a_i' = a_i$ for all $i = 1, 2, 3, 4, 6$ (c.f. 1.9.15(b)), and that the discriminants Δ and Δ' of E and E' are miraculously related by the very simple formula

$$u^{12}\Delta' = \Delta. \tag{1.10.4}$$

(You need not take my word for it, and can verify this for yourselves.) In particular, if $u = 1$, then $\Delta' = \Delta$. In particular, all the changes we did to transform curves into shorter Weierstrass forms when $\operatorname{ch} K \neq 2, 3$, which involved things like completing squares and depressing cubics, did not change Δ (c.f. 1.9.20). We summarize this as:

Proposition 1.10.5. Let K be a field. Two elliptic curves E and E' over K in Weierstrass form are the same iff they are related by an admissible change of coordinates of the form 1.10.2.

Unimportant Remark 1.10.6. Technically speaking, we haven't fully proved 1.10.5; the giveaway is the usage of the imprecise terminology of being "the same". It can be shown that any isomorphism $E \to E'$ of (abstract) elliptic curves in the sense of 1.9.1 taking $\mathscr{O}$ to $\mathscr{O}'$ is induced by an admissible change of coordinates when E and E' are put in Weierstrass form. This again uses the Riemann-Roch Theorem. We will not need this fact, and we won't prove it. The interested reader can find a proof at [8, Proposition III.3.1(b)].

1.10.C Working with Sage

The rest of the lecture was spent discussing how to work with Sage. We learned how to work with various fields in Sage, how to construct elliptic curves over various fields and study their properties, and how to work with explicit admissible changes of coordinates. This is a pain to typeset (well) in LaTeX, so I am not going to do this here, but a version of this discussion can be found in [13, Appendix A], to which I refer the interested reader.

1.11 26/07/01 - Weierstrass Equations III and the j-Invariant

Today, I want to do a few things. Firstly, I want to start by saying something I haven't yet explicitly, but which has been implicit in what we are doing.

Let K be a field and $E \subset \mathbb{P}^2_K$ be an elliptic curve in Weierstrass form (1.9.2). Since $\Delta \neq 0$, the result of 1.9.9 tells us that E is a smooth cubic curve. Equipping it with the point $\mathscr{O} = [0:1:0] \in E(K)$ at infinity, we can then apply 1.7.7 to get that the $\oplus$ operation on $E(K)$ defined there makes $E(K)$ into an abelian group with identity element $\mathscr{O}$. Let's summarize this in

Proposition 1.11.1. Let K be a field, $E \subset \mathbb{P}^2_K$ be an elliptic curve with $\mathscr{O} = [0:1:0] \in E(K)$. Then the operation $\oplus : E(K) \times E(K) \to E(K)$ given by $(P, Q) \mapsto P \oplus Q := \mathscr{O} * (P * Q)$ makes $E(K)$ into an abelian group with identity $\mathscr{O}$. If L/K is any field extension, then the inclusion $E(K) \subset E(L)$ makes $E(K)$ a subgroup of $E(L)$.

Henceforth, we will drop the notation $\oplus$ altogether, and simply refer to the group law using the usual $+$. In particular, we will write $P + Q$ for what we called $P \oplus Q$, and nP for $P + \cdots + P$, where the sum has n terms, for any $n \in \mathbb{Z}_{\geq 0}$. Finally, since an admissible change of coordinates is a linear transformation, it preserves the group law $+$, showing

Proposition 1.11.2. Let K be a field, and E and E' elliptic curves in Weierstrass form over K. An admissible change of coordinates transforming E to E' yields a group isomorphism $E(K) \to E'(K)$, or indeed $E(L) \to E(L')$ for any field extension L/K.

We will use the notation $E \cong_K E'$ to express that two curves are related by an admissible change of coordinates over K. (Note that this notion depends on K; see 1.11.9.) The above proposition is just saying that if $E \cong_K E'$, then $E(K) \cong E'(K)$ as abelian groups (sometimes denoted $E(K) \cong_{\mathsf{Ab}} E'(K)$.)

Let's see what the group operation $+$ means concretely for two points $P, Q \in E(K)$ in the affine patch $Z \neq 0$, so points on the affine curve given by $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$. We draw the line L_{PQ} joining P and Q and take its third point of intersection with E, which is $P * Q$. Either $P * Q = \mathscr{O}$ (this happens iff L_{PQ} is vertical), or $P * Q$ lies in this affine patch. If it's the former, then because $\mathscr{O}$ is an inflection point, we get that $P + Q = \mathscr{O}$, i.e., $P = -Q$. If it's the latter, then we draw the line joining $P * Q$ and $\mathscr{O}$, which in this case is just the vertical line through $P * Q$, and take its other intersection point with E; this is $P + Q$. Finally, if $P = Q$, then L_{PQ} should be taken to mean the tangent line $\mathrm{T}_P E$ as usual. The figure 1.8.4 illustrates this procedure for the elliptic curve $E : y^2 = x(x+1)(x+4)$ over $K = \mathbb{Q}$ with the points $P = (-4, 0)$ and $Q = (-2, 2)$, where $P * Q = (2, 6)$ and so $P + Q = (2, -6)$.

1.11.A Automorphism Groups

We saw last time that an admissible change of coordinates over a field K is given by specifying a quadruple $(u; r, s, t)$, where $u \in K^\times$ and $r, s, t \in K$, with the transformation corresponding to the matrix

$$\begin{bmatrix} u^2 & 0 & r \\ u^2 s & u^3 & t \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.11.3}$$

The set of all such transformations forms a subgroup of $\mathrm{PGL}_3(K)$, which as far as I know doesn't have a standard name, so I'm just going to call it $G(K)$.

Exercise 1.11.4. As a set, $G(K) = \{(u; r, s, t) : u \in K^\times, r, s, t \in K\}$. What is the group operation on $G(K)$ corresponding to matrix multiplication when $(u; r, s, t)$ corresponds to the matrix 1.11.3? Fill in the blanks in $(u; r, s, t) \circ (u'; r', s', t') = (uu'; \dots)$.

An admissible change of coordinates takes on elliptic curve E' into another E, but nothing is stopping us from taking $E = E'$, in which case an admissible change of coordinates corresponds to a symmetry of the elliptic curve E. Such an admissible change of coordinates is called an *automorphism* of

E, and the subgroup of $G(K)$ consisting of all such $(u; r, s, t)$ fixing E is called the *automorphism group of E over K*, denoted $\mathrm{Aut}_K(E)$ (or more precisely $\mathrm{Aut}_K(E, \mathscr{O})$, but let's not worry too much about that). This depends on K in general.

Example 1.11.5. Let $K = \mathbb{Q}$ and $E : y^2 = x^3 + 1$. Consider the following admissible changes of coordinates:

(a) the transformation $(x, y) = (x', -y')$ corresponding to $(u; r, s, t) = (-1; 0, 0, 0)$, and
(b) the transformation $(x, y) = (\omega x', y')$ corresponding to $(u; r, s, t) = (\omega^2; 0, 0, 0)$.

Here ω denotes a primitive cube root of unity. The first transformation is defined over $\mathbb{Q}$ and expresses a symmetry $E(\mathbb{Q}) \to E(\mathbb{Q})$, i.e., belongs to $\mathrm{Aut}_{\mathbb{Q}}(E)$. The second transformation is *not* defined over $\mathbb{Q}$, but is defined over $\Omega = \overline{\mathbb{Q}}$ or $\Omega = \mathbb{C}$ (or even the number field $L = \mathbb{Q}[\omega] = \mathbb{Q}[\sqrt{-3}]$), and expresses a symmetry $E(\Omega) \to E(\Omega)$, i.e., belongs to $\mathrm{Aut}_{\Omega}(E)$.

In general we have a diagram of subgroups of the form

$$\begin{array}{ccccc} \mathrm{Aut}_K(E) & \hookrightarrow & G(K) & \hookrightarrow & \mathrm{PGL}_3(K) \\ \downarrow & & \downarrow & & \downarrow \\ \mathrm{Aut}_{\overline{K}}(E) & \hookrightarrow & G(\overline{K}) & \hookrightarrow & \mathrm{PGL}_3(\overline{K}). \end{array}$$

The group $\mathrm{Aut}_{\overline{K}}(E)$ is sometimes called the *geometric* automorphism group of the elliptic curve E, but we won't use this terminology much. You are invited to explore some more automorphism groups (and geometric automorphism groups) of elliptic curves over some finite fields in 2.3.3.

1.11.B Admissible Changes For Curves in Short Weierstrass Form

Suppose now that the field K has $\mathrm{ch}\, K \neq 2$, and the elliptic curves E and E' are in the short Weierstrass form

$$E : y^2 = x^3 + ax^2 + bx + c \text{ and } E' : (y')^2 = (x')^3 + a'(x')^2 + b'x' + c'.$$

Suppose we have an admissible change of coordinates $(x, y) = (u^2x' + r, u^3y' + u^2sx' + t)$ transforming E into E'. Then a simple inspection shows that we must in fact have $s = t = 0$. In other words, any admissible change of coordinates taking a curve in short Weierstrass form to another of the same form must look like $(x, y) = (u^2x' + r, u^3y')$. That's a much easier form to work with.

Similarly, suppose now that $\mathrm{ch}\, K \neq 2, 3$, and that $a = a' = 0$ as well. Then a very similar argument to the one above tells us that in fact $r = 0$ as well, i.e., the transformation looks like $(x, y) = (u^2x', u^3y')$. Therefore, if for $b, c \in K$ such that $4b^3 + 27c^2 \neq 0$, we let $E_{b,c}$ denote the elliptic curve

$$E_{b,c} : y^2 = x^3 + bx + c,$$

then for any allowable b, c, b', c', we have that $E_{b,c} \cong_K E_{b',c'}$ iff there is a $u \in K^\times$ such that $(b, c) = (u^4b', u^6c')$.

Example 1.11.6. Let $K = \mathbb{C}$ (or indeed any field of characterstic not 2 or 3). The elliptic curves $E : y^2 = x^3 + x$ and $E' : y^2 = x^3 + 1$ over K are not isomorphic (over K, or indeed any field containing K).

This leaves the question of whether two curves can be non-isomorphic over K but isomorphic over some bigger field containing K. We'll answer this question in the next section (1.11.9).

Remark 1.11.7. When $\mathrm{ch}\, K = 3$, a careful version of the above discussion implies that whether or not $a = 0$ in the short Weierstrass form is a (boolean) invariant of the curve, i.e., does not depend on the choice of coordinates. We will revisit this idea soon.

1.11.C The j-Invariant and the Isomorphism Classification of Elliptic Curves

The previous discussion suggests that the quantity b^3/c^2 is an invariant of the isomorphism type of $E_{b,c}$. Actually, this is slightly imprecise since c could be zero without b being zero, causing this "invariant" to

be undefined. Is there a linear combination of b^3 and c^2 we know is never going to vanish, and which we could use as the denominator instead? Of course there is, and it is $\Delta = -16(4b^3 + 27c^2)$. This leads us to

Definition 1.11.8 (The j-Invariant)**.** Let K be a field of characteristic other than 2 or 3. For any $b, c \in K$ with $\Delta = -16(4b^3 + 27c^2) \neq 0$, we let the j-invariant of the curve $E_{b,c}$ be

$$j(E_{b,c}) := \frac{(-48b)^3}{\Delta}.$$

At this point, we have shown that if $b, c, b', c' \in K$ are such that $\Delta\Delta' \neq 0$, and if $E_{b,c} \cong_K E_{b',c'}$, then $j(E_{b,c}) = j(E_{b',c'})$. Does the converse hold? Well, suppose that $j(E_{b,c}) = j(E_{b',c'})$; we're trying to show that there is a $u \in K^\times$ such that $(b, c) = (u^4 b', u^6 c')$. Let's split into a few cases.

(a) First suppose that $bc \neq 0$. Then $j(E_{b,c}) = j(E_{b',c'})$ forces $b'c' \neq 0$ as well (check!). From the equality $b^3/c^2 = (b')^3/(c')^2$, it follows that if let $v := (b'c)/(bc') \in K^\times$, then $(b, c) = (v^2 b', v^3 c')$. Now if there is a square root u of v in K, then this u works and we are done. Therefore, we are certainly done when we can take square roots of elements in K, which happens for instance, if K is algebraically closed. In general, however, we may not be able to do this, and this would give us an example of two curves that are isomorphic over $\overline{K}$ but not K. Such curves are called *twists* of each other. (In this special case of $bc \neq 0$, they are called *quadratic twists*.) For an explicit example, see 1.11.9(a).

(b) Next, suppose that $b \neq 0$ but $c = 0$, which happens iff $j(E_{b,c}) = 1728$ (check!). Then $j(E_{b',c'}) = j(E_{b,c}) = 1728$ forces $b' \neq 0$ and $c = 0$. Now $E_{b,0} \cong_K E_{b',0}$ iff there is a $u \in K^\times$ such that $b = u^4 b'$. Again, if a fourth root of b/b' exists in K (e.g., K algebraically closed), we are done; else we get curves which are *quartic twists* of each other (in general)–see 1.11.9(b).

(c) Finally, suppose that $b = 0$ but $c \neq 0$, which happens iff $j(E_{b,c}) = 0$ (clear!). Then $j(E_{b',c'}) = j(E_{b,c})$ forces $b' = 0$ and $c' \neq 0$. Now $E_{0,c} \cong_K E_{0,c'}$ iff there is a $u \in K^\times$ such that $c = u^6 c'$. Again, if a sixth root of c/c' exists in K (e.g., K algebraically closed), we are done; else, we get curves which are *sextic twists* of each other (in general)–see 1.11.9(c).

Example 1.11.9 (Twists)**.** Let $K = \mathbb{Q}$ for simplicity.

(a) (Quadratic Twists) The elliptic curves $E : y^2 = x^3 + x + 1$ and $E' : (y')^2 = (x')^3 + 4x' + 8$ are quadratic twists of each other: they are isomorphic over $\overline{K} = \overline{\mathbb{Q}}$ (or indeed over the smaller subfield $L = \mathbb{Q}[\sqrt{2}]$) by the change of coordinates $(x, y) = (2^{-1}x', 2^{-3/2}y')$, but they are *not* isomorphic over $\mathbb{Q}$, since there does not exist a $u \in \mathbb{Q}^\times$ such that $(1, 1) = (4u^4, 8u^6)$. Note that both curves E and E' have the same j-invariant, namely $j(E) = j(E') = 2^8 3^3 31^{-1}$.

(b) (Quartic Twists) The elliptic curves $E_i : y^2 = x^3 + 2^i x$ for $i = 0, 1, 2, 3$ are pairwise isomorphic over $\overline{K} = \overline{\mathbb{Q}}$ (or indeed over $L = \mathbb{Q}[2^{1/4}]$), but they are pairwise non-isomorphic over $K = \mathbb{Q}$. Note that all the curves E_i have the same j-invariant $j(E_i) = 1728$. For each $i = 0, 1, 2, 3$ (considered cyclically), the curves E_i and E_{i+1} are quartic twists of each other. (Note that the curve $E_4 : y^2 = x^3 + 16x$ *is* isomorphic to $E_0 : y^2 = x^3 + x$ over $K = \mathbb{Q}$! The curves $E_0 : y^2 = x^3 + x$ and $E_2 : y^2 = x^3 + 4x$ are technically *quadratic* twists of each other, since they are isomorphic over $L = \mathbb{Q}[\sqrt{2}]$ (check!); in general, the degree of a twist is the smallest degree of the field extension of K over which the resulting curves are truly isomorphic.)

(c) (Sextic Twists) The elliptic curves $E_i : y^2 = x^3 + 2^i$ for $i = 0, 1, \ldots, 5$ are pairwise isomorphic over $\overline{K} = \overline{\mathbb{Q}}$ (or indeed over $L = \mathbb{Q}[2^{1/6}]$), but they are pairwise non-isomorphic over $K = \mathbb{Q}$. Note that all the curves E_i have the same j-invariant $j(E_i) = 0$. For each $i = 0, 1, \ldots, 5$ (considered cyclically), the curves E_i and E_{i+1} are sextic twists of each other. Again, E_0 and E_6 are isomorphic over $\mathbb{Q}$, the curves E_0 and E_3 are technically quadratic twists of each other, and E_0, E_2, E_4 and E_1, E_3, E_5 are triples of *cubic twists*.

We have proven

Theorem 1.11.10. Let K be a field with $\operatorname{ch} K \neq 2, 3$. Let $b, c, b', c' \in K$ be such that $\Delta\Delta' \neq 0$, and let $E_{b,c}$ and $E_{b',c'}$ denote the corresponding elliptic curves. If $E_{b,c} \cong_K E_{b',c'}$, then $j(E_{b,c}) = j(E_{b',c'})$. The converse holds if K is algebraically closed, but not in general (1.11.9).

This is really cool: one number determines whether or not two curves are isomorphic! Suppose I hand you two elliptic curves E and E' over $\mathbb{C}$, and ask you to tell me whether they are isomorphic over $\mathbb{C}$, all you have to do is compute their j-invariants and check if they agree. If they do, these are isomorphic; if they don't, they aren't. This basically gives us a complete understanding of isomorphism classes of elliptic curves over algebraically closed fields of characteristic other than 2 or 3. (It still leaves the question of exactly which j-invariants are possible; this I leave for an exercise on the Exercise Sheet for next week.)

While we are here, we might as well prove

Theorem 1.11.11. Let K be a field with $\operatorname{ch} K \neq 2, 3$, and let $\overline{K}$ denote an algebraic closure of K. Let E be an elliptic curve over K. The size of the geometric automorphism group of E is given by

$$\#\operatorname{Aut}_{\overline{K}}(E) = \begin{cases} 2 & j \neq 0, 1728, \\ 4 & j = 1728, \\ 6 & j = 0. \end{cases}$$

Proof. Since any curve E can be put[5] into the shortest Weierstrass form $y^2 = x^3 + bx + c$, and this transformation only changes the (geometric) automorphism group up to isomorphism, we might as well assume that E is in this form. A geometric automorphism then corresponds to a $u \in \overline{K}^\times$ such that $(b, c) = (u^4 b, u^6 c)$. Let's split into the same cases as before:

(a) If $bc \neq 0$, which happens iff $j \neq 0, 1728$, then such a u must satisfy $u^4 = u^6 = 1$ and hence $u^2 = 1$, which implies $u \in \{\pm 1\}$. This gives us precisely two automorphisms $(x, y) \mapsto (x, \pm y)$.
(b) If $b \neq 0$ but $c = 0$, which happens iff $j = 1728$, then such a u must satisfy $u^4 = 1$, which gives us four automorphisms corresponding to $u = \pm 1, \pm i$.
(c) Finally, if $b = 0$ and $c \neq 0$, which happens iff $j = 0$, then such a u must satisfy $u^6 = 1$, which gives us six automorphisms corresponding to $u = \pm\omega^k$ for $k = 0, 1, 2$, where ω is a primitive cube root of unity.

■

Note that the statement of 1.11.11 really needs geometric automorphism groups to be true.

Example 1.11.12. Let $K = \mathbb{Q}$. The curve $E : y^2 = x^3 + x$ has $j(E) = 1728$ and geometric automorphism group $\operatorname{Aut}_{\overline{K}}(E)$ of size 4. In fact, these geometric automorphisms are given explicitly by $(x, y) \mapsto (x, \pm y)$ and $(x, y) \mapsto (-x, \pm iy)$. Of these, only the first two are defined over $\mathbb{Q}$, and hence $\operatorname{Aut}_{\mathbb{Q}}(E) \cong \mathbb{Z}/2$ of size 2.

In fact, a slight strengthening of the above argument (left to the readers) shows easily that the geometric automorphism groups when $j = 1728$ and $j = 0$ are necessarily isomorphic to $\mathbb{Z}/4$ and $\mathbb{Z}/6$.

Exercise 1.11.13. Check this!

1.11.D The Case of Characteristics 2 and 3

So far, we have stuck with the case of characteristics other than 2 or 3. The story in these characteristics is very similar. In general, we have

Definition 1.11.14. Let K be any field (of any characteristic!), and let $a_1, \ldots, a_6 \in K$ be as usual, such that $\Delta \neq 0$, so that the usual equation (1.9.10) defines an elliptic curve E over K. In this case, we define the quantity c_4 associated to the elliptic curve E by

$$c_4 := a_1^4 + 8a_1^2 a_2 - 24 a_1 a_3 + 16 a_2^2 - 48 a_4,$$

[5] Technically, we haven't even defined the j-invariant for curves not in the shortest Weierstrass form. We will rectify this below (1.11.14); see also 1.11.16. For now, you can take this to be a statement about curves in the shortest Weierstrass form.

and the j-invariant of E to be

$$j(E) := \frac{c_4^3}{\Delta}.$$

Note how this specializes to the definition 1.11.8 when $a_1 = a_2 = a_3 = 0$. Note also that c_4 is homogeneous of degree 4 in the a_i with the usual weights, and actually similarly to Δ (1.10.4), transforms very cleanly under admissible changes of coordinates: namely via

$$u^4 c_4' = c_4. \tag{1.11.15}$$

(Again, don't take this on faith–you can check it for yourself!) These two equations (1.11.15 and 1.10.4) combined imply that $j(E)$ is actually invariant under admissible changes of coordinates, proving one half of

Theorem 1.11.16. Let K be any field (of any characteristic). Let $a_i, a_i' \in K$ for $i = 1, 2, 3, 4, 6$ be such that $\Delta\Delta' \neq 0$, and let E and E' denote the corresponding elliptic curves. If $E \cong_K E'$, then $j(E) = j(E')$. The converse holds if K is algebraically closed, but not in general.

To prove this theorem, it only remains to show that if $j(E) = j(E')$, then there is an admissible change of coordinates over $\overline{K}$ taking E to E'. The proof of this statement is very similar to what we did above in $\operatorname{ch} K \neq 2, 3$, and again when the characteristic is 2 or 3, you have to treat the cases $j = 0 = 1728$ and $j \neq 0$ separately. In characteristic 2, you also see *octic twists* showing up. I won't bother to type out the details; the interested readers can either supply the proof themselves, or can they can find it in [8, Proposition III.1.4(b)]. See also [14] for more on twisting in characteristics 2 and 3. Finally, analogously to 1.11.11, in $\operatorname{ch} K = 2, 3$ a curve with $j \neq 0$ still has geometric automorphism group $\mathbb{Z}/2$; when $\operatorname{ch} K = 3$ and $j = 0 = 1728$, you get the geometric automorphism group Dic_3 of size 12, and when $\operatorname{ch} K = 2$ and $j = 0 = 1728$, you get the geometric automorphism group $\mathrm{SL}_2(\mathbb{F}_3) \cong Q_8 \rtimes C_3$ of size 24; see [8, Theorem III.10.1 and Exercise A.1]. All I want to emphasize is that there is nothing particularly deep going on here, and the dedicated reader(s) could work these ideas out for themselves.

1.12 26/07/02 - The Origin of The Name "Elliptic Curves" and Torsion Points

I want to start today by addressing the pressing question of

1.12.A Why are elliptic curves called "elliptic" when they are not ellipses?

The story starts from the attempts by geometers going back to the 18^{th} century to find a formula for the perimeter of an ellipse. For this, suppose that we are given $a, b \in \mathbb{R}$ with $a > b > 0$, and let $C_{a,b} : (x/a)^2 + (y/b)^2 = 1$ be the ellipse with semi-major axis a and semi-minor axis b. We can parametrize the ellipse via $(x(t), y(t)) = (a\cos t, b\sin t)$, and this allows us to express the perimeter P of $C_{a,b}$ via the integral

$$P = 4\int_0^{\pi/2} \sqrt{x'(t)^2 + y'(t)^2}\,\mathrm{d}t = 4\int_0^{\pi/2} \sqrt{a^2\sin^2 t + b^2\cos^2 t}\,\mathrm{d}t.$$

Pulling out the a and expressing the resulting integral in terms of the eccentricity $e := \sqrt{1-(b/a)^2} \in [0,1)$ gives us the expression

$$P = 4a\int_0^{\pi/2} \sqrt{1 - e^2\sin^2 t}\,\mathrm{d}t.$$

As a sanity check, if $e = 0$, then this is just $P = 2\pi a$. In general, the trigonometric substitution $u = \sin t$ lets us rewrite this as

$$P = 4a\int_0^1 \sqrt{\frac{1-e^2u^2}{1-u^2}}\mathrm{d}u = 4a\int_0^1 \frac{1-e^2u^2}{\sqrt{(1-u^2)(1-e^2u^2)}}\mathrm{d}u.$$

This is the integral people could not figure out how to evaluate (where by evaluation we mean finding a formula in terms of elementary functions or constants). It turns out that this is for good reason–it's essentially because there is no expression for the antiderivative in terms of elementary functions. Proving this would take us very far afield, but the basic reason behind that is that if you look at the curve

$$C : v^2 = (1-u^2)(1-e^2u^2),$$

then P amounts to evaluating the integral

$$\int_\gamma \frac{1-e^2u^2}{v}\mathrm{d}u$$

for a suitable path γ on C. The point is that this curve C (or more precisely its smooth compactification), is a kind of object we have met already–it is a smooth projective geometrically integral genus one curve with a rational point, i.e., an elliptic curve in the sense of 1.9.1 (see 1.9.5(b)). Indeed, you don't need any fancy machinery to check that the change of coordinates

$$(x, y) = \left(\frac{\beta}{u-1}, \frac{\beta^2 v}{(u-1)^2}\right)$$

for $\beta = 1/(2(e^2-1))$ transforms C into a curve of the form

$$E_0 : y^2 = f(x)$$

for a suitable monic cubic $f(x) \in \mathbb{R}[x]$ with distinct roots (2.1.12), i.e., an elliptic curve. In fact, we can go one step further and put E_0 into shortest Weierstrass form to get a model of the form

$$E : y^2 = (x - 6(1+e^2))(x + 3(1+6e+e^2))(x + 3(1-6e+e^2)).$$

As you can check, this last curve E has discriminant $\Delta = 2^8 3^{12} e^2(1-e^2)^4$, and hence *for* $e \in (0,1)$ defines an elliptic curve. The reason there isn't an expression of P in terms of elementary functions is ultimately the fact that E has genus one (1.8.17). This is the reason people got interested in curves of the form $y^2 = f(x)$ for cubic $f(x)$, and this is why such curves are called *elliptic* curves–the name comes from *elliptic integrals*, which are so because such integrals arise naturally when you try to solve for the perimeter for an ellipse. See [8, Chapter VI] for more on this connection, and also Challenge Exercise 2.6.19.

1.12.B Torsion in Abelian Groups

Let's quickly review the terminology surrounding torsion.

Definition 1.12.1. Let A be an abelian group, with group operation written additively.

(a) An element $a \in A$ is said to be a *torsion element* if there is an $n \in \mathbb{Z}_{\geq 1}$ such that $n \cdot a = 0$ in A; the least such n is called the *order* of a, and this case a is said to be an n-torsion element.
(b) For each $n \in \mathbb{Z}_{\geq 1}$, the subgroup $A[n] := \{a \in A : na = 0\}$ of elements of order dividing n is called the *n-torsion subgroup* of A. Similarly, the subset $\mathrm{Tors}(A) = \bigcup_{n \in \mathbb{Z}_{\geq 1}} A[n] \subset A$ of all torsion elements is a subgroup of A, and is called the *torsion subgroup* of A.

In the above terminology, the subset $A[n] \subset A$ is a subgroup, for instance, because it is the kernel of the multiplication-by-n homomorphism $[n] : A \to A$, although this is also easy to check directly. Note also that there are many variations of the above basic subgroups. For instance, for any prime p, we can define the subgroup $A[p^\infty] := \bigcup_{m \geq 1} A[p^m] \subset A$ consisting of all elements annihilated by some power of p; this is known as the *p-power-torsion* or *p-primary* subgroup of A.

Example 1.12.2.

(a) We have $\mathrm{Tors}\,\mathbb{Z} = \{0\}$.
(b) If A is a finite group, then $\mathrm{Tors}\,A = A$.
(c) If $A = \mathbb{Q}/\mathbb{Z}$, then also $\mathrm{Tors}\,A = A$.
(d) If $A = \mathbb{Z} \oplus \mathbb{Z}/3$, then $\mathrm{Tors}\,A = 0 \oplus \mathbb{Z}/3 \cong \mathbb{Z}/3$. In fact, we have $A[m] = \{0\}$ if $3 \nmid m$, and $A[m] = \mathrm{Tors}\,A \cong \mathbb{Z}/3$ if $3 \mid m$.

The goal for us now is to study torsion points on elliptic curves. Why might we want to do that? If we look back at Bachet's example from the first lecture (1.1.6), we see that Bachet's procedure is given exactly by sending $P \mapsto -2P$ in the group law on the elliptic curve $y^2 = x^3 + c$. Therefore, the sequence $\{P_0, P_1, P_2, P_3, \dots\}$ produced from a seed point $P = P_0$ is precisely $\{P, -2P, 4P, -8P, \dots.\}$. When does this sequence contain only finitely many distinct terms?

Exercise 1.12.3. Let A be an abelian group, and let $a \in A$. Show that the sequence of elements $\{a, -2a, 4a, -8a, \dots\}$, i.e., given by $a_k = (-2)^k a$ for $k \in \mathbb{Z}_{\geq 0}$, contains only finitely many distinct elements iff a is a torsion element.

Therefore, studying when Bachet-like procedures give rise to infinitely many points is exactly equivalent to studying torsion points on elliptic curves. If we have a good understanding of which points are torsion, and which are not, then we can use this as a tool towards studying rational points on curves. For instance, as soon as we have produced one non-torsion point, we know that the corresponding Diophantine equation has infinitely many rational solutions, a fact which might be hard to verify otherwise.

1.12.C Two Torsion on Elliptic Curves and Supersingularity

Let K be a field, and let $\overline{K}$ be an alegbraic closure of K. Let

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

be an elliptic curve over K (1.9.2). Since $\mathscr{O} \in E(K)$ is an inflection point, we know from §1.7.B that to find the negative of a $P \in E(K) \smallsetminus \{\mathscr{O}\}$ say $P = (x_0, y_0)$, we just do $-P = \mathscr{O} * P$. For this, we have to draw the line $L_{\mathscr{O}P}$ joining $\mathscr{O}$ and P, and take its third intersection point with E. Since $\mathscr{O} = [0 : 1 : 0]$ is the point at infinity in the vertical direction and P lies on the affine chart $Z \neq 0$, the line $L_{\mathscr{O}P}$ is just the vertical line through P which has the (affine) equation $x = x_0$. Substituting this into the equation for E and using that $P \in E(K)$ gives us the quadratic equation for y of the form

$$y^2 + a_1x_0y + a_3y = y_0^2 + a_1x_0y_0 + a_3y_0,$$

which factors as

$$(y - y_0)(y + y_0 + a_1x_0 + a_3) = 0.$$

One root is the expected $y = y_0$ corresponding to P, and the other root gives us the point

$$-P = (x_0, -y_0 - a_1x_0 - a_3). \tag{1.12.4}$$

Note that this formula is valid in all characteristics, including characteristic 2. However, if $\operatorname{ch} K \neq 2$ and we use the short Weierstrass form with $a_1 = a_3 = 0$, then the formula becomes the much easier $-P = (x_0, -y_0)$: you negate the point by negating the y-coordinate.

Now suppose we are trying to study the 2-torsion points on E, i.e., the subgroup $E[2](K) = E(K)[2]$. Note that a point $P \in E(K)$ satisfies $2P = \mathscr{O}$ iff $P = -P$. If $P \neq \mathscr{O}$, then we can write $P = (x_0, y_0)$; in this case, the condition $P = -P$ can be written via 1.12.4 as the single condition

$$2y_0 + a_1x_0 + a_3 = 0. \tag{1.12.5}$$

As you can see from this, things proceed differently from here depending on whether $\operatorname{ch} K = 2$.

Case 1. Suppose first that $\operatorname{ch} K \neq 2$. Then if $P = (x_0, y_0) \in E[2](K)$, we get an expression for y_0 in terms of x_0 as

$$y_0 = -\frac{a_1x_0 + a_3}{2}.$$

We can plug this expression for y back into the Weierstrass equation to get a cubic equation for x, the roots of which correspond precisely to the 2-torsion points on E. The fact that $\Delta \neq 0$ actually implies that the resulting cubic polynomial is separable, i.e., has no repeated roots over an algebraic closure. This is not hard to check (and left as an exercise to the reader!), but let's not bother with it in the general case, and just focus on the case of short Weierstrass equations in which $a_1 = a_3 = 0$. In this case, if $P = (x_0, y_0) \in E[2](K)$, then in fact $y_0 = 0$ and so x_0 is a root of $f(x) = x^3 + a_2x^4 + a_4x + a_6 = x^3 + ax^2 + bx + c$. We have seen in §1.10.A that nonvanishing of Δ is equivalent to the fact that $f(x)$ has three distinct roots over an algebraic closure. However, if we are only looking for roots in K, then exactly one of the following three things happens:

Case 1.a. The polynomial $f(x)$ is irreducible, and so there are no 2-torsion K-points other than $\mathscr{O}$, i.e., $E[2](K) = \{\mathscr{O}\}$. This happens, for instance, if $K = \mathbb{Q}$ and $f(x) = x^3 + 3x + 1$ (c.f. 2.2.5).

Case 1.b. The polynomial $f(x)$ factors into a linear and an irreducible quadratic factor. In this case, there is precisely one nontrivial 2-torsion K-point $T = (x_0, 0)$, where $x_0 \in K$ satisfies $f(x_0) = 0$. In this case, we have $E[2](K) = \{\mathscr{O}, T\}$, with group structure $\mathbb{Z}/2$. This happens, for instance, if $K = \mathbb{Q}$ and $f(x) = x^3 + 1 = (x + 1)(x^2 - x + 1)$.

Case 1.c. The polynomial $f(x)$ splits completely over K. In this case, there are three nontrivial 2-torsion K-points, say T_1, T_2, T_3; here, for $i = 1, 2, 3$ we have $T_i = (x_i, 0)$, where the x_i are the roots of $f(x)$. In this case, $E[2](K) = \{\mathscr{O}, T_1, T_2, T_3\}$ is a group of size 4 in which every nontrivial element has order 2, and this forces $E[2](K)$ to be isomorphic to $\mathbb{Z}/2 \oplus \mathbb{Z}/2$. This happens, for instance, if $K = \mathbb{Q}$ and $f(x) = x(x + 1)(x + 4)$.

Of course, only Case 1.c occurs if K is algebraically closed (of characteristic not two). Therefore, in all subcases of Case 1, we have that $E[2](\overline{K}) \cong \mathbb{Z}/2 \oplus \mathbb{Z}/2$.

Case 2. Finally, suppose that $\operatorname{ch} K = 2$. Here the defining equation for 2-torsion just because $a_1x_0 + a_3 = 0$, and we see that things behave very differently depending on whether or not $a_1 = 0$.

Case 2.a. Suppose that $a_1 \neq 0$. Then the only allowable x-coordinate for a 2-torsion point on E is $x_0 = -a_3/a_1 = a_3/a_1$. Plugging this back into the equation for the elliptic curve tells us that the y-coordinate must satisfy

$$y_0^2 = \frac{a_3^3 + a_1a_2a_3^2 + a_1^2a_3a_4 + a_1^3a_6}{a_1^3}. \tag{1.12.6}$$

If we're looking for roots over an algebraically closed field (or even perfect field!), then then this equation *has a unique* solution; this is because if one solution is $y_0 = \alpha$, then the equation 1.12.6 factors as $(y_0 - \alpha)^2 = 0$, since we're in characteristic two. However, if we are only looking for roots in K, then again exactly one of the following two things happens:

Case 2.a.i The unique root y_0 of 1.12.6 lies in K. In this case, there is precisely one nontrivial 2-torsion K-point $T = (-a_1^{-1}a_3, y_0)$, so $E[2](K) = \{\mathscr{O}, T\}$ with group structure $\mathbb{Z}/2$. This happens, for instance if $K = \mathbb{F}_2$ and $E : y^2 + xy = x^3 + 1$, in which $E[2](\mathbb{F}_2) = \{\mathscr{O}, (0, 1)\}$.

Case 2.a.ii The unique root y_0 of 1.12.6 in $\overline{K}$ *does not* lie in K. In this case, there are no 2-torsion K-points other than $\mathscr{O}$, i.e., $E[2](K) = \{\mathscr{O}\}$. In this case, the "remaining" 2-torsion point is not visible over K, but only over a purely inseparable extension of K. This happens, for instance, if $K = \mathbb{F}_2(t)$ and $E : y^2 + xy = x^3 + t$; the point $T = (0, t^{1/2}) \in E[2](\overline{K}) \smallsetminus \{\mathscr{O}\}$ is the other two-torsion point visible over $L = \mathbb{F}_2(t^{1/2})$, and so $E[2](L) = [2](\overline{K}) = \{\mathscr{O}, T\}$, which is isomorphic to $\mathbb{Z}/2$.

Again, if K is a perfect field (e.g., if K is finite, or if K is algebraically closed of characteristic two), then only Case 2.a.i occurs.

Case 2.b. Suppose finally that $a_1 = 0$. In this case, from the smoothness $\Delta \neq 0$ and the expression for Δ in characteristic two (1.9.13), we see immediately that $a_3 \neq 0$. In this case, there is *no solution* to $a_3 = 0$, even if K is algebraically closed. When this happens, there are no 2-torsion K-points other than $\mathscr{O}$, and $E[2](K) = E[2](\overline{K}) = \{\mathscr{O}\}$. This happens, for instance, when $K = \mathbb{F}_2$ and $E : y^2 + y = x^3$.

These six cases cover all the possibilities for what 2-torsion K-points on an elliptic curve E over a field K can look like. We summarize our discussion in

Theorem 1.12.7 (2-Torsion on Elliptic Curves)**.** Let K be a field, $\overline{K}$ an algebraic closure of K, and E/K an elliptic curve.

(a) If $\operatorname{ch} K \neq 2$, then $E[2](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2$. The subgroup $E[2](K) \subset E[2](\overline{K})$ is either trivial, $\mathbb{Z}/2$, or $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, and all possibilities occur in general (e.g., over $K = \mathbb{Q}$).
(b) If $\operatorname{ch} K = 2$, then $E[2](\overline{K})$ is either isomorphic to $\mathbb{Z}/2$, which happens iff $a_1 \neq 0$, or trivial, which happens iff $a_1 = 0$. In the former case, $E[2](K) \subset E[2](\overline{K})$ is either trivial or $\mathbb{Z}/2$, and only the latter can happen if K is perfect (e.g., finite or algebraically closed).

In characteristic 2, the cases $a_1 \neq 0$ and $a_1 = 0$ are thus genuinely different (this was foreshadowed in 1.10.3). These have different names. Indeed, we have proven the essential content of

Lemma/Definition 1.12.8 (Supersingularity in Characteristic Two)**.** Let K be a field of characteristic two. An elliptic curve E/K is said to be *supersingular* if the following equivalent conditions hold:

(a) In some Weierstrass equation for E, we have $a_1 = 0$.
(b) In *every* Weierstrass equation for E, we have $a_1 = 0$ (c.f. 1.10.3).
(c) For every field L/K, we have that $E[2](L) = \{\mathscr{O}\}$.
(d) If $\overline{K}$ is an algebraic closure of K, then $E[2](\overline{K}) = \{\mathscr{O}\}$.

The elliptic curve E is said to be *ordinary* otherwise.

Therefore, 1.12.7 says that if $\operatorname{ch} K = 2$ and $\overline{K}/K$ as above, then an elliptic curve E/K is ordinary iff $E[2](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/2$ and supersingular iff $E[2](\overline{K}) \cong_{\mathsf{Ab}} \{0\}$.

Example 1.12.9. Let $K = \mathbb{F}_2$.

(a) If $E : y^2 + xy = x^3 + 1$, then $E(\mathbb{F}_2) = \{\mathscr{O}, (0,1), (1,0), (1,1)\}$. This is a group of size 4 with only one 2-torsion point, and hence $E(\mathbb{F}_2) \cong_{\mathsf{Ab}} \mathbb{Z}/4$, with generators $(1,0)$ and $(1,1)$. The curve E is ordinary and $j(E) = 1$.
(b) If $E : y^2 + y = x^3$, then $E(\mathbb{F}_2) = \{\mathscr{O}, (0,0), (0,1)\}$. This is a group of size 3, and hence $E(\mathbb{F}_2) \cong_{\mathsf{Ab}} \mathbb{Z}/3$. The curve E is supersingular and $j(E) = 0$.

Example 1.12.10. Let $K = \mathbb{F}_2(t)$ and let $E : y^2 + xy = x^3 + t$, which has $j(E) = 1/t$. We are now in Case 2.a.ii above, and therefore have $E[2](K) = \{\mathscr{O}\}$. This might suggest that E is supersingular, but this is in fact false: we have that $E[2](L) = \{\mathscr{O}, T\}$ where $L = \mathbb{F}_2(t^{1/2})$ and $T = (0, t^{1/2})$. Indeed, the curve E is ordinary, and we just don't see the nontrivial 2-torsion point over K. This is why it is important to ask for $E[2](L) = \{\mathscr{O}\}$ for *every* field L/K in the definition of supersingularity in 1.12.8.

Unimportant Remark 1.12.11. Note that when $\operatorname{ch} K = 2$, then the formula for c_4 in 1.11.14 simplifies drastically to $c_4 = a_1^4$, and similarly the j-invariant is simply $j(E) = a_1^{12}/\Delta$. This tells us that in characteristic two, supersingularity is equivalent to the vanishing of the j-invariant. This is true also in characteristics 3 and 5, but the reason I want to de-emphasize this is because this should be thought of

as an accident: this equivalence is not true for general $p > 0$. We will define supersingularity for a field of any positive characteristic p below, and 1.12.31 gives an example of supersingular elliptic curve $E/\mathbb{F}_7$ with $j(E) = -1$.

There are many other equivalent definitions of supersingularity, and we will meet some of them on the next Exercise Sheet; see also [7, §IV.4] and [8, Theorem V.3.1].

Unimportant Remark 1.12.12. Note that "supersingularity" of elliptic curves is really unfortunate terminology, since an elliptic curve is always smooth and hence "nonsingular", i.e., doesn't have any singular points. (This is partly why I like to avoid the term "nonsingular" altogether.) The clashing nomenclature comes from the unrelated facts that singular points on curves behave strangely and different than other (i.e., smooth) points and that supersingular elliptic curves are very strange (super = very, singular = strange), even among elliptic curves in positive characteristic.

So far, we have avoided developing explicit formulae for the group law on an elliptic curve. However, we do really need at least a general form of these formulae when discussing three-torsion, so we might as well bite the bullet now.

1.12.D Explicit Formulae for the Group Law

Let K be any field, and let $E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ be an elliptic curve over K (1.9.2). For $i = 1, 2, 3$, let $P_i = (x_i, y_i) \in E(K) \smallsetminus \{\mathscr{O}\}$ be points. Suppose that $P_1 + P_2 = P_3$; in particular, we are assuming that $P_1 + P_2 \neq \mathscr{O}$. In this case, we can easily find an expression relating the coordinates of the P_i.

Indeed, let $L_{P_1P_2}$ denote the line joining P_1 and P_2. (The case $P_1 = P_2$ is allowed, in which case $L_{P_1P_2} = \mathrm{T}_{P_1}E$ is the tangent line to E at P_1 as usual.) The hypothesis $P_1 + P_2 \neq \mathscr{O}$ tells us that $L_{P_1P_2}$ has an equation of the form

$$L_{P_1P_2} : y = \lambda x + \nu,$$

where

$$\lambda := \begin{cases} \dfrac{y_2 - y_1}{x_2 - x_1}, & \text{if } x_1 \neq x_2, \\ \dfrac{3x^2 + 2a_2x + a_4 - a_1y}{2y + a_1x + a_3}, & \text{if } x_1 = x_2, \end{cases}$$

and where

$$\nu := y_1 - \lambda x_1 = y_2 - \lambda x_2.$$

To intersect $L_{P_1P_2}$ with E, we plug in the equation for y into the equation for E to get a cubic equation in x of the form

$$x^3 - (\lambda^2 + a_1\lambda - a_2)x^2 + \cdots = 0.$$

The three roots of this equation are $x = x_1, x_2, x_3$, and so these are related by

$$x_1 + x_2 + x_3 = \lambda^2 + a_1\lambda - a_2.$$

In particular, if we know x_1 and x_2, we can recover x_3 as $x_3 = \lambda^2 + a_1\lambda - a_2 - x_1 - x_2$. The y-coordinate of the third intersection point of $L_{P_1P_2}$ with E is the $\lambda x_3 + \nu$; to get y_3, we then simply apply the negation formula 1.12.4 to get $y_3 = -(\lambda x_3 + \nu) - a_1x_3 - a_3$. In this way, we have found an explicit formula for adding two points on an elliptic curve: the formula is

$$(x_3, y_3) = (\lambda^2 + a_1\lambda - a_2 - x_1 - x_2, -(\lambda x_3 + \nu) - a_1x_3 - a_3),$$

where λ, ν are as above.

In particular, we can take $P_1 = P_2$ to get (after a little algebraic simplification),

Proposition 1.12.13 (The Duplication Formula)**.** Let K be a field, $E : y^2 + a_1xy + a_3x = x^3 + a_2x^2 + a_4x + a_6$ be an elliptic curve over K, and let $P = (x, y) \in E(K)$. If $2P \neq \mathscr{O}$, then the x-coordinate

of $2P$ can be expressed in terms of that of P as

$$x(2P) = \frac{x^4 - b_4x^2 - 2b_6x - b_8}{4x^3 + b_2x^2 + 2b_4x + b_6},$$

where the polynomials b_i are defined by

$$\begin{aligned} b_2 &= a_1^2 + 4a_2, \\ b_4 &= 2a_4 + a_1a_3, \\ b_6 &= a_3^2 + 4a_6, \text{ and} \\ b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2. \end{aligned}$$

Let me emphasize that there is nothing mysterious going on here; this is just algebra. For instance, it is clear from the formula for λ when $x_1 = x_2$ that the denominator of x_3 will look like $(2y + a_1x + a_3)^2$. We can expand this out to get

$$4(y^2 + a_1xy + a_3y) + a_1^2x^2 + 2a_1a_3 + a_3^2.$$

Since P lies on the elliptic curve, we can replace the quantity $y^2 + a_1xy + a_3y$ by $x^3 + a_2x^2 + a_4x + a_6$, and the polynomials b_k essentially come from giving names to the resulting quantities. Similarly, it is clear from the general form of the equation that the numerator will be a quartic polynomial in x and y (of "total weight 8"), and to make it into a polynomial in x only, you again replace every occurence of $y^2 + a_1xy + a_3y$ by $x^3 + a_2x^2 + a_4x + a_6$.

Remark 1.12.14.

(a) Let me remark here, for future use, the polynomial equalities

$$\Delta = -b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \tag{1.12.15}$$

and

$$4b_8 = b_2b_6 - b_4^2, \tag{1.12.16}$$

both of which come from straightforward algebraic checks.

(b) If $\operatorname{ch} K \neq 2$ and we work with the shorter Weierstrass form with $a_1 = a_3 = 0$ and use the labels $(a_2, a_4, a_6) = (a, b, c)$, so the curve E has the form $y^2 = f(x) = x^3 + ax^2 + bx + c$, then the duplication formula can be expressed more succinctly as

$$x(2P) = \frac{f'(x)^2}{4f(x)} - a - 2x = \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4f(x)}. \tag{1.12.17}$$

1.12.E Three Torsion on Elliptic Curves

Let's now study three-torsion on elliptic curves. There are two perspectives we could take here: the geometric perspective, and the algebraic perspective. For the geometric perspective, note that the group law on an elliptic curve can be summarized by the

Observation 1.12.18. Let K be a field, and $E \subset \mathbb{P}^2_K$ be an elliptic curve in Weierstrass form. Then for $P, Q, R \in E(K)$, we have $P + Q + R = \mathscr{O}$ in the group law of $E(K)$ iff P, Q, R are collinear, i.e., there is a line $L \subset \mathbb{P}^2_K$ such that the intersection points of E and L are precisely P, Q, R (counted with multiplicities as usual).

In particular, applying this to $P = Q = R$ immediately gives us that $3P = \mathscr{O}$ in the group law iff P is an inflection point of E. In other words, the 3-torsion points on E are precisely the inflection points on E. This is really cool relationship between the algebraic and geometric perspectives. If we have a good way of getting all the inflection points on E, then we have a handle on 3-torsion, and vice-versa. When $\operatorname{ch} K \neq 2$, there is in fact a good way of getting a handle on all inflection points; this proceeds by calculating something known as the Hessian divisor of the curve. I will not tell this story here, but you can find it, for instance, in [6, Problem 5.23]. See also 1.12.27.

The algebraic perspective is much more direct and mechanical. Suppose that $P = (x_0, y_0) \in E(K) \smallsetminus \{\mathscr{O}\}$ and that $3P = \mathscr{O}$. Then $2P = -P \neq \mathscr{O}$, and so

$$x(2P) = x(-P) = x(P).$$

Conversely, if $x(2P) = x(P)$, then the vertical line $x = x(P)$ intersects E in the points $\mathscr{O}, P$ and $2P$. This means that $2P = \pm P$, but $2P = P$ can't be true since $P \neq \mathscr{O}$, and so we must have $2P = -P$, i.e., $3P = \mathscr{O}$. We summarize this in

Proposition 1.12.19. Let K be a field (of any characteristic), and let $E \subset \mathbb{P}^2_K$ be an elliptic curve in Weierstrass form. Let $P \in E(K) \smallsetminus \{\mathscr{O}\}$. The following conditions are equivalent:

(a) We have $3P = \mathscr{O}$ in the group $E(K)$.
(b) The point P is an inflection point on E.
(c) We have that $x(2P) = x(P)$.

Now, the duplication formula 1.12.13 gives us an easy way to check when $x(2P) = x(P)$. Indeed, simply cross-multiplying and subtracting, we see that $x(2P) = x(P)$ is equivalent to saying that $x = x(P)$ is a root of the polynomial

$$f_3(x) := 3x^4 + b_2x^3 + 3b_4x^2 + 3b_6x + b_8. \tag{1.12.20}$$

The polynomial $f_3(x)$, whose roots correspond to the x-coordinates of the three-torsion points on E, is called the *third division polynomial* of E (the more general division polynomials $f_n(x)$ for $n \in \mathbb{Z}_{\geq 0}$ are defined in Exercise 2.2.10). To study how many roots $f_3(x)$ has, let's first see if it has any repeated roots, which can be checked by computing its discrimimant:

$$\operatorname{disc}(f_3) = -81\Delta^2.$$

You don't have to–and shouldn't–take my word on faith; you can just compute the discriminant yourself, or get a computer algebra software to do it for you. Again, different things happen when $\operatorname{ch} K \neq 3$ and when $\operatorname{ch} K = 3$.

Case 1. Suppose first that $\operatorname{ch} K \neq 3$. Then smoothness $\Delta \neq 0$ implies that $\operatorname{disc}(f_3) \neq 0$. This means that over an algebraic closure $\overline{K}$, the polynomial f_3 factors completely as

$$f_3(x) = 3\prod_{i=1}^{4}(x - \beta_i)$$

for four distinct $\beta_i \in \overline{K}$. For each $i = 1, \ldots, 4$, we can plug $x = \beta_i$ into the equation for E to get the corresponding quadratic equation

$$y^2 + (a_1\beta_i + a_3)y - (\beta_i^3 + a_2\beta_i^2 + a_4\beta_i + a_6) = 0. \tag{1.12.21}$$

Can this quadratic equation 1.12.21 for y have a repeated root? Well, if it does, then then this repeated root y_0 is also a root of its derivative

$$2y + (a_1\beta_i + a_3) = 0.$$

But this was precisely our condition (1.12.5) for (β_i, y_0) to be a 2-torsion point! Since we're assuming it's a 3-torsion point, it *cannot* be 2-torsion (since it is not $\mathscr{O}$), and this tells us that y_0 is not a repeated root of the quadratic equation 1.12.21. Therefore, for each $i = 1, \ldots, 4$, we get two different roots $y = \delta_i^{\pm}$ of this equation, giving us the 8 nontrivial 3-torsion points $(\beta_i, \delta_i^{\pm})$ for $i = 1, \ldots, 4$. In particular, $\#E[3](\overline{K}) = 9$. Up to isomorphism, there are two abelian groups of size 9, and only one of them has the property that every nontrivial element has order 3. This is sufficient to conclude that $E[3](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/3 \oplus \mathbb{Z}/3$.

Case 2. Finally, suppose that $\operatorname{ch} K = 3$. In this case, $f_3(x)$ becomes

$$f_3(x) = b_2x^3 + b_8.$$

Therefore, we see that things behave very differently depending on whether or not $b_2 = 0$.

Case 2.a. Suppose that $b_2 \neq 0$. Then the only allowable x-coordinate for a three-torsion point on E is $x = (-b_2^{-1}b_8)^{1/3} \in \overline{K}$. (Why is this? Well, if β solves $b_2\beta^3 + b_8 = 0$, then in the equation $b_2x^3+b_8$ actually factors completely as $b_2(x-\beta)^3 = 0$, since we are in characteristic 3.) Again, plugging this in above gives us a quadratic equation for y, which doesn't have repeated roots for the same reason as in Case 1. We conclude that in this case, $E[3](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/3$. (Again, we don't need K to be algebraically closed; we only need it to be perfect for this to be true. In particular, something like $K = \mathbb{F}_3$ works just fine.)

Case 2.b. Suppose finally that $b_2 = 0$. In this case, the formula 1.12.15 when $\operatorname{ch} K = 3$ simplifies drastically to $\Delta = b_4^3$, so smoothness (which corresponds to $\Delta \neq 0$ by 1.9.9)implies that $b_4 \neq 0$. But now the relation (1.12.16) along with $b_2 = 0$ forces $b_8 = -b_4^2$, from which we conclude that $b_8 \neq 0$. In particular, the equation $f_3(x) = 0$ has no roots whatsoever, and we conclude that $E[3](\overline{K}) = \{\mathscr{O}\}$.

Note that all of these possibilities occur; we invite the reader to come up with examples illustrating this in Exercise 2.4.6(a).

Remark 1.12.22. Suppose that $\operatorname{ch} K \neq 2, 3$, and we are working with the short Weierstrass form with $a_1 = a_3 = 0$. Then either by using the formula 1.12.17 or by plugging in $a_1 = a_3 = 0$ into 1.12.20, we get the formula

$$f_3(x) = 3x^4 + 4ax^3 + 6bx^2 + 12cx + 4ac - b^2 = 2f(x)f'(x) - f'(x)^2,$$

where $E : y^2 = f(x)$ with $f(x) = x^3 + ax^2 + bx + c$. To check if $f_3(x)$ has any repeated roots, we need to check if it shares any roots with its derivative

$$f_3'(x) = 2f(x)f'''(x) = 12f(x).$$

But any common root of f_3 and f_3', would then be a common root of f and f', which does not exist by smoothness (1.10.A). This is again enough to conclude that the polynomial $f_3(x)$ is separable, i.e., has distinct roots over an algebraic closure $\overline{K}$. This is basically the argument given in [2, Theorem 2.1]. The only disadvantage of this approach is that it does not work when $\operatorname{ch} K = 2$, whereas our proof above does.

We have proven

Theorem 1.12.23 (3-Torsion on Elliptic Curves)**.** Let K be a field, $\overline{K}$ an algebraic closure of K, and E/K an elliptic curve.

(a) If $\operatorname{ch} K \neq 3$, then $E[3](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/3 \oplus \mathbb{Z}/3$.
(b) If $\operatorname{ch} K = 3$, then $E[3](\overline{K})$ is either isomorphic to $\mathbb{Z}/3$, which happens iff $b_2 \neq 0$, or trivial, which happens iff $b_2 = 0$.

An analysis similar to 1.12.7 (but more complicated) can be done to figure out what the subgroup $E[3](K) \subset E[3](\overline{K})$ looks like in general; this is an exercise on the Exercise Sheet for next week (2.4.6). Next, as in 1.12.7, we see from the proof that if K is perfect (e.g., finite) and E has $b_2 \neq 0$, then in fact $E[2](K) \cong_{\mathsf{Ab}} \mathbb{Z}/3$; we don't need to go to $\overline{K}$. Finally, as in 1.12.8, we have proven the essential content of

Lemma/Definition 1.12.24 (Supersingularity in Characteristic Three)**.** Let K be a field of characteristic three. An elliptic curve E/K is said to be *supersingular* if the following equivalent conditions hold:

(a) In some Weierstrass equation for E, we have $b_2 = 0$.
(b) In *every* Weierstrass equation for E, we have $b_2 = 0$.
(c) For every field L/K, we have that $E[3](L) = \{\mathscr{O}\}$.
(d) If $\overline{K}$ is an algebraic closure of K, then $E[3](\overline{K}) = \{\mathscr{O}\}$.

The elliptic curve E is said to be *ordinary* otherwise.

Note that when $\operatorname{ch} K = 3$ and we're in short Weierstrass form with $a_1 = a_3 = 0$, then $b_2 = a_2$. In other words, a curve of the form $y^2 = x^3 + ax^2 + bx + c$ in characteristic 3 is supersingular iff $a = 0$ (c.f. 1.11.7).

Example 1.12.25. Let $K = \mathbb{F}_3$.

(a) If $E : y^2 = x^3 + x^2 + 1$, then $\Delta(E) = j(E) = -1$. It is easy to see that

$$E(\mathbb{F}_3) = \{\mathscr{O}, (0, \pm 1), (-1, \pm 1), (1, 0)\},$$

so that $\#E(\mathbb{F}_3) = 6$. Up to isomorphism, there is a unique abelian group of order six, and so we must have $E(\mathbb{F}_3) \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/3 \cong_{\mathsf{Ab}} \mathbb{Z}/6$. The points $(0, \pm 1)$ are easily seen to have order six, and hence must be the two generators. Note that $x^3 + x^2 + 1 = (x - 1)(x^2 - x - 1) \in \mathbb{F}_3[x]$; therefore, we are in Case 1.b of §1.12.C. Correspondingly,

$$E[2](\mathbb{F}_3) = \{\mathscr{O}, (1, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2,$$

whereas

$$E[2](\overline{\mathbb{F}}_3) = \{\mathscr{O}, (1, 0), (-1 \pm i, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2.$$

Finally, we have that

$$E[3](\mathbb{F}_3) = E[3](\overline{\mathbb{F}}_3) = \{\mathscr{O}, (-1, \pm 1)\}.$$

The curve E is therefore ordinary.

(b) If $E : y^2 = x^3 + x$, then $\Delta(E) = -1$ and $j(E) = 0$. It is easy to see that

$$E(\mathbb{F}_3) = \{\mathscr{O}, (0, 0), (-1, \pm 1)\},$$

so that $\#E(\mathbb{F}_3) = 4$. Since $x^3 + x = x(x^2 + 1) \in \mathbb{F}_3[x]$, we are in Case 1.b of §1.12.C. Correspondingly,

$$E[2](\mathbb{F}_3) = \{\mathscr{O}, (0, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2.$$

This is enough to conclude that $E(\mathbb{F}_3) \cong_{\mathsf{Ab}} \mathbb{Z}/4$, and also

$$E[2](\overline{\mathbb{F}}_3) = \{\mathscr{O}, (0, 0), (\pm i, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2.$$

Finally, we see that

$$E[3](\mathbb{F}_3) = E[3](\overline{\mathbb{F}}_3) = \{\mathscr{O}\},$$

i.e., the curve E is supersingular.

Unimportant Remark 1.12.26. Note that when $\operatorname{ch} K = 3$, the formula for c_4 in 1.11.14 simplifies drastically to $c_4 = b_2^2$ (check!), and similarly the j-invariant is simply $j(E) = b_2^6/\Delta$. This tells us that in characteristic three, supersingularity is again equivalent to the vanishing of the j-invariant. See 1.12.11.

Unimportant Remark 1.12.27. (For those who know about the Hessian divisor on plane curves.) If $F_3 = 3X^4 + b_2X^3Z + 3b_4X^2Z^2 + 3b_6XZ^3 + b_8Z^4$ is the homogenization of the third division polynomial, then the Hessian H of the curve $E = \mathbb{V}(F)$, where

$$F = Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3,$$

satisfies the very simple identity

$$HZ = (2b_2Z + 24X)F + 8F_3.$$

It follows from this formula (check!) that $\mathbb{V}(H)$ always contains $\mathscr{O} = [0 : 1 : 0]$, and further in characteristic not two, the ideal (F, H) contains F_3, telling us again that all the 3-torsion points other than $\mathscr{O}$ must have x-coordinates which are roots of $f_3(x) = F_3(x, 1)$. When $\operatorname{ch} K = 3$, the Hessian satisfies

$$H = -b_2F - (b_2X^3 + b_8Z^3).$$

In this case, the Hessian divisor on E is the vanishing locus of

$$b_2X^3 + b_8Z^3.$$

As before, if $b_2 \neq 0$ (i.e., when E is not supersingular), and we let $\beta := (-b_2^{-1}b_8)^{1/3} \in \overline{K}$, then

$$b_2X^3 + b_8Z^3 = b_2(X - \beta Z)^3,$$

which tells us that the reason for the existence of only 3 inflection points is because the Hessian divisor is nonreduced and has multiplicity 3 at each of the three inflection points (over $\overline{K}$). Finally, if $b_2 = 0$ (i.e., if E *is* supersingular), then $b_8 \neq 0$ as above, and the Hessian divisor is the vanishing locus of Z^3, which is just $9(\mathscr{O})$; in this case, the Hessian divisor is *even more* nonreduced, and is supported only at $\mathscr{O}$ with multiplicity 9, explaining the existence of a unique inflection point on E. Note finally also that when $\operatorname{ch} K = 2$, we have $H \equiv 0$, corresponding to the fact that the vanishing of the Hessian only gives us inflection points when $\operatorname{ch} K \neq 2$.

1.12.F Supersingularity in General

At this point, based on what we have seen already, the following results shouldn't be too surprising.

Theorem 1.12.28. Let K be a field of characterstic zero, and $\overline{K}$ an algebraic closure of K. Then for any elliptic curve E/K and $n \in \mathbb{Z}_{\geq 1}$, we have that

$$E[n](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/n \oplus \mathbb{Z}/n.$$

Theorem 1.12.29. Let K be a field of characteristic $p > 0$, and $\overline{K}$ be an algebraic closure of K. Then for any given elliptic curve E/K, exactly one of the following holds:

(a) For all $n \in \mathbb{Z}_{\geq 1}$, if we write $n = p^r m$ for $r, m \in \mathbb{Z}_{\geq 0}$ with $p \nmid m$, then

$$E[n](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/m \oplus \mathbb{Z}/m \oplus \mathbb{Z}/p^r.$$

(b) For all $n \in \mathbb{Z}_{\geq 1}$, if we write $n = p^r m$ for $r, m \in \mathbb{Z}_{\geq 0}$ with $p \nmid m$, then

$$E[n](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/m \oplus \mathbb{Z}/m.$$

Curves satisfying (a) are known as *ordinary*, and curves satisfying (b) are called *supersingular*.

Note that in either case, if $\operatorname{ch} K \nmid n$, then we always have $E[n](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/n \oplus \mathbb{Z}/n$; the difference comes when trying to analyze p-power torsion. For instance, in the ordinary case for any $r \geq 0$ we have $E[p^r](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/p^r$, whereas in the supersingular case, we have $E[p^r](\overline{K}) \cong_{\mathsf{Ab}} \{0\}$. We won't prove these results here, although one proof of a special case can be found in (the double-starred exercise) 2.2.10. The general algebraic proof of these results can be given by studying the *invariant differential* on elliptic curves; material along these lines can be found in [8]. However, the *real* reason for why we should expect results of this sort (at least 1.12.28) to be true (I think) comes from the picture over $K = \mathbb{C}$, which we will see next time. We end with a couple of examples.

Example 1.12.30. Let $K = \mathbb{F}_5$, and consider the elliptic curve $E : y^2 = x^3 - x$. This has $\Delta(E) = -1$ and $j(E) = -2$. An easy check shows that

$$E(\mathbb{F}_4) = \{\mathscr{O}, (0,0), (\pm 1, 0), (2, \pm 1), (-2, \pm 2)\}.$$

In particular, $\#E(\mathbb{F}_5) = 8$. Now $x^3 - x = x(x+1)(x-1) \in \mathbb{F}_5[x]$ tells us that we are in Case 1.c of §1.12.C, and hence

$$E[2](\mathbb{F}_5) = \{\mathscr{O}, (0,0), (\pm 1, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2.$$

This is enough to conclude (check!) that $E(\mathbb{F}_5) \cong_{\mathsf{Ab}} \mathbb{Z}/4 \oplus \mathbb{Z}/2$, with the four points $(2, \pm 1)$ and $(-2, \pm 2)$ each having order 4.

Next, we note that $f_3(x) = 3x^4 - 6x^2 - 1 = -2(x^4 - 2x^2 - 2) \in \mathbb{F}_5[x]$ is irreducible, which agrees with out observation that $E[3](\mathbb{F}_5) = \{\mathscr{O}\}$. Now a careful study of the general theory above (left as 2.4.6(c)) guarantees that there is an extension $L/\mathbb{F}_5$ of degree at most 8 such that $E[3](L) \cong \mathbb{Z}/3 \oplus \mathbb{Z}/3$, and indeed one can check that we do really need L to be $\mathbb{F}_{5^8} = \mathbb{F}_{390625}$ here: for $i = 0, 1, \ldots, 7$, we have that $E[3](\mathbb{F}_{5^i}) = \{\mathscr{O}\}$ but in fact $E[3](\mathbb{F}_{390625}) \cong_{\mathsf{Ab}} \mathbb{Z}/3 \oplus \mathbb{Z}/3$.

Finally, we see by Lagrange's Theorem that $\#E[5](\mathbb{F}_5) = \{\mathscr{O}\}$, but this is not enough to conclude that E is supersingular over $\mathbb{F}_5$. Indeed, we have $\#E[5](\mathbb{F}_{625}) \cong_{\mathsf{Ab}} \mathbb{Z}/5$ (check!), which proves that E is ordinary.

Example 1.12.31. Let $K = \mathbb{F}_7$, and consider the elliptic curve $E : y^2 = x^3 - x$. This has $\Delta(E) = 1$ and $j(E) = -1$. An easy check shows that

$$E(\mathbb{F}_7) = \{\mathscr{O}, (0,0), (\pm 1, 0), (-2, \pm 1), (-3, \pm 2)\}.$$

In particular, $\#E(\mathbb{F}_7) = 8$. Now $x^3 - x = x(x+1)(x-1) \in \mathbb{F}_7[x]$ tells us that we are in Case 1.c of §1.12.C, and hence

$$E[2](\mathbb{F}_7) = \{\mathscr{O}, (0,0), (\pm 1, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2.$$

This is enough to conclude as above that in fact $E(\mathbb{F}_7) \cong_{\mathsf{Ab}} \mathbb{Z}/4 \oplus \mathbb{Z}/2$, with the four points $(-2, \pm 1)$ and $(-3, \pm 2)$ each having order 4. Next, we note that $f_3(x) = 3x^4 - 6x^2 - 1 \in \mathbb{F}_7[x]$, which factors as

$$f_3(x) = 3(x^2 + x + 3)(x^2 - x + 3) \in \mathbb{F}_7[x].$$

In particular, $E[3](\mathbb{F}_7) = \{\mathscr{O}\}$ (we knew this already), but over $\mathbb{F}_{2401} \cong \mathbb{F}_7[\alpha] := \mathbb{F}_7[x]/(x^4 - 2x^2 - 3x + 3)$, we have $\#E[3](\mathbb{F}_{2401}) = 9$ with $\#E[3](\mathbb{F}_{2401}) \cong_{\mathsf{Ab}} \mathbb{Z}/3 \oplus \mathbb{Z}/3$ (as you can check with Sage). Finally, it follows from $\#E(\mathbb{F}_7) = 8$ and Lagrange's Theorem that $E[7](\mathbb{F}_7) = \{\mathscr{O}\}$. However, this is again not sufficient to conclude that $E/\mathbb{F}_7$ is supersingular. In fact, it turns out that E *is* supersingular.

There are a few ways of proving this, but none within the reach of the tools in this course. One can argue that if $E/\mathbb{F}_p$ is an elliptic curve, then E is supersingular iff $\#E[p](\mathbb{F}_{p^i}) = 1$ for all $i = 1, \ldots, p-1$; this reduces the question to a finite check which can be performed on Sage. (I couldn't find a reference for this fact, but it just involves studying the possible degrees of a point on the group scheme $E[p]$ which in the ordinary case has order p.) Another way to proceed would be to use [7, Proposition IV.4.21] to reduce to checking the coefficient of x^6y^6 in $(y^2 - x^3 + x)^6 \in \mathbb{F}_7[x, y]$, which happens to be zero–this implies supersingularity (see also 2.4.15(a)). Finally, using [7, IV.4.22, IV.4.23.3], we see that $h_7(\lambda) = \lambda^3 + 9\lambda^2 + 9\lambda + 1$, which does have $\lambda = -1$ as a root; again, this is sufficient to conclude that E is supersingular (see also 2.4.15(b) and 2.4.20).

1.13 26/07/06 - The Complex Picture and Introduction to the Nagell-Lutz Theorem

Last time, we mentioned (without proof) that if K is a field of characteristic $p \geq 0$, and $\overline{K}$ an algebraic closure of K, and E/K an elliptic curve, then for all $n \in \mathbb{Z}_{\geq 1}$ such that $p \nmid n$, we have that $E[n](\overline{K}) \cong_{\mathsf{Ab}} \mathbb{Z}/n \oplus \mathbb{Z}/n$. We also saw proofs of this result when $n = 2, 3$. In particular, this holds for all $n \in \mathbb{Z}_{\geq 1}$ when $\operatorname{ch} K = 0$. Today I want to explain the *real reason* we might expect something like this to be true, over our favorite algebraically closed field of characteristic zero, namely $K = \mathbb{C}$.

1.13.A The Story over $K = \mathbb{C}$

The complex exponential $\exp : \mathbb{C} \to \mathbb{C}^\times$ is *singly periodic* with period $2\pi i$: it has the property that for all $z \in \mathbb{C}$ and $n \in \mathbb{Z}$ we have $\exp(z + n \cdot (2\pi i)) = \exp(z)$. (This is what makes the theory of Fourier series work!) But this is not the only kind of periodicity that a complex function can have. Since the complex numbers are "two-dimensional", we can ask if there are complex functions which have two "independent" periods. For various reasons related to those discussed in §1.12.A, mathematicians in the 19^{th} century (such as Weierstrass) got interested in studying these *doubly periodic* complex functions.

To set up the theory, suppose that $\Lambda \subset \mathbb{C}$ is a lattice, such as $\Lambda = \mathbb{Z}[i]$ or $\Lambda = \mathbb{Z}[\zeta_3]$ or $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau$ for some $\tau \in \mathbb{C}$ with $\operatorname{Im} \tau > 0$. (I will not give a formal definition.) A function $f : \mathbb{C} \to \mathbb{C}$ is doubly periodic with periods in Λ iff for all $z \in \mathbb{C}$ and $\omega \in \Lambda$ we have $f(z + \omega) = f(z)$. For this to be true, it is sufficient to require this happens for two independent periods $\omega_1, \omega_2 \in \Lambda$ that form a $\mathbb{Z}$-basis of Λ. Specifying such a function is the same as specifying a function on the quotient group $\mathbb{C}/\Lambda$, which you should think of as being given by the fundamental parallelogram with opposite sides glued in Pac-Man style. The resulting object is topologically a torus. Now Liouville's Theorem from complex analysis implies that there cannot be any nonconstant holomorphic (i.e., complex-differentiable) doubly periodic functions with respect to Λ, but there can indeed be such functions if we allow some points where the function is not defined (i.e., has a pole); such functions are known as *meromorphic* functions.

One of the "simplest" such meromorphic functions, investigated by Weierstrass, is now called the *Weierstrass $\wp$-function*, defined by

$$\wp(z) := \frac{1}{z^2} + \sum_{\omega \in \Lambda \smallsetminus \{0\}} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right). \tag{1.13.1}$$

This depends on Λ, so is probably more properly written as $\wp_\Lambda$. Note that this function is not defined at any points of Λ, but at least the terms of this series make sense when $z \notin \Lambda$.

Fact 1.13.2. The expression 1.13.1 converges to a holomorphic (i.e., complex differentiable) function on $\mathbb{C} \smallsetminus \Lambda$ that is Λ-periodic: for all $z \in \mathbb{C} \smallsetminus \Lambda$ and $\omega \in \Lambda$ we have $\wp(z + \omega) = \wp(z)$. When extended by $\wp(\omega) = \infty$ for all $\omega \in \Lambda$, the extended function $\wp$ is a Λ-periodic meromorphic function on $\mathbb{C}$, i.e., a meromorphic function on $\mathbb{C}/\Lambda$.

The above fact is not magical; it is just a matter of being careful about convergence and invoking the appropriate results from complex analysis. What is much more magical is

Fact 1.13.3. For any lattice $\Lambda \subset \mathbb{C}$, the infinite series defined by

$$g_2 := 60 \sum_{\omega \in \Lambda \smallsetminus \{0\}} \frac{1}{\omega^4} \text{ and } g_3 = 140 \sum_{\omega \in \Lambda \smallsetminus \{0\}} \frac{1}{\omega^6}$$

converge to complex numbers. Further, the function $\wp = \wp_\Lambda$ satisfies the differential equation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3. \tag{1.13.4}$$

This is very interesting, and looks suspiciously like what we have been doing. Specifically, consider the elliptic curve $E_\Lambda \subset \mathbb{P}^2_{\mathbb{C}}$ defined by the equation $y^2 = 4x^3 - g_2 x - g_3$. (This is not quite in

Weierstrass form, but it's close; also, we have to check that $\operatorname{disc}(4x^3-g_2x-g_3) = 16(g_2^3-27g_3^2) \neq 0$; I won't bother with this.) The magical fact 1.13.4 implies that if $z \in \mathbb{C} \smallsetminus \Lambda$, then the point $(x,y) = (\wp(z), \wp'(z))$ belongs to $E_\Lambda(\mathbb{C})$. We can then "complete" this map by defining $\omega \mapsto \mathscr{O} = [0:1:0]$ for all $\omega \in \Lambda$. The resulting map $\mathbb{C} \to E_\Lambda(\mathbb{C})$ is clearly Λ-periodic because $\wp$ is, and hence descends to a map of sets $\mathbb{C}/\Lambda \to E_\Lambda(\mathbb{C})$. The third and craziest magical fact is

Fact 1.13.5. The map $\mathbb{C}/\Lambda \to E_\Lambda(\mathbb{C})$ given by $[z] \mapsto [\wp_\Lambda(z) : \wp'_\Lambda(z) : 1]$ is an isomorphism of (complex Lie) groups.

In other words, the funny group law on the cubic curve $E_\Lambda(\mathbb{C})$ we defined using the chord and tangent process is just the "same" as the group law on the quotient group $\mathbb{C}/\Lambda$ of "addition modulo Λ". Ultimately, the key ingredient in the proof of 1.13.5 is another magical property of the $\wp = \wp_\Lambda$ function which says that if $u, v, w \in \mathbb{C} \smallsetminus \Lambda$ are such that $u+v+w \in \Lambda$ (i.e., $[u],[v],[w] \in \mathbb{C}/\Lambda$ classes that sum to zero in $\mathbb{C}/\Lambda$), then

$$\begin{vmatrix} \wp(u) & \wp'(u) & 1 \\ \wp(v) & \wp'(v) & 1 \\ \wp(w) & \wp'(w) & 1 \end{vmatrix} = 0.$$

In particular, we see immediately what n-torsion in $E_\Lambda(\mathbb{C})$ looks like: to find all points of order n in $\mathbb{C}/\Lambda$, we just divide the fundamental parallelogram into n^2 smaller ones by drawing n slits in each direction, and take the intersection points of these slits, which are the left-and-bottom vertices of the n^2 parallelograms. In particular, of course there are exactly n^2 of these, and they form a subgroup isomorphic to $\mathbb{Z}/n \oplus \mathbb{Z}/n$. (Some nice illustrations can be found in [2, §2.2].)

Because we're not going to prove these results, we are not going to use them. (The interested reader can find proofs in [8, Chapter VI] or in textbooks on complex analysis, such as [15, Chapter 7].) These results belong to a course on the *analytic* theory of elliptic curves; this course is one on the *algebraic* theory. I only wanted to mention them because they give context for a lot of things happening in this here. These ideas form the beginnings of many very beautiful stories in mathematics connecting elliptic curve theory to complex analysis and Riemann surfaces, and much more.

Unimportant Remark 1.13.6. This line of reasoning can be used to give a(n) (analytic) proof of 1.12.28 for $K = \mathbb{C}$, and then for any field K of characteristic zero by invoking the *Lefschetz principle* ([8, §V.6]).

1.13.B The Nagell-Lutz Theorem

Today, I want to start building towards two major theorems we will prove in this course: the Nagell-Lutz Theorem and the Mordell-Weil Theorem. We'll start with the first of these, with a discussion closely following [2, Chapter 2].

For the rest of this course (except perhaps at the very end), we're going to focus on the case of $K = \mathbb{Q}$, in which case we can work with short Weierstrass equations of the form

$$y^2 = x^3 + ax^2 + bx + c$$

with $a, b, c \in \mathbb{Z}$. Recall that if $f(x) := x^3 + ax^2 + bx + c \in \mathbb{Z}[x]$, then the discriminant

$$D = \operatorname{disc}(f) = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$$

is related to the discriminant Δ of E via

$$\Delta = 16D.$$

The Nagell-Lutz Theorem gives a complete algorithmic procedure (given factoring!) to find all torsion points on such an elliptic curve over $\mathbb{Q}$:

Theorem 1.13.7 (Nagell-Lutz)**.** In the above notation, let $a, b, c \in \mathbb{Z}$ be such that $D \neq 0$, and let $E/\mathbb{Q}$ be the corresponding elliptic curve $E : y^2 = x^3 + ax^2 + bx + c$. If $P = (x_0, y_0) \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then in fact $x_0, y_0 \in \mathbb{Z}$ and either $y_0 = 0$ or $y_0^2 \mid D$.

Proofs of this result were published by the Norwegian mathematician Trygve Nagell (1895-1988) and the French mathematician Élisabeth Lutz (1914-2008) in the 1930s. Lutz's paper came out when she was just 23 years old. Nagell was a student of Alex Thue, famous, among many things, for his work on integer points on elliptic curves, and Lutz was a student of André Weil, famous, among many things, for being one of the founders of modern algebraic geometry in the early 20^{th} century, which he studied extensively in jail after refusing to fight for France in World War II. The Nagell-Lutz Theorem (1.13.7) can in some ways be thought of as being similar to the Rational Root Theorem (1.1.2). Before we give a proof, let's look at a few examples (more to be found on Exercise Sheet 4).

Example 1.13.8. Consider the curve $E : y^2 = x^3 + 4$. Then $D = -2^4 3^3$. A nontrivial rational torsion point (x, y) must have either $y = 0$ (not possible, by 1.1.2) or have $y^2 \mid D$. This forces $y \in \{\pm 1, \pm 2, \pm 3, \pm 4, \pm 6, \pm 12\}$. A simple check shows that the only possibility that allows rational x is $y = \pm 2$ with $x = 0$. This tells us that $\operatorname{Tors} E(\mathbb{Q}) = \{\mathscr{O}, (0, \pm 2)\}$, and this forces $\operatorname{Tors} E(\mathbb{Q}) \cong_{\mathsf{Ab}} \mathbb{Z}/3$. Note that this is really a result about the field $K = \mathbb{Q}$; indeed, we have

$$E[2](\overline{\mathbb{Q}}) = \{\mathscr{O}, (2^{2/3}\omega^i, 0)\}_{i=0,1,2} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2.$$

Example 1.13.9. Consider the curve $E : y^2 = x^3 - 16x + 16$, and the point $P = (0, 4) \in E(\mathbb{Q})$ a point of finite order? Well, we have that

$$2P = (4, 4), \quad 3P = (-4, -4), \quad 4P = (8, -20), \quad 5P = (1, -1), \quad \text{and } 6P = (24, 116).$$

However, we see that

$$7P = \left(-\frac{20}{9}, \frac{172}{27}\right).$$

At this point, we can use 1.13.7 to tell us that $7P \notin \operatorname{Tors} E(\mathbb{Q})$ and hence $P \notin \operatorname{Tors} E(\mathbb{Q})$. In other words, the "tangent" process iterated with the seed $P = P_0$ will necessarily give us infinitely many distinct points on E; in particular, E *has* infinitely many rational points! Note that this example also illustrates the important point that the converse to 1.13.7 is false: we have $D = 2^8 \cdot 37$ and $P = (0, 4)$ has integer coordinates $(x_0, y_0) = (0, 4)$ with $y_0^2 \mid D$, but $P \notin \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$.

Example 1.13.10. Let's revisit the example of 1.2.B. We are interested in studying the curves $E : y^2 = x^3 - 432\alpha^2$ for various α. Let's study two cases:

(a) When $\alpha = 1$, we're looking at the curve $y^2 = x^3 - 432$ corresponding to Fermat's Last Theorem for $n = 3$. This has discriminant $D = -2^8 3^9$. Let's apply 1.13.7. Clearly, there are no solutions with $y = 0$. The condition $y^2 \mid D$ gives us lots of possibilities, but only finitely many. It's easy to check them on a computer, and from doing this we get easily that the only values that work are $y = \pm 36$, which correspond to $x = 12$. Therefore, $\operatorname{Tors} E(\mathbb{Q}) = \{\mathscr{O}, (12, \pm 36)\} \cong_{\mathsf{Ab}} \mathbb{Z}/3$. In particular, to prove Fermat's Last Theorem for $n = 3$, it only remains to show that E has no points of infinite order, which as we shall see below is equivalent to $\operatorname{rank} E = 0$. This is what we'll prove later.

(b) When $\alpha = 2$, we're looking at the curve $y^2 = x^3 - 2^6 3^3$. (You might be interested in studying this curve if you care about cubes of integers in arithmetic progression; this is going to be an exercise on the last sheet.) We know from the theory of admissible changes of coordinates that we may change the coefficient c in the equation by a sixth power without affecting the isomorphism class of the curve, so we might as well work with the curve $E : y^2 = x^3 - 27$ instead. The advantage of working with this curve is that it still has integer Weierstrass coefficients, although much smaller, and so 1.13.7 still applies. There *is* a solution $(3, 0)$ with $y = 0$. For solutions with $y \neq 0$, we have $y^2 \mid D = -3^9$, giving us the possibilities $y = \pm 3^i$ for $i = 0, 1, 2, 3, 4$. Checking these options by hand (or by using a computer), we see that none of these options work, and hence $\operatorname{Tors} E(\mathbb{Q}) = \{\mathscr{O}, (3, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2$. Again, it remains to investigate the rank of E, which we will do later.

We'll prove 1.13.7 next time.

1.14 26/07/07 - The Proof of the Nagell-Lutz Theorem

The goal for today is to prove the Nagell-Lutz Theorem (1.13.7). The proof of 1.13.7 is a consequence of the following two propositions:

Proposition 1.14.1. Let $E : y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$ and $D = 2^{-4}\Delta$. Suppose $P = (x_0, y_0) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ is a point such that $y_0 \neq 0$, i.e., $2P \neq \mathscr{O}$. Suppose further that P and $2P$ both have integer coordinates. Then $y_0^2 \mid D$.

Proposition 1.14.2. Let $E : y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$. If $P = (x_0, y_0) \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then $x_0, y_0 \in \mathbb{Z}$.

Given these two results, let's prove 1.13.7.

Proof of 1.13.7, assuming 1.14.1 and 1.14.2. Suppose $P = (x_0, y_0) \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$. It follows from 1.14.2 that $x_0, y_0 \in \mathbb{Z}$. If $y_0 = 0$, we are done. Else $y_0 \neq 0$, which is equivalent to $2P \neq \mathscr{O}$. Since $2P \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ as well, we conclude from 1.14.2 again that $2P$ has integer coordinates. It then follows from 1.14.1 that $y_0^2 \mid D$. ∎

It remains to prove 1.14.1 and 1.14.2. I think of the first of these (1.14.1) as being algebraic trickery, whereas I think there is much more structure and theory behind the proof of 1.14.2 (although a very short proof of 1.14.2 involving algebraic trickery can be given by using 2.2.10).

1.14.A Proof of 1.14.1

Proof of 1.14.1. Recall the duplication formula 1.12.13 for the coordinates of $x(2P)$; in our case, it takes the form

$$x(2P) = \frac{f'(x)^2}{4f(x)} - a - 2x = \frac{g(x)}{4f(x)}, \tag{1.14.3}$$

where

$$g(x) = f'(x)^2 - 4(a + 2x)f(x) = x^4 - 2bx^2 - 8cx + b^2 - 4ac.$$

From $D = \operatorname{disc}(f) \neq 0$ we know that $f(x)$ and $f'(x)$ don't have any common (complex) roots; from the formula $g(x) = f'(x)^2 - 4(a + 2x)f(x)$, it then follows that $f(x)$ and $g(x)$ don't have any common (complex) roots either. In particular, since $\gcd(f(x), g(x)) = 1$ in the Euclidean domain $\mathbb{Q}[x]$, the Euclidean algorithm can be used to produce polynomials $p(x), q(x) \in \mathbb{Q}[x]$ such that

$$p(x)f(x) + q(x)g(x) = 1.$$

In fact, one can check that we may take $p(x) = D^{-1}P(x)$ and $q(x) = D^{-1}Q(x)$, where $P(x), Q(x) \in \mathbb{Z}[x]$ are the polynomials given by

$$P(x) = 3x^3 - ax^2 - 5bx + 2ab - 27c \text{ and } Q(x) = -3x^2 - 2ax + a^2 - 4b.$$

(Here we are using that $a, b, c \in \mathbb{Z}$.) Clearing denominators, we get an identity of the form

$$P(x) \cdot f(x) + Q(x) \cdot g(x) = D, \tag{1.14.4}$$

where $f(x), g(x), P(x), Q(x) \in \mathbb{Z}[x]$ (check!).

Now suppose we are given a $P = (x_0, y_0) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ such that $y_0 \neq 0$ (which is equivalent to $f(x_0) \neq 0$, which is equivalent to $2P \neq \mathscr{O}$), and suppose that P and $2P$ have integer coordinates. In particular, $x_0, y_0 \in \mathbb{Z}$. From $y_0^2 = f(x_0)$, the integrality of $x(2P)$, and the formula 1.14.3, it follows that $y_0^2 = f(x_0) \mid g(x_0)$. It then follows from the identity 1.14.4 evaluated at $x = x_0$ that $y_0^2 \mid D$ as needed. ∎

Example 1.14.5. Note that 1.14.1 has nothing to do with torsion points. Let's revisit the example $E : y^2 = x^3 - 16x + 16$ of 1.13.9, in which $D = 2^{12} \cdot 37$. If we take $P = (0, 4) \in E(\mathbb{Q})$ as in that example, then we see that the fact that $2P$ has integer coordinates implies that $y(P)^2 \mid D$, which is in fact true. The same applies to the y-coordinates of $2P$ and $3P$, but not to the y-coordinate of $4P$, and indeed $8P = (84/25, -52/125)$.

Remark 1.14.6. Here's another perspective on why we might expect an equation of the sort 1.14.4 to be true. We know from the argument above that the coprimality of $f(x)$ and $g(x)$ implies that there is an identity of the form 1.14.4 for some integer polynomials $P(x), Q(x) \in \mathbb{Z}[x]$ and *some* integer $D' \in \mathbb{Z}_{\geq 1}$ instead of D; for instance, we can take D' to be the resultant $\mathrm{Res}_x(f(x), g(x))$. Now if p is a prime with $p \nmid 2D$, then the same equation $y^2 = f(x)$ also defines an elliptic curve over E over p, and the same argument as above–this time over $\mathbb{F}_p$ instead of $\mathbb{Q}$–allows us to conclude that $\overline{f}(x)$ and $\overline{g}(x)$ have no common complex roots over $\overline{\mathbb{F}}_p$, and this is sufficient to conclude that $p \nmid \mathrm{Res}_x(f(x), g(x))$, since the resultant behaves well under reduction mod p. Therefore, the only primes dividing D' can be the primes dividing D (and possibly 2).

Unimportant Remark 1.14.7. The above computation may tempt you to believe that the resultant of $f(x)$ and $g(x)$ with respect to x is actually D. This is incorrect: the resultant (as you can check!) is $\mathrm{Res}_x(f(x), g(x) = D^2$. Therefore, it is somehow stranger than usual that we get the stronger result 1.14.4. (This is remarked also in a footnote to [2, Exercise 2.11].)

1.14.B Proof of 1.14.2

To do this, we need to study more carefully the denominators that appear in rational points on elliptic curves. The first thing we need to study is what denominators even look like. If you have done the exercises (2.1.9), then the following result will not come as a surprise at all.

Proposition 1.14.8. Let $E/\mathbb{Q}$ be an elliptic curve in short Weierstrass form $E : y^2 = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{Z}$. If $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then x and y when expressed as fractions in lowest terms look like

$$(x, y) = \left(\frac{m}{e^2}, \frac{n}{e^3}\right),$$

where $m, n, e \in \mathbb{Z}$ with $e \geq 1$ and $\gcd(m, e) = \gcd(n, e) = 1$.

Proof. For any prime p, the equation

$$y^2 = x^3 + ax^2 + bx + c$$

along with $a, b, c \in \mathbb{Z}$ implies that $v_p(x) < 0$ iff $v_p(y) < 0$, in which case $2v_p(y) = 3v_p(x)$. This–along with unique factorization in $\mathbb{Z}$!–is enough to prove the claim (check!). ∎

Exercise 1.14.9. Convince yourself that the above proof works.

Remark 1.14.10. Note that 1.14.8 shows in particular, that if $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then $x \in \mathbb{Z}$ iff $y \in \mathbb{Z}$. But we knew this already: this follows directly from the Rational Root Theorem (1.1.2).

The above proof shows that it is perhaps wise to look at one prime at a time. Further experimentation with the curves in the exercises would seem to show that if p is a fixed prime and $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ is such that $v_p(x) < 0$, then $v_p(x(kP)) < 0$ for all $k \geq 1$. This is indeed correct. Let's now formalize this intuition, by defining the subset of points of $E(\mathbb{Q})$ with p "in the denominators". While we're doing this, we might as well keep track of how many times p goes in the denominators. For this reason, for an elliptic curve $E/\mathbb{Q}$ in Weierstrass form, prime p, and integer $r \in \mathbb{Z}_{\geq 1}$, we consider the subset

$$E_{p,r}(\mathbb{Q}) := \{(x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\} : v_p(x) \leq -2r\} \cup \{\mathscr{O}\}.$$

Note that by virtue of 1.14.8, the condition $v_p(x) \leq -2r$ may equivalently be stated as $v_p(y) \leq -3r$. If we write $(x, y) = (me^{-2}, ne^{-3})$ as above (1.14.8), then this condition is further equivalent to $p^r \mid e$.

Remark 1.14.11. Note also that if either $x = 0$ or $y = 0$, then $P \notin E_{p,r}(\mathbb{Q})$, by the convention that $v_p(0) = +\infty$. This is the convention we will stick to throughout.

In this way, we get for each prime p a chain of subsets

$$E(\mathbb{Q}) \supset E_{p,1}(\mathbb{Q}) \supset E_{p,2}(\mathbb{Q}) \supset \cdots \supset \{\mathscr{O}\}.$$

This is known as the *p-adic filtration* on $E(\mathbb{Q})$. Note finally also that for each p we have

$$\bigcap_{r \geq 1} E_{p,r}(\mathbb{Q}) = \{\mathscr{O}\}.$$

The first order of business then, is to show

Proposition 1.14.12. Let $E/\mathbb{Q}$ be an elliptic curve in short Weierstrass form $E : y^2 = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{Z}$. Then for each prime p and integer $r \in \mathbb{Z}_{\geq 1}$, the subset $E_{p,r}(\mathbb{Q}) \subset E(\mathbb{Q})$ is a subgroup.

I like to think of 1.14.12 as saying that the denominators are "persistant": if a prime p shows up in the denominator of (the x- and y-coordinates of) P, then it will do so also for $2P, 3P, \ldots$.

Unimportant Remark 1.14.13. If you know something about the p-adic numbers $\mathbb{Q}_p$, then you should think of the $E_{p,r}(\mathbb{Q})$ as the subgroup where the coordinates get very large p-adically. In other words, the $E_{p,r}(\mathbb{Q})$ form (the intersection with $E(\mathbb{Q})$ of) an open neighborhood basis of $\mathscr{O}$ consisting of subgroups of $E(\mathbb{Q}_p)$. We will not use or need this language, but it is helpful to keep in mind for when you later see $\mathfrak{p}$-adic filtrations more generally for $\mathfrak{p}$ a prime of a general number (or even global) field K.

To prove 1.14.12, it is helpful to have some general criterion to figure out when a subset of an abelian group is a subgroup.

Lemma 1.14.14. Let A be an abelian group. Let $A' \subset A$ be a subset such that

(i) we have $0 \in A'$,
(ii) if $a \in A'$ then also $-a \in A'$, and
(iii) if $a_1, a_2, a_3 \in A$ are such that $a_1 + a_2 + a_3 = 0$, and two of the three a_i are in A', then so is the third.

Then $A' \subset A$ is a subgroup.

Exercise 1.14.15. Prove 1.14.14. Is the generalization to a statement of the form "$n-1$ out of n implies all" true for all n?

The following proof has been adapted from the one found in [2, §2.4].

Proof of 1.14.12. Suppose we are given E, p, and r. We use the criterion of 1.14.14(a). I will leave checking conditions (i) and (ii) to the reader. Suppose we are now given points $P_1, P_2, P_3 \in E(\mathbb{Q})$ such that $P_1 + P_2 + P_3 = 0$, and such that $P_1, P_2 \in E_{p,r}(\mathbb{Q})$; we have to show that $P_3 \in E_{p,r}(\mathbb{Q})$ as well. Again, I will leave the case where one (or more) of the P_i is $\mathscr{O}$, or when $P_1 = P_3$ or $P_2 = P_3$ to the reader, and focus on the case where all three are not $\mathscr{O}$ and $P_3 \notin \{P_1, P_2\}$. Recall by 1.12.18 that $P_1 + P_2 + P_3 = 0$ is equivalent to saying that P_1, P_2, P_3 are collinear (counted with suitable multiplicities!); therefore, we have to show that if $P_1, P_2, P_3 \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ are collinear, with $P_3 \notin \{P_1, P_2\}$ and $P_1, P_2 \in E_{p,r}(\mathbb{Q})$, then also $P_3 \in E_{p,r}(\mathbb{Q})$.

At this point, the proof is a calculation best done in (t, s) coordinates as in 1.3.11, so that $(t, s) = (x/y, 1/y)$, and the curve E in these coordinates is

$$s = t^3 + at^2 s + bts^2 + cs^3.$$

Note that this affine patch $Y \neq 0$ does not contain precisely the nontrivial 2-torsion points on $E(\mathbb{Q})$. Note also that the condition of lying in $E_{p,r}(\mathbb{Q})$ when in this patch is totally equivalent to $v_p(t) \geq r \Leftrightarrow$

$v_p(s) \geq 3r$. This automatically includes the point $\mathscr{O} = (0,0)$ when we follow the convention $v_p(0) = +\infty$, as in 1.14.11. Suppose first that P_1, P_2, P_3 are all in this patch; we'll worry about any remaining edge cases at the very end. Therefore, suppose that $P_i = (t_i, s_i)$ with $t_i, s_i \in \mathbb{Q}$ for $i = 1, 2, 3$, and that these points are all collinear; in other words, there is a line ℓ such that $(\ell \cap E)(\mathbb{Q}) = \{P_1, P_2, P_3\}$, counted with multiplicites.

If ℓ is vertical, say given by $t = t_0$ for some $t_0 \in \mathbb{Q}$, then $t_0 = t_1 = t_2 = t_3$. In this case, $P_1 \in \ell(\mathbb{Q}) \cap E_{p,r}(\mathbb{Q})$ implies that $v_p(t_1) \geq r$, and this implies that $v_p(t_3) \geq r$ as needed. Having dealt with this easy case, we may now suppose that ℓ is not vertical; say it has equation

$$\ell : s = \lambda t + \nu$$

for some $\lambda, \nu \in \mathbb{Q}$. We need to figure out an expression for λ. If $t_1 \neq t_2$, then

$$\lambda = \frac{s_2 - s_1}{t_2 - t_1}.$$

Using the fact that

$$s_i = t_i^3 + at_i^2 s_i + bt_i s_i^2 + cs_i^3$$

for $i = 1, 2$, we conclude that

$$s_2 - s_1 = (t_2^3 - t_1^3) + a(t_2^2 s_2 - t_1^2 s_1) + b(t_2 s_2^2 - t_1 s_1^2) + c(s_2^3 - s_1^3).$$

Using that

$$t_2^2 s_2 - t_1^2 s_1 = (t_2^2 - t_1^2)s_2 + t_1^2(s_2 - s_1)$$

and

$$t_2 s_2^2 - t_1 s_1^2 = (t_2 - t_1)s_2^2 + t_1(s_2^2 - s_1^2),$$

we see that

$$(s_2 - s_1)\left(1 - at_1^2 - bt_1(s_1 + s_2) - c(s_1^2 + s_1 s_2 + s_1^2)\right) = (t_2 - t_1)\left((t_1^2 + t_1 t_2 + t_2^2) + as_2(t_1 + t_2) + s_2^2\right),$$

giving us an expression for λ of the form

$$\lambda = \frac{s_2 - s_1}{t_2 - t_1} = \frac{(t_1^2 + t_1 t_2 + t_2^2) + as_2(t_1 + t_2) + bs_2^2}{1 - at_1^2 - bt_1(s_1 + s_2) - c(s_1^2 + s_1 s_2 + s_2^2)}. \tag{1.14.16}$$

Wait–this expression is only valid if the denominator on the right hand side of this formula is nonzero. Is it? Well, except for the 1, every single quantity has positive p-adic valuation. This forces v_p of the denominator on the right to be 0; in particular, it forces the denominator to be nonzero. Finally, what happens if $t_1 = t_2$? Well, the real formula for λ should then involve derivatives, and this approach gives us

$$\lambda = \frac{3t_1^2 + 2at_1 s_2 + bs_1^2}{1 - at_1^2 - 2bt_1 s_1 - 3cs_1^2},$$

but this is of course the same as 1.14.16 when $(t_1, s_1) = (t_2, s_2)$. (Indeed, this is how people took derivatives in the olden days before calculus!) Thus, 1.14.16 is correct in all cases. Now since $P_1, P_2 \in E_{p,r}(\mathbb{Q})$, we conclude from 1.14.16 that $v_p(\lambda) \geq 2r$. From this and $\nu = s_1 - \lambda t_1$, we get that $v_p(\nu) \geq 3r$.

Let's now look at how the three t_i relate to each other. Of course, these t_i are the roots of the cubic equation

$$\lambda t + \nu = t^3 + at^2(\lambda t + \nu) + bt(\lambda t + \nu)^2 + c(\lambda t + \nu)^3,$$

which can be rearranged to be in the form

$$(1 + a\lambda + b\lambda^2 + c\lambda^3)\, t^3 + \nu(a + 2b\lambda + 3c\lambda^2)t^2 + (\cdots)\, t + (\cdots) = 0,$$

where the $(\cdots)$ is stand-in for some coefficient we don't care about. Arguing as above, we note that $v_p(1 + a\lambda + b\lambda^2 + c\lambda^3) = 0$; in particular, this leading coefficient is nonzero. This implies by Vieta's formulae that this equation *does* genuinely have three roots t_i which are related by

$$t_1 + t_2 + t_3 = -\frac{\nu(a + 2b\lambda + 3c\lambda^2)}{1 + a\lambda + b\lambda^2 + c\lambda^3}. \tag{1.14.17}$$

This equation (1.14.17) is key. Note that since the denominator on the right side has $v_p = 0$, we conclude that the right hand side has p-adic valuation at least $v_p(\nu) \geq 3r \geq r$. Since t_1 and t_2 have p-adic valuation at least r as well, it follows from 1.14.17 that so does t_3, finishing the proof of this case.

It remains to discuss what happens when some of the P_i don't lie on the affine patch $Y \neq 0$, i.e., are nontrivial 2-torsion points. Note that the condition $P_1, P_2 \in E_{p,r}(\mathbb{Q})$ implies that $y(P_i) \neq 0$ for $i = 1, 2$ (see above), so that P_1, P_2 are not 2-torsion. Therefore, the only possibility we need to worry about is when P_3 is 2-torsion. In general, it is possible for two points which are *not* nontrivial 2-torsion points to sum to one (e.g., if $P_1 = P_2$ has order 4); however, we actually see from the above proof that this *cannot* happen if $P_1, P_2 \in E_{p,r}(\mathbb{Q})$, because of the equation 1.14.17. Indeed, the only way P_3 could be a two-torsion point is if its (t, s) coordinates blew-up (i.e., were "infinite"), and this would require us to have $1 + a\lambda + b\lambda^2 + c\lambda^2 = 0$ (check!). Therefore, the fact that this quantity is nonzero and the equation 1.14.17 is *sufficient* to conclude that P_3 is not a two-torsion point either, and the above proof applies. ■

Exercise 1.14.18. Convince yourself that the argument at the end of this proof works, and that we have actually proven what we claimed to.

At this point, we have proven 1.14.12. We will use this result to finish the proof of 1.14.2 and hence 1.13.7 next time.

1.15 26/07/08 - Finishing the Proof of Nagell-Lutz, Introduction to Descent and the Mordell-Weil Theorem

The first order of business today is to wrap up the proof of the Nagell-Lutz Theorem (1.13.7) by finishing proving that torsion points have integer coordinates (1.14.2). After this, we will start building the theory needed to prove the Mordell-Weil Theorem: the machinery of descent.

1.15.A Completing the Proof of Nagell-Lutz

Let's recall where we are: $E : y^2 = x^3 + ax^2 + bx + c$ is an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$, and for each prime p and integer $r \in \mathbb{Z}_{\geq 1}$ we have defined a subgroup $E_{p,r}(\mathbb{Q}) \subset E(\mathbb{Q})$ consisting of points $P = (x, y) \in E(\mathbb{Q})$ with $v_p(x) \leq -2r$ (along with $\mathscr{O}$). The end of the proof last time suggested that torsion points don't like being in the subgroups $E_{p,r}(\mathbb{Q})$. This is indeed correct, and we will see how to formalize this momentarily. However, note first that the above proof (1.14.12) gives us something stronger than we were going for; we have actually shown

Proposition 1.15.1. In the above notation, suppose that $P_1, P_2, P_3 \in E_{p,r}(\mathbb{Q})$ (counted with multiplicities!) sum to zero, i.e., are such that $P_1 + P_2 + P_3 = \mathscr{O}$. Then their t-coordinates have the property that

$$v_p(t_1 + t_2 + t_3) \geq 3r.$$

Technically, we haven't checked this in every single case; there are some edge cases left to check, which I leave to the reader.

Exercise 1.15.2. Figure out what edge cases needed to be checked, and verify that this result is true also in these edge cases.

This suggests that something interesting is happening with the t-coordinates. For this, we can consider the map $t : E_{p,r}(\mathbb{Q}) \to p^r\mathbb{Z}_{(p)}$, where $\mathbb{Z}_{(p)} := \{q \in \mathbb{Q} : v_p(q) \geq 0\}$ is the subring of $\mathbb{Q}$ consisting of of rational numbers with "no p in the denominator". This ring has many fascinating properties; for instance, it has unique factorization, and in fact the only prime element it has is p (up to units, of which there are lots!). If you haven't seen these facts before, you should try proving them yourself:

Exercise 1.15.3. Show that the ring $\mathbb{Z}_{(p)}$ above has the claimed properties. [DVR]

Now could the map $t : E_{p,r}(\mathbb{Q}) \to p^r\mathbb{Z}_{(p)}$ be a group homomorphism? Not quite, because even if $P_1, P_2, P_3 \in E_{p,r}(\mathbb{Q})$ sum to zero, their t-coordinates need not sum to zero in $p^r\mathbb{Z}_{(p)}$, which is certainly a necessary condition for a map to be a group homomorphism; however, 1.15.1 suggests how to fix this. We post-compose with the quotient by the subgroup $p^{3r}\mathbb{Z}_{(p)} \subset p^r\mathbb{Z}_{(p)}$ to get a map

$$[t] : E_{p,r}(\mathbb{Q}) \to p^r\mathbb{Z}_{(p)}/p^{3r}\mathbb{Z}_{(p)}. \tag{1.15.4}$$

This map $[t]$ certainly sends $\mathscr{O}$ to 0, and since negation in (t, s) coordinates is simply $-(t, s) = (-t, -s)$ (check!), it certainly preserves negatives. Therefore, in light of 1.15.1, we see that that 1.15.4 is actually a group homomorphism, thanks to

Lemma 1.15.5. Let A and B be an abelian groups and $\phi : A \to B$ be a map of sets such that

(i) we have $\phi(0) = 0$,
(ii) if $a \in A$, then $\phi(-a) = -\phi(a)$, and
(iii) if $a_1, a_2, a_3 \in A$ are such that $a_1 + a_2 + a_3 = 0$, then also $\phi(a_1) + \phi(a_2) + \phi(a_3) = 0$.

Then ϕ is a group homomorphism.

Exercise 1.15.6. Prove 1.15.5 (c.f. 1.14.14).

We have now produced a homomorphism $[t]$ out of $E_{p,r}(\mathbb{Q})$. What is the kernel of $[t]$? It's the set of all points whose t-coordinates actually lie in $p^{3r}\mathbb{Z}_{(p)}$, but we have a name for such points: these are

precisely the elements of $E_{p,3r}(\mathbb{Q})$. Therefore, $\ker[t] = E_{p,3r}(\mathbb{Q})$, and the First Isomorphism Theorem gives us an *injective homomorphism* of abelian groups

$$E_{p,r}(\mathbb{Q})/E_{p,3r}(\mathbb{Q}) \hookrightarrow p^r\mathbb{Z}_{(p)}/p^{3r}\mathbb{Z}_{(p)}.$$

Since the right side is isomorphic to $p^r\mathbb{Z}/p^{3r}\mathbb{Z}$ (check!), it is cyclic of order p^{2r}. This is sufficient to deduce

Proposition 1.15.7. In the above notation, if $P \in E_{p,r}(\mathbb{Q})$, then $p^{2r}P \in E_{p,3r}(\mathbb{Q})$.

I like to think of 1.15.7 to be a strengthening of 1.14.12; while the latter says that the denominators are "persistant", this one says that the denominators actually get *worse* as you keep adding points–unless you get to $\mathscr{O}$ first. We are now ready to prove 1.14.2.

Proof of 1.14.2. Suppose that the result is not true, and pick a point $P \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ such that $x(P) \notin \mathbb{Z}$. In fact, pick P to have least possible order $n \in \mathbb{Z}_{\geq 2}$ among such points. (This is possible by the well-ordering principle!) We will use this point P to get a contradiction.

The condition that $x(P) \notin \mathbb{Z}$ just says that there is a prime p such that $P \in E_{p,1}(\mathbb{Q})$; pick such a prime p, and let $r := -v_p(x(P))/2$ (which belongs to $\mathbb{Z}_{\geq 1}$ by 1.14.8), so that $P \in E_{p,r}(\mathbb{Q}) \smallsetminus E_{p,r+1}(\mathbb{Q})$. Now we split into two cases:

(a) Suppose that $p \mid n$, and let $n = pm$ for some $m \in \mathbb{Z}_{\geq 1}$. Let $Q := pP$. Since P has exact order n, the point Q has exact order $m < n$. On the one hand, since $E_{p,r}(\mathbb{Q})$ is a subgroup (1.14.12) and $P \in E_{p,r}(\mathbb{Q})$, we must also have $Q \in E_{p,r}(\mathbb{Q}) \subset E_{p,1}(\mathbb{Q})$. On the other hand, since $m < n$ and n was chosen to be minimal, if $Q \neq \mathscr{O}$, then Q has integer coordinates, and hence $Q \notin E_{p,1}(\mathbb{Q})$. The only way out of this crisis is if $Q = \mathscr{O}$, which means that $m = 1$ and $n = p$. But if this happens, then using that $[t]$ is a homomorphism, we see that

$$0 = [t](Q) = [t](pP) = p \cdot [t](P) \in p^r\mathbb{Z}_{(p)}/p^{3r}\mathbb{Z}_{(p)},$$

forcing $p \cdot t(P) \in p^{3r}\mathbb{Z}_{(p)}$, which is equivalent to $P \in E_{p,3r-1}(\mathbb{Q})$. This contradicts that $P \notin E_{p,r+1}(\mathbb{Q})$ since $3r - 1 \geq r + 1$.

(b) Finally, suppose that $\gcd(p, n) = 1$. Then Bézout's Lemma implies there are $a, b \in \mathbb{Z}$ such that $an + bp^{2r} = 1$. Applying this to P and recalling that $nP = \mathscr{O}$, we conclude that $P = b(p^{2r}P)$. Now by 1.15.7, we have that $p^{2r}P \in E_{p,3r}(\mathbb{Q})$. Since $E_{p,3r}(\mathbb{Q})$ is a subgroup (1.14.12), this implies that $P = b(p^{2r}P) \in E_{p,3r}(\mathbb{Q})$, which is again a contradiction to $P \notin E_{p,r+1}(\mathbb{Q})$.

■

This completes the proof of 1.14.2 and hence the Nagell-Lutz Theorem (1.13.7).

1.15.B The Mordell-Weil Theorem and The Machinery of Descent

The following discussion has been adapted from [2, §3.1]. The primary goal for the rest of the course is to prove, as promised, the famous "finite basis" theorem of Mordell.

Theorem 1.15.8 (Mordell). Let $E/\mathbb{Q}$ be an elliptic curve. The group $E(\mathbb{Q})$ is finitely generated.

As noted before, this theorem is equivalent to saying that there are finitely many "seed points" in $E(\mathbb{Q})$ such that every other rational point on E can be obtained by iteratively applying the chord-and-tangent processes to the seeds. This theorem was first conjectured by French mathematician Henri Poincaré [TODO] in around 1901, and proven by British mathematician Louis Mordell in 1922. Later, André Weil (the advisor of Élisabeth Lutz mentioned above) generalized the result to abelian varieties over number fields in 1928 in his doctoral thesis; this is why the theorem is sometimes also known as the Mordell-Weil theorem, although the case of elliptic curves over $\mathbb{Q}$ is essentially due to Mordell.

To prove 1.15.8, we will need criteria to study when a group is finitely generated. Suppose that A is a finitely generated abelian group. What can we then say about $A/2A$? Well, if we pick $n \in \mathbb{Z}_{\geq 1}$ and

generators $a_1, \ldots, a_n \in A$ of A, then the 2^n elements $\sum_{i \in I} a_i$ as I runs over subsets of $\{1, \ldots, n\}$ form coset representatives for all of $A/2A$, and hence $A/2A$ is a *finite* group of size at most 2^n. Therefore, for A to be finitely generated, we certainly need $A/2A$ to be finitely generated. So the first thing we need to show is

Theorem 1.15.9 (Weak Finite Basis)**.** Let $E/\mathbb{Q}$ be an elliptic curve. Then $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.

Is this condition enough? Sadly, no. Even if we ask for A/pA to be finite for *all* primes p, this is still not enough: the group $A = \mathbb{Q}$ (the additive group of rational numbers) has the property that $A/pA = 0$ for all primes p, and yet A is *not* finitely generated.

Exercise 1.15.10. If you haven't seen this before, show that the additive group of rational numbers $A = \mathbb{Q}$ is *not* finitely generated.

What is missing is a notion of "size". To motivate this, let us look back at our proof of Fermat's Last Theorem for $n = 4$ (1.3.A). Fermat's proof proceeded by *infinite descent* in which we started with one solution and produced a "smaller" one. The well-ordering principle then tells us that this procedure cannot go on forever. For this to work, we need the notion of the "size" of a solution; fortunately, there, we could just use the size of integers.

In general, given an abelian group A, suppose we have a notion of "size" (also known as "height") on A as measured by a function $h : A \to [0, \infty)$. For something like what we did in 1.3.A to work for this size function h, we need it to satisfy some properties. Let's think through what some of these might be.

(a) Firstly, we need to ensure that for any bound say $C_1 \in \mathbb{R}_{\geq 0}$, there are only finitely many points of size at most C_1. Indeed, if there were infinitely many points of small size, then no proof along the lines of 1.3.A is going to work.

(b) Secondly, the height function should work well with translation on the group (i.e., must interact with the group structure). In other words, we need to ask that for all $b \in A$, the map $\tau_b : A \to A$ given by $a \mapsto a + b$ doesn't change heights too much, i.e., that we have $h(a + b) \approx h(a)$. Actually, it suffices to give an upper bound, i.e., to ask that for all $b \in A$ there is a constant $C_2(b) \in \mathbb{R}_{\geq 0}$ (possibly depending on b) such that for all $a \in A$ we have $h(a + b) \leq 2h(a) + C_2(b)$.

(c) Finally, we need to ensure that the size of an element does scale with taking multiples; that $h(na) \gtrsim nh(a)$ for all $a \in A$ and $n \in \mathbb{Z}_{\geq 1}$. Actually, for the purposes of our discussion, we will deal with a function h that grows quadratically and not linearly, so we might replace the scaling factor n on the right side by n^2. Again it suffices to prove things up to overall constants. Therefore, we might require that there be some $C_3 \in \mathbb{R}_{\geq 0}$ such that for all $a \in A$ we have $h(2a) \geq 4h(a) - C_3$.

If the above conditions don't feel very motivated yet, I sure hope they do after you see

Theorem 1.15.11 (Abstract Descent)**.** Let A be an abelian group, and let $h : A \to [0, \infty)$ be a function. Suppose that h satisfies the three conditions:

(a) (Finiteness) For each $C_1 \in \mathbb{R}_{\geq 0}$, the set $\{a \in A : h(a) \leq C_1\}$ is finite.
(b) (Translation) For each $b \in A$, there is a $C_2(b) \in \mathbb{R}_{\geq 0}$ (possibly depending on b), such that for all $a \in A$, we have $h(a + b) \leq 2h(a) + C_2(b)$.
(c) (Scaling) There is a $C_3 \in \mathbb{R}_{\geq 0}$ such that for all $a \in A$, we have $h(2a) \geq 4h(a) - C_3$.

If further $A/2A$ is finite, then A is finitely generated.

A function $h : A \to [0, \infty)$ on an abelian group A satisfying (a), (b), and (c) as above is called a *height function*. To motivate the proof of 1.15.11, suppose we take a finite set $S \subset A$ of coset representatives for $A/2A$. This means that given any $a \in A$, there is a $b_1 \in S$ and $a_1 \in A$ such that $a = b_1 + 2a_1$. Now we hope that a_1 has smaller "size" than a (as measured by h), and indeed

$$h(a_1) \leq \frac{1}{4}h(2a_1) + \frac{1}{4}C_3 \leq \frac{1}{4}\left(2h(a) + C_2(-b_1)\right) + \frac{1}{4}C_3 = \frac{1}{2}h(a) + \frac{1}{4}\left(C_2(-b_1) + C_3\right),$$

where the first step uses (c) and the second uses (b). Since there are only finitely many choices for $b_1 \in S$, we might as well take $C_2 := \max_{b \in S} C_2(-b)$ and let $C := (C_2 + C_3)/4$ to get the more uniform bound

$$h(a_1) \leq \frac{1}{2}h(a) + C. \tag{1.15.12}$$

This tells us that (up to the constant term C), the height of a_1 is less than half that of a. In particular, we can iterate this procedure to get points of smaller and smaller height, until we land in a set of height small enough that is independent of the $a \in A$ we began with. Since the set of such points will be finite by (a), this will be enough to prove that A is finitely generated. Let's make this into a formal proof.

Proof. Let $S \subset A$ be a finite set of coset representatives for $A/2A$, and let $C_2 := \max_{b \in S} C_2(-b)$ and $C := (C_2 + C_3)/4$ as above. Let $T := \{a \in A : h(a) \leq 2C + 1\}$, which is finite by (a). (We'll see where this bound in the definition of T comes from in just a moment.) We claim that the finite subset $S \cup T$ generates A.

To see this, suppose we are given an $a \in A$; we need to show that a lies in the subgroup of A generated by $S \cup T$. As above, since S forms a complete set of representatives, there is a $b_1 \in S$ and $a_1 \in A$ such that $a = b_1 + 2a_1$. Iterating this process, we obtain recursively the elements $a_n \in A$ and $b_n \in S$ such that for each $n \in \mathbb{Z}_{\geq 1}$ we have $a_{n-1} = b_n + 2a_n$ (where $a_0 := a$). By an iteration of 1.15.12, we see immediately that for each $n \geq 1$,

$$h(a_n) \leq \frac{1}{2^n} h(a) + \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{n-1}}\right) C \leq \frac{1}{2^n} h(a) + 2C.$$

This tells us that if $N \gg 1$ (specifically, $N \geq \log_2 h(a)$ if $h(a) \neq 0$; if $h(a) = 0$ then $a \in T$ and there is nothing to show), then $a_N \in T$. It then follows from the unfolding

$$a = b_1 + 2b_2 + \cdots + 2^{N-1} b_N + 2^N a_N$$

that a lies in the subgroup of A generated by $S \cup T$, as needed. ∎

Therefore, to prove 1.15.8 for an elliptic curve $E/\mathbb{Q}$ by invoking 1.15.11 for $A = E(\mathbb{Q})$, we need to do two things: we need to prove the weak finite basis theorem (1.15.9), and then we need to come up with a function $h : E(\mathbb{Q}) \to [0, \infty)$ that satisfies properties (a), (b), (c) of 1.15.11. We'll tackle the second of these issues first next time.

Remark 1.15.13. Why is 1.15.11 called the "descent" theorem? Well, the terminology really does come from Fermat's method of infinite descent which we saw in 1.3.A. The idea is that we start with a solution (here, a), and produce from it a "smaller" solution (here, a_1). We keep continuing this way to get smaller and smaller solutions (here, a_n). If we get to a point where the smaller solution is smaller than is possible for that particular Diophantine equation (as happened in 1.3.A), then we are done–we have shown there are no solutions. What happens more generally is that we keep iterating this procedure to get to a point where the last iterate in this process is so small that it lies in some finite set of solutions that we understand (here, T). Then, up to the finitely many terms b_i which go into getting to this small solution, this finite set of "small" solutions generates every other.

Unimportant Remark 1.15.14. There is nothing special about the number 2 above, and it could have been replaced by any integer $m \geq 2$ (c.f. [8, Thm. VIII.3.1]). It just that the choice of $m = 2$ is very convenient, as we shall see next time.

1.16 26/07/09 - The Height Machine

Recall that last time we described a general "descent" procedure (1.15.11) which can be used to prove that a group is finitely generated. For this, we needed the notion of "size" or "height" on the abelian group. The goal for today is today is to define a *height function* on the set of rational points of an elliptic curve over $\mathbb{Q}$, and check that it satisfies the required properties for 1.15.11. This would then reduce the proof of Mordell's Theorem (1.15.8) to the weaker 1.15.9. We follow the treatment in [2, §3.2-3.3].

The right notion of height on the set of rational points of an elliptic curve is a little more subtle than one might expect. One could hope that a naive notion, say the absolute value of the x-coordinate might work. However, a little experimentation (2.2.3) shows that this definition fails 1.15.11(c) (why?). Nonetheless, anyone who has experimented with rational points on elliptic curves (see, e.g., 2.1.9(b)) knows that what blows up very quickly as you take multiples of points on an elliptic curve is the size of the numerators and denominators of the coordinates. The right definition of height, as it turns out, is indeed complexity-theoretic.

Definition 1.16.1.

(a) Let $q \in \mathbb{Q}$, and write q in lowest terms as $q = u/v$ for $u, v \in \mathbb{Z}$ with $v \geq 1$ and $\gcd(u, v) = 1$. Define the *height* $H(q)$ of q to be

$$H(q) := \max\{|u|, |v|\} \in \mathbb{Z}_{\geq 1},$$

and the *logarithmic height* $h(q)$ of q to be

$$h(q) := \log H(q) \in [0, \infty).$$

(b) Let $E/\mathbb{Q}$ be an elliptic curve in Weierstrass form. Given a point $P \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, we define the *height* (resp. *logarithmic height*) of P by $H(P) := H(x(P))$ (resp. $h(P) := h(x(P)) = \log H(P)$). By convention, we also define $H(\mathscr{O}) = 1$ so that $h(\mathscr{O}) = 0$.

Observation 1.16.2. Given any $C_1 \in \mathbb{R}_{\geq 0}$, the set $\{q \in \mathbb{Q} : h(q) \leq C_1\}$ is finite.

Remark 1.16.3.

(a) Roughly speaking, the logarithmic height of a rational number is approximately the number of bits a computer might need to store it. I am also being deliberately vague about the base of the logarithm above–it doesn't really matter, as long as you pick one base and stick with it throughout.

(b) There is nothing really special about the function $u/v \mapsto \max\{|u|, |v|\}$; we could easily have chosen something different such as $u/v \mapsto |u| + |v|$ or $u/v \mapsto \sqrt{u^2 + v^2}$. This wouldn't give the same exact H or h, but the resulting H and h would be comparable thanks to the inequalities

$$\frac{1}{2}(|u| + |v|) \leq \max\{|u|, |v|\} \leq |u| + |v| \text{ and } \frac{1}{\sqrt{2}}\sqrt{u^2 + v^2} \leq \max\{|u|, |v|\} \leq \sqrt{u^2 + v^2},$$

and we would get the same results (which you can later check!). It is sometimes advantageous to use these different kinds of heights instead; one such example of where the ℓ_2 height given by $u/v \mapsto \sqrt{u^2 + v^2}$ is more useful can be found in [16]. For another example of where the ℓ_1 height $u/v \mapsto |u| + |v|$ is more useful, see the proof of 1.16.6 (the part after 1.16.7).

(c) Similarly, there is nothing really special about taking the height of the x coordinate of a point, as opposed to that of its y-coordinate. See 1.16.4.

Lemma 1.16.4. Let $E/\mathbb{Q}$ be an elliptic curve in short Weierstrass form $y^2 = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{Z}$. There is an absolute constant $C_4 \in \mathbb{R}_{\geq 0}$ depending only on a, b, c with the following property: for any $P = (m/e^2, n/e^3) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ as in 1.14.8 with $m, n, e \in \mathbb{Z}$ and $e \geq 1$ and $\gcd(m, e) = \gcd(n, e) = 1$, we have that

$$|m| \leq H(P), \quad e^2 \leq H(P), \quad \text{and } |n| \leq C_4 \cdot H(P)^{3/2}.$$

In particular,

$$h(y(P)) \leq \frac{3}{2}h(P) + \log C_4. \tag{1.16.5}$$

Proof. The inequalities $|m| \leq H(P)$ and $e^2 \leq H(P)$ come from the definition of $H(P)$. To see the third one, note that by $P \in E(\mathbb{Q})$ we have the equality

$$n^2 = m^3 + am^2e^2 + bme^4 + ce^6.$$

Now using the triangle inequality and the first two inequalities, we see that

$$|n|^2 \leq |m|^3 + |a| \cdot |m|^2 \cdot e^2 + |b| \cdot |m| \cdot (e^2)^2 + |c| \cdot (e^2)^3 \leq (1 + |a| + |b| + |c|) \cdot H(P)^3,$$

so it suffices to take $C_4(E) := \sqrt{1 + |a| + |b| + |c|}$. The last inequality follows immediately upon taking logarithms. ∎

In fact, it is true that $h(y(P)) \asymp (3/2)h(P)$, i.e., there is also a constant making the inequality 1.16.5 go the other way. If you're curious, you can try to prove this; we won't need it, and so we won't prove it (see also 2.5.8). We are now ready to prove what we wanted.

Theorem 1.16.6. Let $E/\mathbb{Q}$ be an elliptic curve in short Weierstrass form

$$E : y^2 = x^3 + ax^2 + bx + c$$

with $a, b, c \in \mathbb{Z}$. The function

$$h : E(\mathbb{Q}) \to [0, \infty)$$

defined by 1.16.1(b) satisfies the conditions (a), (b), and (c) of 1.15.11. In other words, we have that:

(a) For each $C_1 \in \mathbb{R}_{\geq 0}$, the set $\{P \in E(\mathbb{Q}) : h(P) \leq C_1\}$ is finite.
(b) For each $Q \in E(\mathbb{Q})$, there is a $C_2(Q) \in \mathbb{R}_{\geq 0}$ such that for all $P \in E(\mathbb{Q})$, we have the inequality $h(P + Q) \leq 2h(P) + C_2(Q)$.
(c) There is a $C_3 \in \mathbb{R}_{\geq 0}$ such that for all $P \in E(\mathbb{Q})$, we have $h(2P) \geq 4h(P) - C_3$.

If you're still not too sure about where the general form of the conditions (a), (b), and (c) comes from, you will see the real source in the course of the following proof.

Proof. The condition (a) is automatic from 1.16.2 and the fact that each x-value corresponds to at most two y-values. Therefore, we have to show (b) and (c).

Let's show (b). If $Q = \mathscr{O}$, we can take $C_2(\mathscr{O}) = 0$. Hence suppose $Q \neq \mathscr{O}$. Note that it suffices to choose the $C_2(Q)$ such that the proposed inequality holds for all but finitely many P, say all except P_i, for we can always replace $C_2(Q)$ by the larger of $C_2(Q)$ and $\max_i\{2h(P_i) - h(P_i + Q)\}$. Therefore, it suffices to find a $C_2(Q)$ that works for all $P \in E(\mathbb{Q}) \smallsetminus \{O, \pm Q\}$.

At this point, we just have to formula-crunch. Suppose that $Q = (x_0, y_0)$ and $P = (x, y)$. Since $P \neq \pm Q$, we have that $x \neq x_0$. Using the formulae in 1.12.D, we see that

$$x(P + Q) = \left(\frac{y - y_0}{x - x_0}\right)^2 - a - x - x_0 = \frac{(y - y_0)^2 - (x + x_0 + a)(x - x_0)^2}{(x - x_0)^2}.$$

Now the numerator contains a term of the form $y^2 - x^3$, which we can replace by $ax^2 + bx + c$, since $P \in E(\mathbb{Q})$. After clearing denominators of the rational coefficients, we see a general formula of the form

$$x(P + Q) = \frac{A_0y + A_1x^2 + A_2x + A_3}{A_4x^2 + A_5x + A_6}$$

for some $A_0, \ldots, A_6 \in \mathbb{Z}$ that depend only on a, b, c and Q but not on P. Next, we can use 1.14.8 to write P as $(m/e^2, n/e^3)$, where $m, n, e \in \mathbb{Z}$ and $e \geq 1$ with $\gcd(m, e) = \gcd(n, e) = 1$. This allows us to rewrite $x(P + Q)$ as

$$x(P + Q) = \frac{A_0ne + A_1m^2 + A_2me^2 + A_3e^4}{A_4m^2 + A_5me^2 + A_6e^4}.$$

This fraction may not be in its lowest terms, but we are only looking for an upper bound on the height, so we can certainly conclude that

$$H(P+Q) \leq \max\left\{|A_0ne + A_1m^2 + A_2me^2 + A_3e^4|, |A_4m^2 + A_5me^2 + A_6e^4|\right\}.$$

Again, the triangle inequality, along with the inequalities in 1.16.4, give us the estimates

$$|A_0ne + A_1m^2 + A_2me^2 + A_3e^4| \leq (|A_0C_4| + |A_1| + |A_2| + |A_3|) \cdot H(P)^2$$

and

$$|A_4m^2 + A_5me^2 + A_6e^4| \leq (|A_4| + |A_5| + |A_6|) \cdot H(P)^2,$$

so it suffices to take

$$C_2(Q) := \log \max\{|A_0C_4| + |A_1| + |A_2| + |A_3|, |A_4| + |A_5| + |A_6|\},$$

which again depends only on a, b, c and Q but not on P.

Finally, let's show (c). As in (b), we can ignore finitely many P_i, for we can always replace C_3 by the larger of C_3 and $\max_i\{4h(P_i) - h(2P_i)\}$. Therefore, it suffices to find a C_3 that works for all $P \in E(\mathbb{Q}) \smallsetminus E[2](\mathbb{Q})$. Again, at this point it is just formula-crunching, although there are some interesting ideas involved. Suppose that $P = (x, y)$. The duplication formula (1.12.13 and 1.14.3) tells us that the x-coordinate of $2P$ is

$$x(2P) = \frac{x^4 - 2bx^2 - 8cx + b^2 - 4ac}{4(x^3 + ax^2 + bx + c)}.$$

Let us write $x(P) = u/v$ in lowest terms, so that $u, v \in \mathbb{Z}$ with $v \geq 1$ and $\gcd(u, v) = 1$. (In the language of 1.14.8, we have $u = m$ and $v = e^2$.) Then the duplication formula looks like

$$x(2P) = \frac{u^4 - 2bu^2v^2 - 8cuv^3 + (b^2 - 4ac)v^4}{4(u^3v + au^2v^2 + buv^3 + cv^4)}.$$

For simplicity, let's call the numerator of this expression $\Phi(u, v)$, and the denominator $\Psi(u, v)$; both are integers, and we need to give an upper bound on their greatest common divisor. For this, we invoke

Lemma 1.16.7. Let $u, v \in \mathbb{Z}$ with $v \geq 1$ and $\gcd(u, v) = 1$. In the above notation,

$$\gcd(\Phi(u, v), \Psi(u, v)) \mid 4D,$$

where $D = \operatorname{disc}(f) = -4a^3c + a^2b^2 + 18abc - 4b^3 - 27c^2$ is as in 1.13.B.

Assuming this result for the moment, we finish the proof as follows: since we have bound the cancellation, we get a lower bound on the height of the form

$$H(2P) \geq \frac{1}{4|D|} \max\{|\Phi(u, v)|, |\Psi(u, v)|\} = \frac{v^4}{4|D|} \max\{|g(x)|, |4f(x)|\} \geq \frac{v^4}{4|D|} \cdot \frac{1}{2}(|g(x)| + |4f(x)|).$$

(This is where using the ℓ_1 height might have been a little easier.) We need to compare this with

$$H(P)^4 = \max\{u^4, v^4\} = v^4 \max\{x^4, 1\}.$$

The ratio $H(2P)/H(P)^4$ cancels out the v^4 term (this is why there is 4 in (c)!), and we get that

$$\frac{H(2P)}{H(P)^4} \geq \frac{1}{8|D|} \frac{|g(x)| + |4f(x)|}{\max\{x^4, 1\}}.$$

Let's take a second to study the real function

$$\beta : \mathbb{R} \to [0, \infty), \quad x \mapsto \beta(x) := \frac{|g(x)| + |4f(x)|}{\max\{x^4, 1\}}.$$

Since $g(x)$ is a monic quartic polynomial and $f(x)$ is cubic, we have that

$$\lim_{x \to \pm 1} \beta(x) = 1.$$

Further, since $g(x)$ and $f(x)$ have no common roots (this uses the smoothness of E, and was observed in the proof of 1.14.1), the function β never attains the value 0. These two facts, combined with the observation that β is continuous and the Extreme Value Theorem from calculus, are enough to conclude that β stays a positive distance away from the x-axis, i.e., that

$$\xi := \inf_{x \in \mathbb{R}} \beta(x) \in (0, \infty).$$

This gives us an inequality of the form

$$\frac{H(2P)}{H(P)^4} \geq \frac{\xi}{8D},$$

and so it suffices to take $C_3 := -\log(\xi/8|D|)$. To finish the proof of this theorem, it remains to give the

Proof of 1.16.7. Recall the polynomials

$$P(x) = 3x^3 - ax^2 - 5bx + 2ab - 27c \text{ and } Q(x) = -3x^2 - 2ax + a^2 - 4b$$

from the proof of 1.14.1, which had the property (1.14.4) that

$$P(x) \cdot 4f(x) + 4Q(x) \cdot g(x) = 4D.$$

Writing $x = u/v$ and multiplying throughout by v^7 gives us an equation of the form

$$\left(v^3 P\left(\frac{u}{v}\right)\right) \cdot \Psi(u, v) + v \cdot \left(v^2 Q\left(\frac{u}{v}\right)\right) \cdot \Phi(u, v) = 4Dv^7.$$

The quantities in parentheses are integers. Therefore, if we let $g := \gcd(\Phi(u, v), \Psi(u, v))$, then $g \mid 4Dv^7$. But since g also divides

$$\Phi(u, v) = u^4 - 2bu^2v^2 - 8cuv^3 + (b^2 - 4ac)v^4,$$

and $\gcd(u, v) = 1$, we conclude that $\gcd(g, v) = 1$. It follows from $g \mid 4Dv^7$ then that $g \mid 4D$. This finishes the proof of 1.16.7, ... ■

... and hence also of 1.16.6. ■

At this point, we have reduced the proof of Mordell's Theorem (1.15.8) to the 1.15.9. Next time, I will make a few more remarks about height before moving on to the proof of 1.15.9.

1.17 26/07/10 - The Néron-Tate Canonical Height

Last time, we discussed the height function $h : E(\mathbb{Q}) \to [0, \infty)$ on an elliptic curve $E/\mathbb{Q}$ and used this to reduce the proof of the Mordell Theorem (1.15.8) to the weaker 1.15.9. Today, I want to spend a little bit more time talking about the theory of heights before we move on the proof of the weak Mordell Theorem.

1.17.A The Approximate Parallelogram Law for the Naive Height

For this, we will need one more result about heights. This is

Proposition 1.17.1. Let $E : y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$. Then for all $P, Q \in E(\mathbb{Q})$, we have that

$$h(P + Q) + h(P - Q) = 2h(P) + 2h(Q) + O(1), \tag{1.17.2}$$

where the constant implicit the $O(1)$ depends on a, b and c but not on P, Q.

At this point, it's helpful to introduce the "big O" notation from complexity theory. The above proposition is just saying that there is a constant $C_5 \in \mathbb{R}_{\geq 0}$, depending only on a, b, c, such that for all $P, Q \in E(\mathbb{Q})$, we have that

$$|h(P + Q) + h(P - Q) - 2h(P) - 2h(Q)| \leq C_5.$$

In general, we will use this notation to mean that there is a constant (we will make to sure to emphasize what exactly it depends on) such that the quantity supposed to be $O(1)$ is bounded in absolute value by that constant.

Proof. The proof of 1.17.2 is very similar to that of 1.16.6 and is left as Exercise 2.5.5. ■

Remark 1.17.3. What does the identity 1.17.2–except for the extra $O(1)$ term–remind you of? It should remind you of both the polynomial identity $(p+q)^2 + (p-q)^2 = 2p^2 + 2q^2$ in $\mathbb{Z}[p, q]$, and of the identity

$$\left\|\vec{P} + \vec{Q}\right\|^2 + \left\|\vec{P} - \vec{Q}\right\|^2 = 2\left\|\vec{P}\right\|^2 + 2\left\|\vec{Q}\right\|^2$$

for vectors $\vec{P}, \vec{Q}$ in an inner product space, which says that the sum of the squares of the four sides of a parallelogram is the sum of the squares of the its diagonals. For this reason, we will refer to 1.17.2 as the (approximate) parallelogram law.

Let's now derive some consequences of 1.17.1.

Corollary 1.17.4. Let $E/\mathbb{Q}$ be an elliptic curve as in 1.17.1.

(a) Fix a $Q \in E(\mathbb{Q})$. For each $P \in E(\mathbb{Q})$, we have that $h(P + Q) \leq 2h(P) + O(1)$, where the constant in $O(1)$ depends on a, b, c, and Q, but not P.
(b) Fix an $m \in \mathbb{Z}$. For each $P \in E(\mathbb{Q})$, we have that $h(mP) = m^2 h(P) + O(1)$, where the constant in $O(1)$ depends on a, b, c, and m but not P.

Proof.

(a) Follows immediately from 1.17.1, since $h(P - Q) \geq 0$.
(b) Since $h(P) = h(-P)$, it suffices to prove this for $m \in \mathbb{Z}_{\geq 0}$. We proceed by induction, with the cases $m = 0, 1$ being clear. If $m \geq 1$, then applying 1.17.2 to the pair of points (mP, P) gives us that
$$h((m + 1)P) + h((m - 1)P) = 2h(mP) + 2h(P) + O(1),$$
where the absolute $O(1)$ depends only on a, b, c. Now using the inductive hypothesis to write
$$h((m - 1)P) = (m - 1)^2 h(P) + O(1) \text{ and } h(mP) = m^2 h(P) + O(1),$$

this time with the $O(1)$s depending on a, b, c, and m but not P, and putting everything together gives us

$$h((m+1)P) = \left(2m^2 + 2 - (m-1)^2\right) + O(1)$$

as needed.

■

You might start to see the advantage of the big O notation in this proof.

Remark 1.17.5.

(a) Note that 1.17.4(a) is just a restatement of 1.16.6(b); the reason I point this out is because I want to emphasize that it is a consequence of the approximate parallelogram law (1.17.1).
(b) Similarly, the condition 1.16.6(c) is one half of the statement of 1.17.4(b). It says in particular that, in general, point doubling on elliptic curve *quadruples* the number of digits of the numerator and denominator; in particular, this applies to the Bachet formula of 1.1.6. This fulfils a promise made in 1.1.6.

1.17.B The Néron-Tate Canonical Height

The above theory says that the (naive) height function h behaves approximately like a quadratic function of the point, except for these annoying $O(1)$s floating around. It is at least theoretically interesting to be able to get an actual quadratic function out of the height function. For that: how do we get rid of the $O(1)$s? *Can* we get rid of them?

It was an insight of John Tate (1925-2019) that we indeed can. To motivate this result, suppose that $E/\mathbb{Q}$ is an elliptic curve as above and that $\hat{h} : E(\mathbb{Q}) \to [0, \infty)$ is a function that

(a) differs from h by a bounded amount, i.e., is such that for all $P \in E(\mathbb{Q})$, we have that $\hat{h}(P) = h(P) + O(1)$, where the constant in $O(1)$ depends only on a, b, c and not P, and
(b) is genuinely quadratic, i.e., satisfies that for all $m \in \mathbb{Z}$ and $P \in E(\mathbb{Q})$ we have that $\hat{h}(mP) = m^2\hat{h}(P)$.

Tate's observation was that these two conditions are enough to pin down such an $\hat{h}$ uniquely–if it exists. Indeed, suppose that $\hat{h}$ is such a function. Then (a) is saying that there is a $C \in \mathbb{R}_{\geq 0}$ such that for all $P \in E(\mathbb{Q})$ we have

$$\left|\hat{h}(P) - h(P)\right| \leq C.$$

If this is true for all P, this is in particular true if we replace P by $2^N P$ for any $N \in \mathbb{Z}_{\geq 0}$. Doing this, dividing throughout by 4^N, and using property (b) then gives us that for all $P \in E(\mathbb{Q})$ and $N \in \mathbb{Z}_{\geq 0}$ we have that

$$\left|\hat{h}(P) - \frac{1}{4^N} h(2^N P)\right| \leq \frac{C}{4^N}.$$

At this point, it doesn't matter what C is at all. If such an $\hat{h}$ exists, then we may take the limit as $N \to \infty$ to conclude that

$$\hat{h}(P) = \lim_{N \to \infty} \frac{1}{4^N} h(2^N P).$$

Tate observed that this can be taken as a definition of $\hat{h}$ that does actually do what we want.

Lemma/Definition 1.17.6 (The Néron-Tate Canonical Height)**.** Let $E : y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$. For each $P \in E(\mathbb{Q})$, the limit

$$\hat{h}(P) = \lim_{N \to \infty} \frac{1}{4^N} h(2^N P) \tag{1.17.7}$$

exists. The resulting function $\hat{h} : E(\mathbb{Q}) \to [0, \infty)$ is called the *Néron-Tate canonical height on the elliptic curve* E.

Remark 1.17.8.

(a) The other person whose name is attached to the quanitity $\hat{h}$ was the French mathematician André Néron (1922-1985), who had discovered this quantity earlier in terms of a more complicated expression in terms of local height functions from arithmetic intersection theory. It was Tate was realized that Néron's definition could be made much simpler via the limiting process (1.17.7).
(b) Our definition 1.17.7 is the original one due to Tate, and is consistent with the one used by Sage and the LMFDB, although it differs from the definition in [8, Prop. VIII.9.1] by a factor of 2: Silverman's height is half of ours. This won't really matter for what follows.

Proof. It suffices to show that the sequence $4^{-N}h(2^N P)$ is Cauchy. For this, note first that 1.17.4(b) implies that there is a $C \in \mathbb{R}_{\geq 0}$ such that for all $P \in E(\mathbb{Q})$ we have

$$|h(2P) - 4h(P)| \leq C.$$

Now suppose we are given $P \in E(\mathbb{Q})$ and integers $N \geq M \geq 0$. Then we estimate

$$\begin{aligned}\left|\frac{1}{4^N}h(2^N P) - \frac{1}{4^M}h(2^M P)\right| &= \left|\sum_{n=M}^{N-1}\left(\frac{1}{4^{n+1}}h(2^{n+1}P) - \frac{1}{4^n}h(2^n P)\right)\right| \\ &\leq \sum_{n=M}^{N-1}\frac{1}{4^{n+1}}\left|h(2^{n+1}P) - 4h(2^n P)\right| \\ &\leq C \cdot \sum_{n=M}^{N-1}\frac{1}{4^{n+1}} \leq C \cdot \sum_{n=M}^{\infty}\frac{1}{4^{n+1}} = \frac{C}{3}\cdot\frac{1}{4^M}.\end{aligned}$$

This proves that the sequence $N \mapsto 4^{-N}h(2^N P)$ is Cauchy, and hence that $\hat{h}(P)$ exists. ∎

This canonical height has all the properties we would like it to.

Theorem 1.17.9. Let $E : y^2 = x^3 + ax^2 + bx + c$ be an elliptic curve over $\mathbb{Q}$ with $a, b, c \in \mathbb{Z}$, and let $\hat{h} : E(\mathbb{Q}) \to [0, \infty)$ be the Néron-Tate canonical height on E as defined in 1.17.6. Then

(a) For each $P \in E(\mathbb{Q})$, we have that $\hat{h}(P) = h(P) + O(1)$, where the constant implicit in the $O(1)$ depends on a, b, c but not P. In particular, for each $C_1 \in \mathbb{R}_{\geq 0}$, the set $\{P \in E(\mathbb{Q}) : \hat{h}(P) \leq C_1\}$ is finite.
(b) For each $m \in \mathbb{Z}$ and $P \in E(\mathbb{Q})$ we have that $\hat{h}(mP) = m^2\hat{h}(P)$.

The properties (a) and (b) characterize the function $\hat{h}$ uniquely. This function $\hat{h}$ further satisfies the following properties.

(c) For any $P \in E(\mathbb{Q})$ we have that $\hat{h}(P) = 0$ iff $P \in \operatorname{Tors} E(\mathbb{Q})$.
(d) For all $P, Q \in E(\mathbb{Q})$, we have that $\hat{h}(P+Q) + \hat{h}(P-Q) = 2\hat{h}(P) + 2\hat{h}(Q)$. In other words, the *exact* parallelogram law holds for $\hat{h}$.
(e) The map $\langle -, - \rangle : E(\mathbb{Q}) \times E(\mathbb{Q}) \to \mathbb{R}$ given by

$$\langle P, Q \rangle := \hat{h}(P+Q) - \hat{h}(P) - \hat{h}(Q)$$

is a bilinear pairing.

Proof.

(a) Taking $M = 0$ in the inequality used in the proof of 1.17.6, we see that for all $P \in E(\mathbb{Q})$ and $N \in \mathbb{Z}_{\geq 0}$ we have that

$$\left|\frac{1}{4^N}h(2^N P) - h(P)\right| \leq \frac{C}{3}.$$

Taking the limit as $N \to \infty$ shows us that for all $P \in E(\mathbb{Q})$ we have $\left|\hat{h}(P) - h(P)\right| \leq C/3$ as needed. The second part of (a) follows immediately from the first and 1.16.6(a).

(b) The statement of 1.17.4(b) says that for any $m \in \mathbb{Z}$, there is a constant $C' \in \mathbb{R}_{\geq 0}$ (depending only on a, b, c, m) such that for all $P \in E(\mathbb{Q})$ we have that

$$\left|h(mP) - m^2 h(P)\right| \leq C'.$$

Replacing P by $2^N P$ for $N \geq 0$ and dividing throughout by 4^N tells us that for all N we have

$$\left|\frac{1}{4^N} h(2^N(mP)) - m^2 \cdot \frac{1}{4^N} h(2^N P)\right| \leq \frac{C'}{4^N}.$$

Taking the limit as $N \to \infty$ shows that $\hat{h}(mP) = m^2 \hat{h}(P)$ as needed.

We see in this way how the limiting procedure gets rid of all of the $O(1)$s automatically. The fact that (a) and (b) characterize $\hat{h}$ uniquely was discussed already at the beginning of §1.17.B.

(c) Note first that it follows immediately from 1.17.7 that $\hat{h}(\mathscr{O}) = 0$ (no matter what the convention for $h(\mathscr{O})$ is, as long as $h(\mathscr{O}) \in [0, \infty)$!). Next, if $P \in E(\mathbb{Q})$ is a point of order $m \in \mathbb{Z}_{\geq 1}$, then we get from (b) that

$$0 = \hat{h}(\mathscr{O}) = \hat{h}(mP) = m^2 \hat{h}(P),$$

which shows $\hat{h}(P) = 0$. Conversely, if $\hat{h}(P) = 0$, then again (b) shows that for each $m \in \mathbb{Z}_{\geq 1}$ we have $\hat{h}(mP) = 0$. Then taking $C_1 = 0$ in (a), we conclude that the set $\{mP : m \in \mathbb{Z}_{\geq 1}\}$ is finite, which can only happen if $P \in \operatorname{Tors} E(\mathbb{Q})$.

(d) This follows from 1.17.1 the same way that (b) followed from 1.17.4(b); the details are left to the reader as a(n) (easy) exercise (1.17.10).

(e) Surprisingly, this is a consequence of the parallelogram law (d); since this is fun to figure out, I leave it as Exercise 2.5.9.

■

Exercise 1.17.10. Show that 1.17.9(d) follows from 1.17.1 the same way that 1.17.9(b) follows from 1.17.4(b).

Remark 1.17.11.

(a) Note that the 1.17.9(c) was the promised relationship between torsion points and heights on the elliptic curves.

(b) The results 1.17.9(c) and (e) (see also 2.5.7) combined tell us that the pairing $\langle -, - \rangle$ of 1.17.9(e) gives rise to a nondegenerate bilinear pairing on the Mordell-Weil lattice

$$\operatorname{MW}(E/\mathbb{Q}) := E(\mathbb{Q})/\operatorname{Tors} E(\mathbb{Q}),$$

which (thanks to 1.15.8) is a finitely generated torsion-free abelian group and hence (as we shall see below) is a finitely generated free abelian group. This bilinear pairing further extends to a (positive-definite) inner product on the finite-dimensional real vector space $V := \mathbb{R} \otimes_{\mathbb{Z}} E(\mathbb{Q}) \xrightarrow{\sim} \mathbb{R} \otimes_{\mathbb{Z}} \operatorname{MW}(E/\mathbb{Q})$ (see 2.5.10), known as the *Néron-Tate height pairing*. Now $\operatorname{MW}(E/\mathbb{Q})$ sits inside of the inner product space V as a full-rank lattice, and hence has a real invariant associated to it, namely its covolume. Concretely, if $r = \operatorname{rank} E = \dim_{\mathbb{R}} V$ and we pick an integral basis $P_1, \ldots, P_r$ of $\operatorname{MW}(E/\mathbb{Q})$, then this covolume is just the determinant of the matrix $[\langle P_i, P_j \rangle]_{1 \leq i,j \leq r}$, which is independent of the choice of basis (technically, this is the *square* of the covolume). This squared covolume has a name: it is called the *regulator* of the elliptic curve $E/\mathbb{Q}$, and is denoted either $\operatorname{Reg}(E)$ or $R(E/\mathbb{Q})$. This is an invariant of E which measures its arithmetic complexity. For instance, if $\operatorname{rank} E = 1$, then the regulator equals the height of *the* generator of $\operatorname{MW}(E/\mathbb{Q})$ (which is unique up sign), and hence measures how "big" the rational solutions on E are. As another example, if the rank of E is say 2, and we have one "small" generator of $\operatorname{MW}(E/\mathbb{Q})$, then the regulator dictates how big the other generator has to be. Therefore, if we can find an independent way of calculating the regulator (as can be done, conditional on the strong Birch and Swinnerton-Dyer conjecture; see [8, Conjecture C.16.5(b)]), then this gives us an estimate of how far we have to search to get generators of the Mordell-Weil group. As a final application, we will see later how the height pairing can be used to test the independence of points on an elliptic curve (2.5.3).

That's all I have to say about heights for now, but we will return to them after we haven seen the proof of the Mordell Theorem (2.6.4).

1.18 26/07/13 - The Weak Mordell Theorem

The goal for today is to start making our way through the proof of the weak Mordell Theorem (1.15.9), which says that if $E/\mathbb{Q}$ is an elliptic curve, then the group $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. The basic strategy for this is simple: we will try to find an abelian group B and a group homomorphism $\alpha : E(\mathbb{Q}) \to B$ with the property that $\ker \alpha = 2E(\mathbb{Q})$. If we manage to do this, then α would induce an injection $E(\mathbb{Q})/2E(\mathbb{Q}) \hookrightarrow B$. If we further manage to show that $\operatorname{Im} \alpha$ is finite, then we would have succeeded.

Since we're working over $\mathbb{Q}$, we might as well assume we are working with the shorter Weierstrass form $E : y^2 = f(x)$, where $f(x) = x^3 + ax^2 + bx + c \in \mathbb{Z}[x]$ is a monic cubic polynomial with distinct (complex) roots. (We could work with the shortest Weierstrass form, but we'll see why the slightly longer form is useful today and next time.) Now here, at least when it comes to 2-torsion, there are essentially three things that can happen (as we have seen already in §1.12.C):

(a) either $f(x)$ splits completely into linear factors over $\mathbb{Q}$,
(b) or $f(x)$ splits into a linear and an irreducible monic quadratic polynomial,
(c) or the cubic polynomial $f(x)$ is irreducible.

It is possible to give a uniform proof that covers all three cases simultaneously (as is done in [1, §15]); but we'll give different proofs (and hopefully have more fun).

The cases (a), (b), and (c) are in some ways very different. It is possible to give "elementary" proofs of (a) and (b), but all the proofs of (c) I know of use at least some algebraic number theory (nothing more than the basics of number field theory). Having said this, I find the proof of (c) conceptually the cleanest, and so let's do this case first.

1.18.A The Case of Irreducible $f(x)$

The following proof has been taken from [1, §15].

Suppose that $f(x)$ is irreducible. The basic idea is to work with the cubic number field $L := \mathbb{Q}[x]/(f(x))$. Let $\beta \in L$ denote the class of x, which is a root of $f(x)$. We will prove

Theorem 1.18.1. Let $E : y^2 = f(x)$ be an elliptic curve over $\mathbb{Q}$ with $f(x) = x^3 + ax^2 + bx + c \in \mathbb{Z}[x]$ a monic irreducible polynomial. Let $L := \mathbb{Q}[x]/(f(x)) = \mathbb{Q}[\beta]$ as above, and consider the map

$$\alpha : E(\mathbb{Q}) \to L^\times/(L^\times)^2 \quad P \mapsto \begin{cases} [1] & \text{if } P = \mathscr{O}, \text{ and} \\ [x - \beta] & \text{if } P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}. \end{cases}$$

(a) The map $\alpha : E(\mathbb{Q}) \to L^\times/(L^\times)^2$ is a group homomorphism.
(b) We have that $\ker \alpha = 2E(\mathbb{Q})$.
(c) The image $\operatorname{Im} \alpha$ is finite.

Note that the map α is well-defined since for any $x \in \mathbb{Q}$ we have that $x - \beta \neq 0$ in L. Again, as we shall see below, the proofs of (a) and (b) are "elementary", but the proof of (c) needs some theorems from algebraic number theory. Indeed, the proofs of (a) and (b) work in much more generality, and only (c) really uses that we are working with number fields.

Proof.

(a) We invoke 1.15.5. The condition (i) is automatic by definition. The condition (ii) asks us to show that if $P \in E(\mathbb{Q})$, then $\alpha(P) \cdot \alpha(-P) = [1]$ in $L^\times/(L^\times)^2$, but in fact this is clear because $\alpha(P) = \alpha(-P)$ and so $\alpha(P) \cdot \alpha(-P) = \alpha(P)^2 = [1]$, since squares are trivial in $L^\times/(L^\times)^2$. For condition (iii), we have to show using 1.12.18 that if $P_1, P_2, P_3 \in E(\mathbb{Q})$ are collinear (with multiplicities, as usual!), then $\alpha(P_1) \cdot \alpha(P_2) \cdot \alpha(P_3) = [1]$ in $L^\times/(L^\times)^2$. Now one of P_i, say P_1, is $\mathscr{O}$, then the result is clear from what we have done already, since then $P_2 = -P_3$. Therefore, we need only to worry about the case in which none of P_1, P_2, P_3 are $\mathscr{O}$.
When this is the case, the line ℓ through P_1, P_2, P_3 is not vertical, and so has the equation $\ell : y = \lambda x + \nu$ for some $\lambda, \nu \in \mathbb{Q}$. In the case, the x-coordinates x_1, x_2, x_3 of P_1, P_2, P_3 are precisely the

roots of the polynomial

$$g(x) := f(x) - (\lambda x + \nu)^2,$$

so that $g(x)$ factors as $g(x) = \prod_{i=1}^{3}(x - x_i)$. We wish to evaluate the product of the $x_i - \beta$, but this is precisely

$$\prod_{i=1}^{3}(x_i - \beta) = -\prod_{i=1}^{3}(\beta - x_i) = -g(\beta) = -f(\beta) + (\lambda\beta + \nu)^2 = (\lambda\beta + \nu)^2,$$

since $f(\beta) = 0$. Since none of the $x_i - \beta$ are zero, we have that $\lambda\beta + \nu \neq 0$ in L, which is enough to conclude that the product of the $x_i - \beta$ is a square in L. This says exactly that $\alpha(P_1)\alpha(P_2)\alpha(P_3) = [1]$ in $L^\times/(L^\times)^2$ as needed.

(b) Since α is a group homomorphism, and every element in the codomain squares to one, it follows automatically that $2E(\mathbb{Q}) \subset \ker\alpha$. Indeed, for any $P_1 \in 2E(\mathbb{Q})$, we can find a $P_0 \in E(\mathbb{Q})$ such that $P_1 = 2P_0$; then $\alpha(P_1) = \alpha(2P_0) = \alpha(P_0)^2 = [1]$, since again squares are trivial in $L^\times/(L^\times)^2$. The more interesting direction involves showing that $\ker\alpha \subset 2E(\mathbb{Q})$. To do this, we have to show that if $P_1 \in E(\mathbb{Q})$ is such that $\alpha(P_1) = [1]$, then there is a P_0 such that $P_1 = 2P_0$. If $P_1 = \mathscr{O}$, then we can just take $P_0 = \mathscr{O}$. Hence suppose that $P_1 \neq \mathscr{O}$, and write it as $P_1 = (x_1, y_1) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$. At this point, the proof is just messing around with the algebra. The condition $\alpha(P_1) = [1]$ is saying that there are $u, v, w \in \mathbb{Q}$ such that $x_1 - \beta = (u\beta^2 + v\beta + w)^2$. Now we must have $u \neq 0$, since β does not satisfy a nonzero linear or quadratic equation over $\mathbb{Q}$ (check!). Consider the $t \in \mathbb{Q}$ defined by

$$t := \frac{v}{u} - a.$$

Given this choice of t, if we expand the product $(t-\beta)(u\beta^2+v\beta+w)$ and use that $-\beta^3 = a\beta^2+b\beta+c$, we get that

$$\begin{aligned}(t - \beta)(u\beta^2 + v\beta + w) &= -u\beta^3 + (tu - v)\beta^2 + (tv - w)\beta + tw \\ &= (au + tu - v)\beta^2 + (bu + tv - w)\beta + (cu + tw).\end{aligned}$$

By our choice of t, the coefficient of β^2 in this expression vanishes, and so if we let $(\lambda, \nu) := (bu + tv - w, cu + tw)$, then we get that

$$(t - \beta)(u\beta^2 + v\beta + w) = \lambda\beta + \nu.$$

If we square this and use that $(u\beta^2 + v\beta + w)^2 = x_1 - \beta$, we get that β satisfies

$$(t - \beta)^2(x_1 - \beta) = (\lambda\beta + \nu)^2,$$

or equivalently that β is a root of the monic cubic equation

$$(x - x_1)(x - t)^2 + (\lambda x + \nu)^2.$$

However, $f(x)$ is the minimal polynomial of β, and so it is the *only* monic cubic polynomial that β satisfies (check!)–this means that we must have the equality of polynomials

$$f(x) = (x - x_1)(x - t)^2 + (\lambda x + \nu)^2 \in \mathbb{Q}[x].$$

At this point, we are basically done: if we let L be the line $L : y = \lambda x + \nu$, then the above equality tells us that the line L is tangent to E at the point $P_0 := (t, \lambda t + \nu)$ and passes through the point $P_1' := (x_1, \lambda x_1 + \nu)$, so that, in particular, $P_1' + 2P_0 = \mathscr{O}$ in $E(\mathbb{Q})$. However, there are precisely two points on E with x-coordinate exactly equal to x_1, namely $\pm P_1$, which forces $P_1' \in \{\pm P_1\}$. If $P_1' = -P_1$, then the equation $P_1' + 2P_0 = \mathscr{O}$ is saying precisely what we wanted–that $P_1 = 2P_0$. If, however, $P_1' = P_1$, then we have that $P_1 = 2(-P_0)$, so that P_1 is still the doubling of another rational point on $E(\mathbb{Q})$, finishing the proof.

(c) This is the part that does need a little algebraic number theory, so if you haven't seen this yet, you can return to this proof later. For those who are familiar with number fields, the idea is very simple: even though the group $L^\times/(L^\times)^2$ is not finite, we can only really have "finitely many primes" showing up in the image of α. To say this more precisely, consider for each finite set S of primes of $\mathscr{O}_L$ the subgroup

$$L(S, 2) := \{[a] \in L^\times/(L^\times)^2 : \text{ for all primes } \mathfrak{q} \notin S \text{ we have } v_{\mathfrak{q}}(a) \equiv 0 \pmod 2\}.$$

Standard finiteness results in algebraic number theory show that for each finite set S of primes, the subgroup $L(S,2)$ is finite (2.5.15). Therefore, if we can find a finite set S such that $\operatorname{Im}\alpha \subset L(S,2)$, then we would be done.
To find such an S, we need to study the primes dividing $x-\beta$ more carefully. For this, we first note that since $\beta \in \mathscr{O}_L \subset L$ is a root of $f(x)$, it can be factored as

$$f(x) = (x-\beta)\cdot q(x) = (x-\beta)(x^2+rx+s) \in L[x]$$

for some monic quadratic polynomial $q(x) = x^2+rx+s \in L[x]$. Since E is smooth, the polynomial f does not have a repeated root at β, i.e., $q(\beta) \neq 0$. Therefore, there are only finitely many primes dividing $q(\beta) \in \mathscr{O}_L$. If we let $S := \{\mathfrak{p} : \mathfrak{p} \mid q(\beta)\}$ be this finite set, then we will show that $\operatorname{Im}\alpha \subset L(S,2)$, finishing the proof.
It remains to show that if $P = (x,y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then $\alpha(P) \in L(S,2)$; in other words, that if $\mathfrak{p}$ is prime of $\mathscr{O}_L$ and $v_{\mathfrak{p}}(x-\beta) \equiv 1 \pmod 2$, then $\mathfrak{p} \mid q(\beta)$. To see this, write $P = (m/e^2, n/e^3)$ as in 1.14.8, so that $m,n,e \in \mathbb{Z}$ with $\gcd(m,e) = \gcd(n,e) = 1$. Then the fact that $P \in E(\mathbb{Q})$ can be expressed by the equation

$$n^2 = (m-e^2\beta)(m^2+rme^2+se^4) \tag{1.18.2}$$

obtained by clearing denominators. Now if $\mathfrak{p}$ is a prime of S such that $v_{\mathfrak{p}}(x-\beta) \equiv 1 \pmod 2$, then certainly also $v_{\mathfrak{p}}(m-e^2\beta) = v_p(x-\beta) + 2v_{\mathfrak{p}}(e) \equiv 1 \pmod 2$. The advantage of clearing denominators is that since $\beta \in \mathscr{O}_L$, we must have $v_{\mathfrak{p}}(m-e^2\beta) \geq 0$, so that if it is odd, we must have that $\mathfrak{p} \mid m-e^2\beta$, or equivalently

$$m \equiv e^2\beta \pmod{\mathfrak{p}}.$$

Now from the equation 1.18.2 it follows immediately that $v_{\mathfrak{p}}(m^2+rme^2+se^4) \equiv 1 \pmod 2$ as well, and hence as above that $\mathfrak{p} \mid m^2+rme^2+se^4$. Using that $m \equiv e^2\beta \pmod{\mathfrak{p}}$, this can be written as

$$\mathfrak{p} \mid e^4(\beta^2+r\beta+s) = e^4 q(\beta).$$

Now since $\mathfrak{p} \mid m-e^2\beta$, we cannot have that $\mathfrak{p} \mid e$; indeed, if $\mathfrak{p} \mid e$, then $\mathfrak{p} \mid m$ also, but $\gcd(m,e) = 1$. Therefore, $\mathfrak{p} \nmid e$, and so from $\mathfrak{p} \mid e^4 q(\beta)$, we conclude that $\mathfrak{p} \mid q(\beta)$, finishing the proof.

■

Example 1.18.3. Let's work through the example $E : y^2 = x^3 - 432$ of §1.2.B carefully. The polynomial $f(x) = x^3 - 432 \in \mathbb{Q}[x]$ is irreducible, and so we are indeed in case (c). The number field $L = \mathbb{Q}[\beta] := \mathbb{Q}[x]/(x^3-432)$ is isomorphic to $\mathbb{Q}[\sqrt[3]{2}]$ via the map $\beta \mapsto 6\sqrt[3]{2}$. For convenience, let us fix this isomorphism and let $\gamma := \sqrt[3]{2}$, so that $\beta := 6\gamma$.

(a) As above, we have a group homomorphism $\alpha : E(\mathbb{Q}) \to L^\times/(L^\times)^2$ given on points $P = (x,y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ by $P \mapsto [x-\beta]$.

(b) Consider the point $P_1 = (12,36) \in E(\mathbb{Q})$. Note that $\alpha(P_1) = [12-\beta]$. Since

$$12-\beta = (\gamma^2+2\gamma-2)^2,$$

we have that $\alpha(P_1) = [1]$. Let us use the theory above to construct a point P_0 such that $2P_0 = P_1$. We can take $(u,v,w) := (1/36, 1/3, -2)$, and hence $t = 12$, which gives $(\lambda,\nu) = (6,-36)$ (check!). Indeed, we have that

$$x^3 - 432 = (x-12)^3 + (6x-36)^2.$$

The line $y = 6x-36$ meets the curve E *thrice* at P_1, so that $P_0 = P_1' = P_1$ above and $P_1 = 2(-P_1)$, which is indeed correct and shows that $P_1 \in 2E(\mathbb{Q})$.

(c) We can factor $f(x)$ as $f(x) = (x-\beta)(x^2+\beta x+\beta^2)$, so that $q(x) = x^2+\beta x+\beta^2 \in L[x]$. In particular, $q(\beta) = 3\beta^2 = 108\gamma^2$, which factors as an ideal (check!) as

$$(q(\beta)) = (\gamma)^8(\gamma+1)^9.$$

Therefore, we can take $S := \{(\gamma),(\gamma+1)\}$. At this point, some algebraic number theory of L shows that $L(S,2)$ has size 16 (2.5.17), and hence that $E(\mathbb{Q})/2E(\mathbb{Q})$ injects into a group of size 16, as in in particular finite.

The above example shows that if we can get a better handle on $\operatorname{Im}\alpha$, then this can allow us to prove statements about rational points on the curve E. In particular, doing this for the curve $E : y^2 = x^3 - 432$ can allow us to prove Fermat's Last Theorem for $n = 3$. This is indeed the case. Since this is fun to figure out on your own, I won't spoil it; this argument can be found in the form of an extended exercise (2.5.17) on Exercise Sheet 5. This is all I am going to say about Fermat's Last Theorem for $n = 3$, although we will visit the case of $n = 4$ later.

We have now finished the proof of the Mordell Theorem 1.15.8 in the special case that $E[2](\mathbb{Q}) = \{\mathscr{O}\}$. Let's now discuss the other cases, the proofs of which will not use any serious algebraic number theory.

1.18.B The Case of Completely Split $f(x)$

Let's start this argument today; we'll finish it next time. Suppose now that $f(x) \in \mathbb{Z}[x]$ splits completely over $\mathbb{Q}$ into linear factors, and so can be written as

$$f(x) = (x - e_1)(x - e_2)(x - e_3)$$

for pairwise distinct integers $e_1, e_2, e_3 \in \mathbb{Z}$. In this case, if we let $T_i := (e_i, 0)$ for $i = 1, 2, 3$, then

$$E[2](\mathbb{Q}) = \{\mathscr{O}, T_1, T_2, T_3\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2$$

and the elliptic discriminant is

$$\Delta = 16(e_1 - e_2)^2(e_1 - e_3)^2(e_2 - e_3)^2.$$

We want to do something similar to what we did above. The key idea here is that, in fact, most of the proof above goes through.

Let $L := \mathbb{Q}[\beta] := \mathbb{Q}[x]/(f(x))$ as above, which this time is *not* a field, and in fact (thanks to the Chinese Remainder Theorem) splits as

$$L \xrightarrow{\sim} \mathbb{Q} \times \mathbb{Q} \times \mathbb{Q},$$

with the map given by sending the polynomial $g(\beta) \in L$ to the triple $(g(e_1), g(e_2), g(e_3))$. Now we can still ask for $L^\times$ (the group of *units* of L, and not the nonzero elements of L), and the group $L^\times/(L^\times)^2$, which again splits as

$$L^\times/(L^\times)^2 \xrightarrow{\sim} \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2.$$

Finally, the map

$$\alpha : E(\mathbb{Q}) \to L^\times/(L^\times)^2 \xrightarrow{\sim} \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$$

still *mostly* makes sense. Indeed, we still send $\mathscr{O}$ to $[1] \leftrightarrow ([1], [1], [1])$ of course, and a point $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ to $[x - \beta] \leftrightarrow ([x - e_1], [x - e_2], [x - e_3])$. This does make sense when $x \neq e_1, e_2, e_3$, but in fact something goes wrong when $x \in \{e_1, e_2, e_3\}$. Indeed, if say $P = T_1$, then the "second and third components" $e_1 - e_2$ and $e_1 - e_3$ still make sense, but the "first component" is "broken", since 0 is not allowed as an element of $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$. Therefore, we need to come up with some sort of patch. I encourage you to think about how to patch this until next time, when we will finish the proof of this case.

1.19 26/07/14 - Weak Mordell for Split Two Torsion

The first order of business today is to wrap up the proof of the weak Mordell theorem (1.15.9) in the case of full rational two-torsion.

Recall the set-up: $E/\mathbb{Q}$ is an elliptic curve in short Weierstrass form $y^2 = f(x)$, where $f(x) \in \mathbb{Z}[x]$ is a monic cubic polynomial that splits completely into linear factors over $\mathbb{Q}$ as $f(x) = \prod_{i=1}^{3}(x - e_i)$ for pairwise distinct $e_i \in \mathbb{Z}$. Again, we let $T_i = (e_i, 0)$ for $i = 1, 2, 3$ be the rational two-torsion points, and note that the elliptic discriminant is $\Delta = 16\prod_{1\le i<j\le 3}(e_i - e_j)^2$. Finally, note that T_i is the unique point on $E(\mathbb{Q})$ with x-coordinate e_i.

Continuing the discussion from last time (§1.18.B), we let $L := \mathbb{Q}[\beta] := \mathbb{Q}[x]/(f(x))$. As noted before, L is *not* a field; rather, it is a split algebra in the sense that the Chinese Remainder Theorem tells us that there is a $\mathbb{Q}$-algebra isomorphism

$$L \xrightarrow{\sim} \mathbb{Q} \times \mathbb{Q} \times \mathbb{Q}, \quad g(\beta) \mapsto (g(e_1), g(e_2), g(e_3)),$$

with the inverse map given by taking a triple (a_1, a_2, a_3) of rational numbers to the class $g(\beta)$ of the unique quadratic polynomial $g(x)$ satisfying $g(e_i) = a_i$ for $i = 1, 2, 3$; an explicit formula for g can be given by Lagrange interpolation, but we won't need it.[6]

From the above isomorphism we get also that $L^\times/(L^\times)^2 \xrightarrow{\sim} \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$, and as before we consider the map

$$\alpha : E(\mathbb{Q}) \to L^\times/(L^\times)^2 \xrightarrow{\sim} \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$$

given by taking $\mathscr{O}$ to $[1] \leftrightarrow ([1], [1], [1])$ and $P = (x, y) \in E(\mathbb{Q}) \smallsetminus E[2](\mathbb{Q})$ to $[x - \beta] \leftrightarrow ([x - e_1], [x - e_2], [x - e_3])$. We need to figure out what the right values of $\alpha(T_i)$ for $i = 1, 2, 3$ should be to make the same theorem as last time work in this setting. We will see what they need to be in

Theorem 1.19.1. In the above setup, consider the map

$$\alpha : E(\mathbb{Q}) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2 \times \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$$

given by

$$P \mapsto \begin{cases} ([1], [1], [1]), & \text{if } P = \mathscr{O}, \\ ([x - e_1], [x - e_2], [x - e_3]), & \text{if } P = (x, y) \in E(\mathbb{Q}) \smallsetminus E[2](\mathbb{Q}), \\ ([(e_1 - e_2)(e_1 - e_3)], [e_1 - e_2], [e_1 - e_3]), & \text{if } P = T_1, \\ ([e_2 - e_1], [(e_2 - e_1)(e_2 - e_3)], [e_2 - e_3]), & \text{if } P = T_2, \text{ and} \\ ([e_3 - e_1], [e_3 - e_2], [(e_3 - e_1)(e_3 - e_2)]), & \text{if } P = T_3. \end{cases}$$

(a) The map α is a group homomorphism.
(b) We have that $\ker \alpha = 2E(\mathbb{Q})$.
(c) The image $\operatorname{Im} \alpha$ is finite.

We will see in the proof below where the right choices of the values of the $\alpha(T_i)$ for $i = 1, 2, 3$ come from.

Proof.

(a) Of course, we invoke 1.15.5. As in the proof of 1.18.1, we are reduced to showing that if $P_1, P_2, P_3 \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ are collinear (as usual with multiplicities), then $\alpha(P_1) \cdot \alpha(P_2) \cdot \alpha(P_3) = ([1], [1], [1])$. First suppose that none of the P_i are two-torsion points, which is the same as saying that the x-coordinates x_i of the P_i satisfy $x_i \neq e_j$ for $i, j = 1, 2, 3$. The line ℓ joining the P_i is not vertical, and hence has equation $\ell : y = \lambda x + \nu$ for some $\lambda, \nu \in \mathbb{Q}$. Then x_i are the roots of the polynomial

$$g(x) := f(x) - (\lambda x + \nu)^2.$$

[6] This is the reason the Lagrange interpolating polynomial shows up in the proof of this result found in [13, §2.8].

We wish to show for $j = 1, 2, 3$ that the product $\prod_{i=1}^{3}(x_i - e_j)$ is a nonzero square, but indeed this product is

$$\prod_{i=1}^{3}(x_i - e_j) = -g(e_j) = (\lambda e_j + \nu)^2.$$

Since $T_j \notin \ell(\mathbb{Q})$, we have that $\lambda e_j + \nu \neq 0$, and so we are done.
Suppose now that some of the P_i coincide with the T_js. If all three did, then we must have (up to permutation) that $P_i = T_i$ for $i = 1, 2, 3$. This gives us the condition $\alpha(T_1)\alpha(T_2)\alpha(T_3) = ([1], [1], [1])$, and this is already enough to tell us what the "patches" should be; this is where the values of the $\alpha(T_i)$ for $i = 1, 2, 3$ come from. Let's now check that this works in the remaining cases too.
If exactly two of the P_is were T_js, say P_1 and P_2, then we could not have $P_1 = P_2$, since this would force $P_3 = \mathscr{O}$. Therefore, P_1 and P_2 must be distinct; up to relabelling, this again forces $P_i = T_i$ for $i = 1, 2, 3$, and we are back in the case above.
It only remains to deal with the case in which exactly one of the P_is coincides with the T_js, say $P_1 = T_1$ and $P_2, P_3 \in E(\mathbb{Q}) \smallsetminus E[2](\mathbb{Q})$, which means again that $x_i \neq e_j$ for $i = 2, 3$ and $j = 1, 2, 3$. In this case, if the line $L : y = \lambda x + \nu$ joining the P_i were to be horizontal (i.e., $\lambda = 0$), then $T_1 \in L(\mathbb{Q})$ would force $\nu = 0$ as well, and again we'd have $P_i = T_i$ for $i = 1, 2, 3$ (up to relabelling); since we are assuming we are not in this case, we must have that $\lambda \neq 0$. The fact that $T_1 \in \ell(\mathbb{Q})$ implies that $\nu = -\lambda e_1$. We conclude from this that the polynomial $g(x) = f(x) - (\lambda x + \nu)^2$ factors as $g(x) = (x - e_1)h(x)$, where

$$h(x) := (x - e_2)(x - e_3) - \lambda^2(x - e_1) = (x - x_2)(x - x_3).$$

Now let us compute the product $\alpha(P_1)\alpha(P_2)\alpha(P_3)$. In the second factor, this is precisely (the class of)

$$(e_1 - e_2)(x_2 - e_2)(x_3 - e_2) = (e_1 - e_2) \cdot h(e_2) = \lambda^2(e_1 - e_2)^2,$$

which is indeed a nonzero square. Similarly, the last factor is $\lambda^2(e_1 - e_3)^2$. Finally, the first factor is exactly

$$(e_1 - e_2)(e_1 - e_3) \cdot (x_2 - e_1)(x_3 - e_1) = h(e_1)^2.$$

Again, we know that $h(e_1) \neq 0$, and we are done.

(b) This part of the proof is not just similar to that of 1.18.1(b); it is identical. Indeed, the point is that the proof there did not really use that L is a number field. Let's breeze through the proof again. As before, $\ker\alpha \subset 2E(\mathbb{Q})$ is clear. If $P_1 = (x_1, y_1) \in \ker\alpha \smallsetminus \{\mathscr{O}\}$, then there are $u, v, w \in \mathbb{Q}$ with $x_1 - \beta = (u\beta^2 + v\beta + w)^2$ and $u \neq 0$. As before, we let $t := (v/u) - a$, where $a = -(e_1 + e_2 + e_3)$, and we let λ, ν be determined by $(t - \beta)(u\beta^2 + \nu\beta + w) = \lambda\beta + \nu$. The polynomial $(x - x_1)(x - t)^2 + (\lambda x + \nu)^2$ is a monic cubic polynomial satisfied by $\beta \in L$, but this implies that this must be equal to $f(x)$ by properties of L. The rest of the proof is identical.[7]

(c) This part of the proof is even easier than last time. Again, the goal is to find a finite set S of (rational!) primes such that if $\mathbb{Q}(S, 2) \subset \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ is the subgroup generated by -1 and the $p \in S$, then $\operatorname{Im}\alpha \subset \mathbb{Q}(S, 2) \times \mathbb{Q}(S, 2) \times \mathbb{Q}(S, 2)$. We claim that we can take S to be the set of primes dividing Δ.
Since this is clear for $\alpha(T)$ for $T \in E[2](\mathbb{Q})$, we only need to show that if $P = (x, y) \in E(\mathbb{Q}) \smallsetminus E[2](\mathbb{Q})$ and p is a prime such that for some j we have $v_p(x - e_j) \equiv 1 \pmod 2$, then $p \mid \Delta$. Without loss of generality, suppose that $v_p(x - e_1) \equiv 1 \pmod 2$. Then the equation $y^2 = \prod_{j=1}^{3}(x - e_j)$ forces that $v_p(x - e_j) \equiv 1 \pmod 2$ for at least one of $j = 2, 3$. Without loss of generality, suppose this true for $j = 2$.
As in the proof of 1.18.1, we now clear denominators using 1.14.8 to write $P = (m/e^2, n/e^3)$ for some $m, n, e \in \mathbb{Z}$ with $e \geq 1$ and $\gcd(m, e) = \gcd(n, e) = 1$. Then $P \in E(\mathbb{Q})$ translates to

$$n^2 = (m - e^2e_1)(m - e^2e_2)(m - e^2e_3).$$

The condition $v_p(m - e^2e_j) \equiv 1 \pmod 2$ for $j = 1, 2$ now implies that $p \mid m - e^2e_j$ for $j = 1, 2$, which forces that $p \mid e^2(e_1 - e_2)$. But again $p \nmid e$ because $p \mid m - e^2e_1$ and $\gcd(m, e) = 1$. Therefore, we conclude that $p \mid e_1 - e_2 \mid \Delta$ as needed.

■

[7] If you wish, you can phrase this proof without using properties of L and rather using the Lagrange interpolating polynomial to translate directly between $E(\mathbb{Q})$ and $(\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3$ without referring to L at all, as is done in [13, §2.8].

This finishes the proof of 1.19.1, and hence 1.15.9 and 1.15.8 in the case of full rational two-torsion.

Remark 1.19.2.

(a) Note that in the proof of 1.19.1, the image of α is contained in the "hyperplane" in $(\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^3$ consisting of triples for which the product of the coordinates is trivial (i.e., a square). This allows us to express α (up to non-canonical choices) as a map to $(\mathbb{Q}^\times/(\mathbb{Q}^\times)^2)^2$ instead (as is done in [8, §X.1]). I don't like this formulation because it hides the symmetry between the e_i evident in the statement in 1.19.1.

(b) One might rightfully ask where the maps α in the 1.18.1 and 1.19.1 *really* come from. The answer to this question is that they come from Galois cohomology. This is beyond the scope of this course, but the interested reader can find more about this in [8, Chapters VIII, X].

(c) Technically, to finish the proof of 1.15.9 and hence 1.15.8, we still need to deal with one more case: the case in which $f(x)$ has exactly one rational root. This can be done very similarly to 1.18.1 above (as is done in [1, §15]); I leave this to the interested reader. We will follow a different approach–namely that of descent via two-isogeny, as in [1, §14] and [2, §3.4-3.5]–to give a proof of this case (and a second proof of the case of full two-torsion) next time.

(d) To what extent does the proof of 1.19.1 (and even 1.18.1) rely on the fact that our base field is $\mathbb{Q}$? We only really needed this fact in the statement (c) (of both theorems). Therefore, (i) the proofs are basically correct as stated for any number field K in place of $\mathbb{Q}$, and (ii) we have proven something quite general in parts (a) and (b). For instance, we have shown already the result of 1.19.3.

Corollary 1.19.3. Let K be a field of characteristic not two, $e_1, e_2, e_3 \in K$ be pairwise distinct, and let E/K be the elliptic curve with Weierstrass equation $y^2 = \prod_{i=1}^{3}(x - e_i)$.

(a) A point $P = (x, y) \in E(K) \smallsetminus E[2](K)$ lies in $2E(K)$ iff each of $x - e_1, x - e_2, x - e_3$ are squares in K.

(b) The point $T_1 = (e_1, 0)$ lies in $2E(K)$ iff $e_1 - e_2$ and $e_1 - e_3$ are squares in K. Similar results hold for the points $T_i = (e_i, 0)$ for $i = 2, 3$.

In particular, if K is an algebraically closed field of characterstic other than 2, and E/K an elliptic curve, then the duplication map $[2] : E(K) \to E(K)$ is surjective.

Proof. This follows from the proof of parts (a) and (b) of 1.19.1 above. The second statement follows from the first since every element in an algebraically closed field is a square. ∎

Unimportant Remark 1.19.4. In fact, it can be shown that if K is *any* algebraically closed field, E/K an elliptic curve, and $n \in \mathbb{Z}_{\geq 1}$, then the *multiplication-by-n* map $[n] : E \to E$ is surjective on K points (combine [8, II.2.3, III.4.2(a)] with the fact that surjectivity of this map of schemes is equivalent to surjectivity on K-points–when K is algebraically closed [TODO]).

Corollary 1.19.5. Let K be a field of characteristic other than 2 such that -1 is not a square in K (e.g., $K = \mathbb{Q}$ or $K = \mathbb{F}_p$ for p a prime with $p \equiv 3 \pmod 4$). If E/K is an elliptic curve, then $E[4](K) \ncong_{\mathsf{Ab}} \mathbb{Z}/4 \oplus \mathbb{Z}/4$, i.e., it is not possible for all the 4-torsion on E to be defined over K.

Proof. Suppose that K is a field of characteristic other than two, and E/K is an elliptic curve such that $E[4](K) \cong_{\mathsf{Ab}} \mathbb{Z}/4 \oplus \mathbb{Z}/4$. In particular, $E[2](K) \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2$, and hence E has fully split two-torsion over K. This means that E can be put into Weierstrass form $y^2 = f(x)$ where $f(x) = (x - e_1)(x - e_2)(x - e_3)$ for some pairwise distinct $e_i \in K$. Since $E[4](K) \cong_{\mathsf{Ab}} \mathbb{Z}/4 \oplus \mathbb{Z}/4$, at least two of the points $T_i = (e_i, 0)$ for $i = 1, 2, 3$ belong to $2E(K)$; without loss of generality, suppose these are T_1 and T_2. Since $T_1 \in 2E(K)$, 1.19.3(b) tells us that $e_1 - e_2$ is a square in K. Since $T_2 \in 2E(K)$, 1.19.3(b) tells us that $e_2 - e_1$ is a square in K. Since $e_1 \neq e_2$, these facts combined tell us that -1 is a square in K. ∎

Unimportant Remark 1.19.6. The argument of 1.19.5 can be generalized to prove that if K is a field, E/K an elliptic curve, and $n \in \mathbb{Z}_{\geq 1}$ such that $E[n](K) \cong_{\mathsf{Ab}} \mathbb{Z}/n \oplus \mathbb{Z}/n$, then in fact K contains a

primitive n^{th} root of unity. This shows in particular that if $C \subset \mathbb{P}^2_{\mathbb{R}}$ is a smooth cubic curve, then not all of its 9 complex inflection points can be real (which can also be proven by other means; see 2.2.9(e) and 2.4.12(b)). The cleanest modern approach to proving this involves using the Weil pairing. We will not explain what that is, but the interested reader can find this topic in [8, §III.8].

1.20 26/07/15 - Descent via Two-Isogeny

The following discussion has been adapted from [1, §14] and [2, §3.4].

We will now give another proof of 1.15.9 for elliptic curves E over $\mathbb{Q}$ under the additional hypothesis that there is at least one nontrivial rational 2-torsion point $T \in E[2](\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$. Up to an admissible change of coordinates, we can then assume that $T = (0,0)$ and E has the Weierstrass form

$$E_{a,b} : y^2 = x(x^2 + ax + b)$$

for some $a, b \in \mathbb{Z}$. The discriminant and j-invariant of $E_{a,b}$ are easily seen to be (c.f. 1.9.21)

$$\Delta(E_{a,b}) = 16b^2(a^2 - 4b) \text{ and } j(E_{a,b}) = 256\frac{(a^2-3b)^3}{b^2(a^2-4b)}.$$

so $E_{a,b}$ defines an elliptic curve iff $b \neq 0$ and $a^2 - 4b \neq 0$. Let us make some easy observations.

Observation 1.20.1. In the above setup,

(a) the point $T = (0,0) \in E_{a,b}(\mathbb{Q})$ is the only point on $E_{a,b}$ with x-coordinate 0, and
(b) the 2-torsion on $E_{a,b}$ is fully split (i.e., we have that $E_{a,b}[2](\mathbb{Q}) \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2$) iff $x^2 + ax + b \in \mathbb{Q}[x]$ has two rational roots, which happens iff $a^2 - 4b$ is a perfect square.

In fact, we didn't really use any properties of $K = \mathbb{Q}$ so far other than $\operatorname{ch} K \neq 2$.

We wish to study the duplication map $[2] : E \to E$ on E and say something about the cokernel $\operatorname{coker}([2] : E(\mathbb{Q}) \to E(\mathbb{Q}))$. To do this, we'll break up the duplication map $[2] : E(\mathbb{Q}) \to E(\mathbb{Q})$ into two simpler maps. Let's take a second to review some general algebra formalism that will enable us to do this.

1.20.A Quick Algebra Review

Let's review some basic terminology involving exact sequences.

Definition 1.20.2. Let $A \xrightarrow{f} B \xrightarrow{g} B$ be a sequence of homomorphisms of abelian groups.

(a) We say that this sequence is a *complex* iff $gf = 0$, which happens iff $\operatorname{Im} f \subset \ker g$.
(b) We say that this sequence is an *exact complex* or *exact at* B iff $\operatorname{Im} f = \ker g$.

A longer sequence of homomorphisms of abelian groups is said to be a complex if the composite of every consecutive pair of morphisms is zero, and it is said to be exact iff it is exact at each term. We will need two basic facts about these objects, which are summarized in

Proposition 1.20.3.

(a) Suppose that $A \xrightarrow{f} B \xrightarrow{g} C$ is an exact sequence. If A and C are finite, then so is B.
(b) Suppose that $A \xrightarrow{f} B \xrightarrow{g} C$ is a sequence of homomorphisms of abelian groups (not necessarily a complex). Then there is an exact sequence of the form

$$0 \to \ker(f) \to \ker(gf) \to \ker(g) \to \operatorname{coker}(f) \to \operatorname{coker}(gf) \to \operatorname{coker}(g) \to 0.$$

Proof. I will leave the proofs of both of these facts as an exercise (2.5.11), and only remark here that the proof of (a) uses nothing more than the First Isomorphism Theorem. ∎

The above proposition suggests a strategy for studying (co)kernels of complicated morphisms. Suppose that $h : A \to C$ is a homomorphism of abelian groups, and we are trying to show that $\ker(h)$ (resp. $\operatorname{coker}(h)$) is finite. One strategy for how we might do this involves breaking h up into two "simpler" maps $A \xrightarrow{f} B$ and $B \xrightarrow{g} C$ such that $h = gf$. If we can then show that $\ker(f)$ and $\ker(g)$ are finite (resp. $\operatorname{coker}(f)$ and $\operatorname{coker}(g)$ are finite), then 1.20.3 will be enough to conclude that $\ker(h)$ (resp. $\operatorname{coker}(h)$) is finite.

1.20.B Isogenies

Let's carry out the strategy of §1.20.A for the duplication map $h = [2] : E(\mathbb{Q}) \to E(\mathbb{Q})$. We will start by giving a definition.

Definition 1.20.4 (Isogenies). Let K be a field, and let E_1 and E_2 be elliptic curves over K.

(a) An *isogeny defined over* K (also known as K-*isogeny*) from E_1 to E_2 is a morphism $\phi : E_1 \to E_2$ such that $\phi(\mathscr{O}_1) = \mathscr{O}_2$, where $\mathscr{O}_i$ denotes the origin in E_i for $i = 1, 2$.
(b) We say that E_1 and E_2 are *isogenous over* K (also known as K-*isogenous*) iff there is a nonconstant isogeny $E_1 \to E_2$.

Clearly, the composite of two isogenies is an isogeny. Here's an important

Remark 1.20.5. We haven't quite defined what a morphism of curves is in this course, but for our purposes it suffices to think of a morphism as being a map given by some polynomial functions (or rational functions, when defined).[8] The precise definition of a morphism can be found in, say, [8, Ch. I]; we will not need it, since our isogenies will be incredibly explicit.

The above definition is slightly strange, but before we say anything more, let's give a few examples.

Example 1.20.6. Let K be a field and $E = E_1$ and E_2 be elliptic curves over K.

(a) The constant map $E_1 \to E_2$ with value $\mathscr{O}$ is always an isogeny, known as the 0-isogeny.
(b) The identity map $[1] : E \to E$ is an isogeny.
(c) The duplication map $[2] : E \to E$ is an isogeny. This is clear by the general form of the duplication formula (1.12.13) (although what happens with the y-coordinate needs a little bit of thought).
(d) Indeed, for any integer $n \in \mathbb{Z}$, the multiplication-by-n morphism $[n] : E \to E$ is an isogeny, as is easily seen by a quick induction using the ideas of 1.12.D (see also 2.2.10). The cases $n = 0, 1, 2$ correspond to (a), (b), and (c) above.
(e) An admissible change of coordinates $\phi : E_1 \to E_2$ is an isogeny. In particular, automorphisms are isogenies. This already shows that there is a difference in general between isogenies defined over a field K and those defined over a bigger field (say $\overline{K}$) in general.

For an example of a map of elliptic curves that is *not* an isogeny, consider the example of translation $\tau_P : E \to E$ by a nontrivial point $P \in E(K) \smallsetminus \{\mathscr{O}\}$; this is a morphism, but it is not an isogeny because it doesn't preserve $\mathscr{O}$.

We'll give many more examples soon, but two can also be found on the exercise sheets (2.5.13 and 2.5.14). Let's now list some facts about isogenies (without proof).

Fact 1.20.7.

(a) An isogeny of elliptic curves $\phi : E_1 \to E_2$ is automatically a group homomorphism. More precisely, if $\phi : E_1 \to E_2$ is an isogeny over K, then for each extension L/K, the map $\phi(L) : E_1(L) \to E_2(L)$ induced by ϕ on the set of L-points is a group homomorphism. (This is very surprising–the fact that *any* polynomial map preserving $\mathscr{O}$ is automatically a homomorphism for the complicated group operation obtained from the chord-and-tangent process! This is part of some sort of "rigidity" package for elliptic curves.)
(b) The relation of being isogenous over K is an equivalence relation. Reflexivity and transitivity are basically clear (one has to argue that a composite of nonconstant isogenies is nonconstant, but this is not hard). What is more surprising is that this relationship is also symmetric; this is harder. Therefore, you should think of two isogenous curves as being "similar" to each other. This is ultimately where the name "isogeny" comes from (iso = same, genus = type).
(c) Associated to every isogeny $\phi : E_1 \to E_2$ is an integer called its *degree*; an isogeny of degree $n \in \mathbb{Z}_{\geq 0}$ is said to be an n-*isogeny*. Roughly speaking, the degree of an isogeny is the size

[8] It would be somewhat strange if in our algebraic theory of curves we somehow had the exponential function showing up!

of its kernel (as measured over an algebraic closure $\overline{K}$ of K), i.e., $\deg \phi = \#E_1[\phi](\overline{K})$. (This is not strictly correct; it is correct in characteristic zero, but not in positive characteristic due to the presence of purely inseparable field extensions.) This degree is multiplicative: if $\phi : E_1 \to E_2$ and $\psi : E_2 \to E_3$ are isogenies, then so is $\psi \circ \phi$ and

$$\deg(\psi \circ \phi) = \deg(\psi) \cdot \deg(\phi).$$

We will not prove these facts, and hence we will also not use them in what follows. However, these results are good to know and will serve as important motivation for what we do below. The interested reader can find a proof of these facts in [8].

Now let $K = \mathbb{Q}$ (actually we will only use that $\operatorname{ch} K \neq 2$), let E/K be an elliptic curve, and consider the duplication map $[2] : E \to E$. By 1.20.6(c), this is an isogeny. By 1.20.7(c) and 1.12.7, we have that $\deg[2] = 4$. Therefore, if we can "break up" $[2]$ into two steps, i.e., as a composite of two isogenies $\phi : E \to E'$ and $\phi' = \hat{\phi} : E' \to E$, each of degree 2, then we could hope to apply the strategy of §1.20.A to reduce the proof of 1.15.9 to showing that $\operatorname{coker} \phi(K)$ and $\operatorname{coker} \phi'(K)$ are finite. This procedure is known as "descent by two-isogeny", and this is exactly what we will do.

At this point, I might as well give you some explicit formulae. Let $a, b \in K$ with $b(a^2 - 4b) \neq 0$, and let $E_{a,b} : y^2 = x(x^2 + ax + b)$ as above, with 2-torsion point $T = (0, 0) \in E[2](K) \smallsetminus \{\mathscr{O}\}$. We define $a', b' \in K$ by the formula

$$(a', b') := (-2a, a^2 - 4b), \tag{1.20.8}$$

and we let E' be the elliptic curve $E' = E_{a',b'} : y'^2 = (x')^3 + a'(x')^2 + b'x'$, which has its own two-torsion point $T' = (0, 0) \in E'[2](K) \smallsetminus \{\mathscr{O}'\}$. Next, we let $\phi : E \to E'$ be defined by the formula

$$(x, y) \mapsto \left(x + a + \frac{b}{x}, y \cdot \left(1 - \frac{b}{x^2} \right) \right)$$

when $x \neq 0$ and by $\mathscr{O}, T \mapsto \mathscr{O}'$. Note that since the point (x, y) lies on E, we can also write the first coordinate of this map as

$$x + a + \frac{b}{x} = \frac{y^2}{x^2}.$$

Finally, we let $\phi' : E' \to E$ be defined by essentially the same formula except for a few powers of 2; more precisely, by the formula

$$(x', y') \mapsto \left(\frac{1}{4} \left(x' + a' + \frac{b'}{x'} \right), \frac{1}{8} \cdot y' \cdot \left(1 - \frac{b'}{(x')^2} \right) \right)$$

when $x' \neq 0$ and by $\mathscr{O}', T' \mapsto \mathscr{O}$. We will soon show that these formulae work and that $\phi' \circ \phi = [2]_E$ (1.20.13), but before that, I should probably spend some time explaining *where* these formulae come from.

One quick note before we do that: if we iterate the map 1.20.8, then we get immediately that

$$(a'', b'') := (-2a', (a')^2 - 4b') = (4a, 16b).$$

Therefore, the resulting curve $E'' : (y'')^2 = (x'')^3 + a''(x'')^2 + b''$ is isomorphic to E via the map $(x'', y'') \mapsto (x''/4, y''/8)$. In other words, the map ϕ' is precisely the same kind of map out of E' that ϕ is out of E, just composed with the isomorphism $E'' \xrightarrow{\sim} E$ to take values in E. This means that ϕ and ϕ' play very symmetric roles. In particular, if we can show that $\operatorname{coker} \phi'(K) = E(K)/\phi'(E'(K))$ is finite, this would automatically mean by symmetry that $\operatorname{coker} \phi(K)$ is finite as well. If we can further prove that $\phi' \circ \phi = [2]_E$, then we have reduced the more difficult problem of showing that $\operatorname{coker}[2](K) = E(K)/2E(K)$ is finite to the simpler problem of showing that $\operatorname{coker} \phi'(K)$ is finite.

1.20.C Motivational Aside

It is now time to explain where the above formulae actually come from. Suppose now that we are working over $K = \mathbb{C}$, and that $E/\mathbb{C}$ is an elliptic curve, which by the discussion in 1.13.A looks like $\mathbb{C}/\Lambda$ for

some lattice $\Lambda \subset \mathbb{C}$. Up to scaling Λ (i.e., homothety, which doesn't change the isomorphism type of the quotient $\mathbb{C}/\Lambda$), we might as well assume that it is generated by 1 and τ with $\operatorname{Im}\tau > 0$, so that $\Lambda = \mathbb{Z} \oplus \mathbb{Z}\tau \subset \mathbb{C}$. As in 1.13.A, we can "visualize" the points of $E(\mathbb{C}) \cong \mathbb{C}/\Lambda$ as being the points of a fundamental parallelogram in the lattice Λ, e.g., the parallelogram generated by the vectors 1 and τ.

We are interested in studying the duplication map

$$[2] : E(\mathbb{C}) \to E(\mathbb{C}).$$

The first thing to note is that this map is surjective. This is clear in the plane-mod-lattice picture, since this map just maps each "quarter" of the fundamental parallelogram in the domain isomorphically to the fundamental parallelogram in the codomain. Algebraically, this is also a consequence of 1.19.3, since $K = \mathbb{C}$ is algebraically closed. Since the duplication map $[2] : E(\mathbb{C}) \to E(\mathbb{C})$ is a surjective group homomorphism, it is essentially given by the quotient by its kernel, which in this case is $E[2](\mathbb{C}) = \{\mathscr{O}, T_1, T_2, T_3\}$, where T_1, T_2, T_3 correspond to the points $1/2, \tau/2$, and $(1+\tau)/2$ in the fundamental parallelogram. This subgroup is evidently isomorphic to $\mathbb{Z}/2 \oplus \mathbb{Z}/2$.

But now, instead of directly taking the quotient by $E[2](\mathbb{C})$, we could break this operation up into two steps. We could first take the quotient by the subgroup $\{\mathscr{O}, T\}$ where $T = T_2$, and then take the quotient by the image of $E[2](\mathbb{C})$ in that. Geometrically, this amounts to considering the lattice $\Lambda' := \mathbb{Z} \oplus \mathbb{Z}\tau'$, where $\tau' := \tau/2$ and the "vertical slide" map $\phi : \mathbb{C}/\Lambda \to \mathbb{C}/\Lambda'$, which gives rise to a "degree 2" map of the corresponding elliptic curves $\phi : E(\mathbb{C}) \to E'(\mathbb{C})$, where $E'(\mathbb{C}) \cong \mathbb{C}/\Lambda'$. After having done this, we would then consider the lattice $\Lambda'' := (1/2)\mathbb{Z} \oplus \mathbb{Z}\tau'$ and the "horizontal slide" map $\psi : \mathbb{C}/\Lambda' \to \mathbb{C}/\Lambda''$, which would give rise to another "degree 2" map of the corresponding elliptic curves $\psi : E'(\mathbb{C}) \to E''(\mathbb{C})$, where $E''(\mathbb{C}) \cong \mathbb{C}/\Lambda''$. Finally, we would note that the lattices Λ'' and Λ are homothetic, and so a simple scaling by two gives us an isomorphism $\mathbb{C}/\Lambda'' \xrightarrow{\sim} \mathbb{C}/\Lambda$, which corresponds to the isomorphism $E''(\mathbb{C}) \to E(\mathbb{C})$ given essentially by $(x'', y'') \mapsto (x''/4, y''/8)$. The composite $\mathbb{C}/\Lambda' \to \mathbb{C}/\Lambda$ of ψ with this isomorphism is what we are calling ϕ', and this corresponds exactly to the map $\phi' : E'(\mathbb{C}) \to E(\mathbb{C})$. Geometrically, it is evident that the composition $\phi' \circ \phi$ is multiplication by 2 on $E(\mathbb{C})$, since that's what we designed this composite operation to be. See 1.20.9 for an illustration.

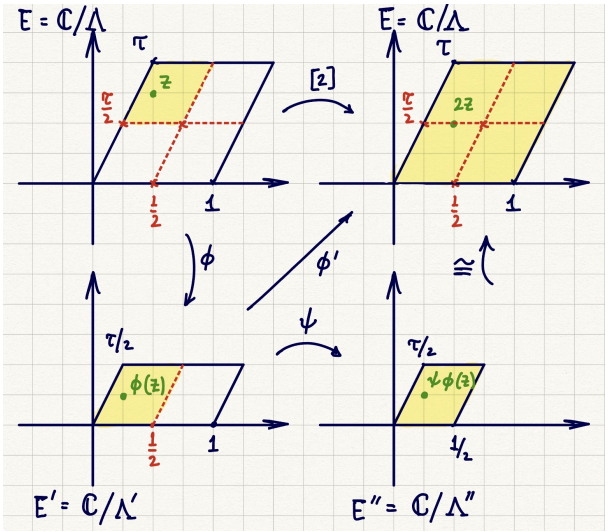


Figure 1.20.9: A diagrammatic representation of the two-isogenies ϕ and ϕ'.

Let's now reconcile this picture with the algebra, so again we have $a, b \in K$ with $b(a^2 - 4b) \neq 0$ and E/K is the elliptic curve $y^2 = x^3 + ax^2 + bx$. This has a 2-torsion point $T = (0,0) \in E[2](K) \smallsetminus \{\mathscr{O}\}$, we wish to take the "quotient" of E by the translation-by-T map $\tau_T : E \to E$. This warrants an exercise.

Exercise 1.20.10. In the above setup, show that the translation-by-T map $\tau_T : E(K) \to E(K)$ is given by the formula

$$P \mapsto P + T = \begin{cases} T & \text{if } P = \mathscr{O}, \\ \mathscr{O} & \text{if } P = T, \text{ and} \\ \left(\dfrac{b}{x}, -\dfrac{by}{x^2}\right) & \text{if } P = (x,y) \in E(K) \smallsetminus \{\mathscr{O}, T\}. \end{cases}$$

Since T is a 2-torsion point, τ_T is an involution, as is also clear from the formula.

If the quotient $\phi : E \to E'$ by $\{\mathscr{O}, T\}$ is given by the formula $\phi(x, y) = (\xi, \eta)$, where $\xi, \eta \in K(E)$, then we would certainly need ϕ to have the property that $\phi(P) = \phi(P + T)$ for all P. In particular, we would need ξ and η to be invariant under the operation τ_T. This suggests what we might take ξ and η to be, as functions of P: we could try, for instance,

$$\xi(P) := x(P) + x(P + T) + a \text{ and } \eta(P) := y(P) + y(P + T). \tag{1.20.11}$$

(The additional a in the formula for ξ is simply to make the formulae a little cleaner, and serves no essential function.) Rewriting these in terms of 1.20.10 gives us precisely the formulae

$$(\xi, \eta) = \left(x + \frac{b}{x} + a, y - \frac{by}{x^2}\right).$$

Now it is a simple check to see that these quantities satisfy the equation

$$\eta^2 = \xi^3 - 2a\xi^2 + (a^2 - 4b)\xi.$$

Exercise 1.20.12. Check this! You will need to use that $y^2 = x^3 + ax^2 + bx$.

This is where the formulae for the elliptic curve E' and the map ϕ come from. Again, it is clear from the geometric picture (1.20.9) that repeating this procedure should just return E up to a change of coordinates involving scaling by $u = 2$, and this is indeed the case, as can be checked easily algebraically.

All of the above discussion was by way of motivation; we will not use these ideas (especially those involving complex lattices) in any serious way. Let's now return to the algebraic formalism.

1.20.D Back to the Properties of Our Two-Isogenies

Let's summarize what we want in

Theorem 1.20.13. Let K be a field of characteristic other than 2, and let $a, b \in K$ with the property $b(a^2 - 4b) \neq 0$. Let $E = E_{a,b}$ be the elliptic curve over K given by $E : y^2 = x^3 + ax^2 + bx$. Let $T := (0, 0) \in E[2](K) \smallsetminus \{\mathscr{O}\}$.

Let $a', b' \in K$ be defined by $(a', b') = (-2a, a^2 - 4b)$, and let E'/K be the elliptic curve $E' : y^2 = x^3 + a'x^2 + b'x$, and again let $T' := (0, 0) \in E'[2](K) \smallsetminus \{\mathscr{O}'\}$.

Consider the map $\phi : E \to E'$ which on K-points is given by the formula

$$P \mapsto \phi(P) := \begin{cases} \mathscr{O}' & \text{if } P = \mathscr{O}, T, \text{ and} \\ \left(x + a + \dfrac{b}{x}, y\left(1 - \dfrac{b}{x^2}\right)\right) & \text{if } P = (x, y) \in E(K) \smallsetminus \{\mathscr{O}, T\}, \end{cases}$$

and indeed by the same formula for every field extension L/K. Similarly, consider the map $\phi' : E' \to E$ which on K-points is given by the formula

$$P' \mapsto \phi'(P') := \begin{cases} \mathscr{O} & \text{if } P' = \mathscr{O}', T', \text{ and} \\ \left(\dfrac{1}{4}\left(x' + a' + \dfrac{b}{x}\right), \dfrac{1}{8} \cdot y'\left(1 - \dfrac{b'}{(x')^2}\right)\right) & \text{if } P' = (x', y') \in E'(K) \smallsetminus \{\mathscr{O}', T'\}, \end{cases}$$

and indeed by the same formula for every field extension L/K. Then:

(a) The maps ϕ and ϕ' are well-defined, i.e., the formula for ϕ above does take points on E to E', and similarly for ϕ'.
(b) The map $\phi(K) : E(K) \to E(K)$ induced by ϕ is a group homomorphism with kernel $\{\mathscr{O}, T\}$ (and the same is true for K replaced by any extension L/K). A similar result is true for ϕ'.
(c) We have that $\phi' \circ \phi = [2]_E$, and similarly $\phi \circ \phi' = [2]_{E'}$.

Proof. Note that the formulae for ϕ and ϕ' at least make sense thanks to 1.20.1(a).

(a) This is just a simple algebraic check which is left to the reader (1.20.12 and 1.20.14).

(b) This follows from 1.20.7(a), but since we haven't proven that, we will give a direct proof. Invoking 1.15.5 as in the proofs of 1.18.1 and 1.19.1, it suffices to show that if $P_1, P_2, P_3 \in E(K) \smallsetminus \{\mathscr{O}\}$ are collinear (with multiplicities), then so are $\phi(P_1), \phi(P_2), \phi(P_3)$. Suppose first that none of the P_i are T, and that the line ℓ containing the P_i (which is not vertical) has the equation $\ell : y = \lambda x + \nu$ for some $\lambda, \nu \in K$. Note that since $T \notin \ell(K)$, we must have $\nu \neq 0$. Our goal is to produce $\lambda', \nu' \in K$ such that if ℓ' is the line with equation $\ell' : y' = \lambda' x' + \nu'$, then $\phi(P_1), \phi(P_2)$, and $\phi(P_3)$ are precisely the intersection points of ℓ' with E'. We claim that it suffices to take

$$(\lambda', \nu') := \left(\lambda - \frac{b}{\nu}, \nu - a\lambda + b\frac{\lambda^2}{\nu}\right).$$

It is a straightforward algebraic check to see that this works, and this is left as an exercise (1.20.14); however, see also 1.20.15 below. Finally, it remains only to deal with the case in which exactly one of the P_is is T, say $P_1 = T$ and $P_2, P_3 \notin T$. We have to show in this case that $T + P_2 + P_3 = \mathscr{O}$ in $E(K)$ implies that $\phi(T) + \phi(P_2) + \phi(P_3) = \mathscr{O}'$ in $E'(K)$. Since $\phi(T) = \mathscr{O}'$ by definition, this is equivalent to showing that $\phi(P_2) = -\phi(P_3)$. But indeed, since ϕ obviously plays well with negatives, we have that

$$\phi(P_2) = \phi(-(P_3 + T)) = -\phi(P_3 + T) = -\phi(P_3),$$

where the last equality follows simply because evidently $\phi(P_3) = \phi(P_3 + T)$ (check; see 1.20.10 and 1.20.11).

(c) One can check easily using the results of §1.12.D (e.g., 1.12.13) that both the maps $\phi' \circ \phi$ and $[2]_E$ are given on K-points by the formulae

$$P \mapsto \begin{cases} \mathscr{O} & \text{if } P = \mathscr{O} \text{ or } y(P) = 0, \text{ and} \\ \left(\dfrac{(x^2-b)^2}{4y^2}, \dfrac{(x^2-b)(x^4+2ax^3+6bx^2+2abx+b^2)}{8y^3}\right) & \text{if } P = (x,y) \in E(K) \smallsetminus E[2](K), \end{cases}$$

and indeed that this recipe works on L points for any extension L/K. The proof of the second fact is symmetric. We leave the details to the reader (1.20.14).

■

Exercise 1.20.14. Fill in the gaps in the above proof, i.e., perform all the algebraic checks left to the reader. (Hint: If you get stuck, it may help to look at [2, §3.4].)

Remark 1.20.15. The formulae for (λ', ν') in the proof of part (b) of 1.20.13 may seem to have come out of nowhere. Let me try to explain *how* one might come up with such formulae (which the authors of [2] do not do), so suppose we are in the setting there: $P_1, P_2, P_3 \in E(K) \smallsetminus \{\mathscr{O}\}$ are points which form the intersection of E with the line $\ell : y = \lambda x + \nu$. Let $P_i = (x_i, y_i)$ and $\phi(P_i) = (x_i', y_i')$ for $i = 1, 2, 3$, so that $x_1 x_2 x_3 \neq 0$ and for each i we have

$$x_i' = x_i + a + \frac{b}{x_i} \text{ and } y_i' = y_i\left(1 - \frac{b}{x_i^2}\right).$$

Then x_1, x_2, x_3 are roots of the polynomial

$$g(x) = f(x) - (\lambda x + \nu)^2 = x^3 - (\lambda^2 - a)x^2 + (b - 2\lambda\nu)x - \nu^2.$$

Our first goal is to find $\lambda', \nu' \in K$ so that x_1', x_2', x_3' are roots of the polynomial

$$g'(x) = f'(x) - (\lambda' x + \nu')^2 = x^3 - ((\lambda')^2 - a')x^2 + (b' - 2\lambda'\nu')x - (\nu)'^2,$$

which should exist if we believe 1.20.7(a).[9] From Vieta's formulae applied to $g(x)$, we note that

$$x_1 + x_2 + x_3 = \lambda^2 - a, \quad x_1x_2 + x_1x_3 + x_2x_3 = b - 2\lambda\nu, \text{ and } \quad x_1x_2x_3 = -\nu^2.$$

This allows us to conclude that

$$\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} = \frac{b - 2\lambda\nu}{\nu^2},$$

[9] The notation $f'(x)$ and $g'(x)$ here does *not* stand for the derivatives of $f(x)$ and $g(x)$ respectively; sorry about this overloaded notation!

since $\nu \neq 0$. If indeed x_i' are to be roots of $g'(x)$, then we should also have that

$$x_1' + x_2' + x_3' = (\lambda')^2 - a'.$$

But we know how to express the x_i' in terms of the x_i. Indeed, we get from this the equation

$$(\lambda')^2 - a' = \sum_{i=1}^{3} x_i' = \sum_{i=1}^{3} \left(x_i + a + \frac{b}{x_i} \right) = (\lambda^2 - a) + 3a + b \left(\frac{b - 2\lambda\nu}{\nu^2} \right),$$

which along with $a' = -2a$ suggests that we might take

$$\lambda' := \lambda - \frac{b}{\nu}$$

as above. Having obtained this formula for λ', we can easily obtain ν' by simply asking for

$$\nu' := y_i' - \lambda' x_i'$$

for any i and doing some simplification (1.20.14); miraculously, all choices of i give the same formula

$$\nu' = \nu - a\lambda + b\frac{\lambda^2}{\nu}.$$

This is one way one might derive a formula for λ' and ν'.

We have now reduced the proof of 1.15.9 in the case of a nontrivial two-torsion point to showing that the cokernel $\operatorname{coker} \phi(\mathbb{Q})$ is finite, for if we do this, by symmetry, the cokernel $\operatorname{coker} \phi'(\mathbb{Q})$ would also be finite, and hence so would be $\operatorname{coker}[2](\mathbb{Q}) = E(\mathbb{Q})/2E(\mathbb{Q})$ by 1.20.3(b). Of course, our strategy for this will be to produce a group homomorphism α out of $E(\mathbb{Q})$ with the property that $\ker \alpha = \phi'(E'(\mathbb{Q}))$ and that $\operatorname{Im} \alpha$ is finite. This is what we will start with next time.

1.21 26/07/16 - Finishing the Proof of Mordell's Theorem and the Structure of Finitely Generated Abelian Groups

The first order of business today is to finish the proof of Mordell's Theorem.

1.21.A Wrapping up Descent via Two-Isogeny

The following discussion has been adapted from [2, §3.4].

Recall the set-up we are in: we have $K = \mathbb{Q}$ and $a, b \in \mathbb{Q}$ are such that $b(a^2 - 4b) \neq 0$, and $E/\mathbb{Q}$ is the curve $y^2 = x^3 + ax^2 + bx$, which has the nontrivial rational two-torsion point $T = (0,0) \in E[2](\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$. By definition, $(a', b') := (-2a, a^2 - 4b)$, and $E'/\mathbb{Q}$ is the curve $(y')^2 = (x')^3 + a'(x')^2 + b'x'$, which as its own $T' = (0,0) \in E'[2](\mathbb{Q}) \smallsetminus \{\mathscr{O}'\}$. We have defined maps $\phi : E \to E'$ and $\phi' : E' \to E$ such that $\phi' \circ \phi = [2]_E$ (1.20.13). To prove 1.15.9 (and hence 1.15.8), it only remains to show by 1.20.3 that $E(\mathbb{Q})/\phi'(E'(\mathbb{Q}))$ is finite. This is achieved by

Theorem 1.21.1. In the above setup, let $\alpha : E(\mathbb{Q}) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ be the map given by

$$P \mapsto \begin{cases} [1] & \text{if } P = \mathscr{O}, \\ [b] & \text{if } P = T, \text{ and} \\ [x] & \text{if } P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}, T\}. \end{cases}$$

Then:

(a) The map α is a group homomorphism.
(b) We have that $\ker \alpha = \phi'(E'(\mathbb{Q}))$.
(c) The image $\operatorname{Im} \alpha$ is finite.

Note that the definition of α makes sense thanks to 1.20.1(a). As in 1.19.1, we see in the proof below where the right choice of the value of $\alpha(T)$ comes from.

Proof.

(a) As in the proofs of 1.18.1, 1.19.1, and 1.20.13, we invoke 1.15.5 to reduce to showing that if $P_1, P_2, P_3 \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$ are collinear (with multiplicities as usual), then $\alpha(P_1)\alpha(P_2)\alpha(P_3) = [1]$. First suppose that none of the P_i are T, so that the x-coordinates x_i of the P_i are all nonzero (1.20.1(a)). If $\ell : y = \lambda x + \nu$ is the line containing the P_i for some $\lambda, \nu \in \mathbb{Q}$, then $T \notin \ell(\mathbb{Q})$ implies $\nu \neq 0$. As in previous proofs, x_1, x_2, x_3 are the roots of

$$g(x) = (x^3 + ax^2 + bx) - (\lambda x + \nu)^2,$$

and so Vieta's formulae tell us that $x_1x_2x_3 = \nu^2$, showing that $\alpha(P_1)\alpha(P_2)\alpha(P_3) = [1]$.
Finally, as in the proof of 1.20.13, it remains to deal with the case in which $P_1 = T$ but $P_2, P_3 \neq T$, which means as above that $x_2x_3 \neq 0$. This time, we have using 1.20.10 that

$$\alpha(P_2) = \alpha(-(P_3 + T)) = \alpha(P_3 + T) = \left[\frac{b}{x_3}\right] = \frac{1}{\alpha(T)\alpha(P_3)},$$

where the last equality uses our choice of $\alpha(T)$ (this is where that comes from!). This finishes the proof.

(b) We have to show that if $P \in E(\mathbb{Q})$, then $\alpha(P) = [1]$ iff there is a $P' \in E'(\mathbb{Q})$ such that $\phi'(P') = P$. If $P = \mathscr{O}$, both sides hold and we are done. Next, let's consider the case of $P = T$. We see that $\alpha(T) = [1]$ iff b is a square, in which case $P' \in \{(a \pm 2\sqrt{b}, 0)\} = E'[2](\mathbb{Q}) \smallsetminus \{\mathscr{O}', T'\}$ have the property that $\phi'(P') = T$. Conversely, if there is such a P', then $P' \neq \mathscr{O}', T'$. If we write $P' = (x', y') \in E'(\mathbb{Q}) \smallsetminus \{\mathscr{O}', T'\}$ with $x' \neq 0$, then it follows from the formula (1.20.13) that $(x')^2 + a'x' + b' = 0$. Since x' is rational, the discriminant $a'^2 - 4b' = 16b$ is then a perfect rational square, and hence $\alpha(T) = [b] = [1]$.[10]

[10] The idea behind this proof is that $\alpha(T) = [1]$ iff $b = ((a')^2 - 4b')/16$ is a square. On the one hand, by 1.20.1(b) is

Finally, suppose that $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}, T\}$, which implies $x \neq 0$ (1.20.1(a)). If there is a $P' \in E'(\mathbb{Q})$ with $\phi'(P') = P$, then $P' \neq \mathscr{O}', T'$, and so writing $P' = (x', y')$, we see that

$$x = \frac{1}{4}\left(x' + a' + \frac{b'}{x'}\right) = \left(\frac{y'}{2x'}\right)^2,$$

which shows $\alpha(P) = [x] = [1]$. The converse is more interesting. Suppose that $\alpha(P) = [1]$, which says precisely that x is a square, say $x = u^2$ for $u \in \mathbb{Q}^\times$. The general theory (1.20.C) says there should be two points $P_1, P_2 \in E'(\mathbb{Q})$ with the property that $\phi'(P_i) = P$ for $i = 1, 2$, and a little fiddling around tells us what the coordinates of the P_i should be. Explicitly, let $x_1, x_2, y_1, y_2 \in \mathbb{Q}$ be defined by

$$x_1 := 2x + a + \frac{2y}{u}, \quad y_1 := 2ux_1, \quad x_2 := 2x + a - \frac{2y}{u}, \quad \text{and} \quad y_2 := -2ux_2.$$

We claim that $P_i := (x_i, y_i) \in E'(\mathbb{Q})$ with $\phi'(P_i) = P$ for $i = 1, 2$. Indeed, note that

$$x_1x_2 = (2x + a)^2 - \frac{4y^2}{x} = a^2 + \frac{4}{x}\left(x^3 + ax^2 - y^2\right) = b'.$$

Since $b' \neq 0$, this implies in particular that $x_1x_2 \neq 0$. Therefore, to show that $P_1 \in E'(\mathbb{Q})$, we have to show that

$$\frac{y_1^2}{x_1^2} = x_1 + a' + \frac{b'}{x_1}.$$

The left side of this is clearly $4u^2 = 4x$, and since $b' = x_1x_2$, the right side is $x_1 + a' + x_2 = 4x + 2a + a' = 4x$. Of course, a symmetric argument shows that $P_2 \in E'(\mathbb{Q})$. Finally, we show that $\phi'(P_1) = P$ simply by using the formula in 1.20.13. Indeed, the x-coordinate of $\phi'(P_1)$ is

$$x_1 + a' + \frac{b'}{x_1} = \left(\frac{y_1}{2x_1}\right)^2 = u^2 = x,$$

and the y-cooordinate is

$$\frac{1}{8} \cdot y_1 \cdot \left(1 - \frac{b'}{x_1^2}\right) = \frac{1}{8} \cdot 2ux_1 \cdot \left(1 - \frac{x_2}{x_1}\right) = \frac{u}{4}(x_1 - x_2) = \frac{u}{4} \cdot \frac{4y}{u} = y,$$

finishing the proof (the other case is, of course, symmetric).

(c) Of course, we will find a set S of primes such that $\operatorname{Im}\alpha \subset \mathbb{Q}(S, 2)$. We claim that it suffices to take $S := \{p : p \mid b\}$. Indeed, suppose that $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}, T\}$ and that p is a prime such that $v_p(x) \equiv 1 \pmod 2$; we have to show that $p \mid b$. As in 1.18.1 and 1.19.1, we use 1.14.8 to clear denominators and write P as $(m/e^2, n/e^3)$ for $m, n, e \in \mathbb{Z}$ with $\gcd(m, e) = \gcd(n, e) = 1$; then $P \in E(\mathbb{Q})$ implies after clearing denominators that

$$n^2 = m(m^2 + ame^2 + be^4).$$

Now $v_p(x) \equiv 1 \pmod 2$ implies $v_p(m) \equiv 1 \pmod 2$; since $m \in \mathbb{Z}$, this implies using the above equation that both $p \mid m$ and $p \mid m^2 + ame^2 + be^4$. These together imply that $p \mid be^4$, but again $p \nmid e$ since $\gcd(m, e) = 1$, and hence we conclude that $p \mid b$ as needed.

■

This completes the proof of 1.15.9 and hence of the Mordell Theorem (1.15.8). Let's take a second here to think how magical this is: we put a group structure on the set of points on a cubic curve, and then said something about this group structure, to conclude that (over $\mathbb{Q}$), there are finitely many "seed points" such that every other rational point can be obtained from these by repeated applications of the chord-and-tangent process.

It is clear from the above proof that the size of these "Mordell" or "Mordell-Weil" groups is controlled by the number of prime factors of b. Specifically, in the notation of 1.21.1, we see the the image $\operatorname{Im}\alpha$ of α consists of things that look like $[b_1]$, where $b_1 \in \mathbb{Z}$ is a squarefree divisor of b. Let's give a characterization using the above proof of which such b_1 occur.

equivalent to saying that $E'[2](\mathbb{Q}) \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2$. On the other hand, it is clear from the geometric picture of §1.20.C that the points P' of E' mapping to T are exactly the nontrivial 2-torsion points other than T', which exist over $\mathbb{Q}$ iff $E'[2](\mathbb{Q})$ is fully split over $\mathbb{Q}$ as above.

Corollary 1.21.2. Continuing the notation of 1.21.1, if we write $b = b_1 b_2$ for some squarefree $b_1 \in \mathbb{Z}$, then $[b_1] \in \operatorname{Im}\alpha$ iff there is a rational solution (u, v, w) to $w^2 = b_1 u^4 + a u^2 v^2 + b_2 v^4$ with not all u, v, w zero.

I learned the following very short proof from Tom Fisher.

Proof. If $b_1 = 1$ or b_2 is a square, then both sides of the implication are true (check!). Hence suppose that neither $[b_1]$ nor $[b_2]$ is [1]. Then $[b_1] \in \operatorname{Im}\alpha$ iff there is a $P = (x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}, T\}$ such that $x = b_1 t^2$ for some $t \in \mathbb{Q}^\times$. If this happens, then we have that

$$y^2 = b_1 t^2 \cdot (b_1^2 t^4 + a b_1 t^2 + b),$$

so that $(u, v, w) = (t, 1, y/(b_1 t))$ is a solution to $w^2 = b_1 u^4 + a u^2 v^2 + b_2 v^4$. Conversely, if (u, v, w) is a solution to $w^2 = b_1 u^4 + a u^2 v^2 + b_2 v^4$ with (u, v, w) not all zero, then in fact $uv \neq 0$, and then we can reverse-engineer what the (x, y) should be; explicitly,

$$\left(b_1 \left(\frac{u}{v}\right)^2, b_1 \left(\frac{u}{v}\right)\left(\frac{w}{v^2}\right) \right) \in E(\mathbb{Q})$$

shows us that $[b_1] \in \operatorname{Im}\alpha$. ∎

Remark 1.21.3. The equations (even curves) of the form $w^2 = b_1 u^4 + a u^2 v^2 + b_2 v^4$ are known as *homogeneous spaces*. The existence of points on these homogeneous spaces determines the rank of the corresponding elliptic curves. This can be done in lots of specific cases, as we shall see next time, but sadly there is no general algorithm known to be able to do this in all cases. This is essentially why there is no algorithmic/constructive proof of the Mordell Theorem (1.15.8).

1.21.B The Structure Theorem for Finitely Generated Abelian Groups

We have proved that if $E/\mathbb{Q}$ is an elliptic curve, then the group $E(\mathbb{Q})$ is finitely generated. Let's spend just a little bit of time reviewing what finitely generated abelian groups look like.

Clearly, a group of the form $A = \mathbb{Z}^{\oplus r} \oplus T$ for $r \in \mathbb{Z}_{\geq 0}$ and T a finite abeian group is a finitely generated abelian group. It is a fact–that needs proof–that these are all.

Theorem 1.21.4. Let A be a finitely generated abelian group. There are unique integers $r, s \in \mathbb{Z}_{\geq 0}$ and unique integers $2 \leq a_1, \ldots, a_s \in \mathbb{Z}_{\geq 2}$ such that $a_1 \mid a_2 \mid \cdots \mid a_s$ and such that

$$A \cong_{\mathsf{Ab}} \mathbb{Z}^{\oplus r} \oplus \mathbb{Z}/a_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/a_s\mathbb{Z}.$$

The integer r is called the *rank* of A, and the integers a_i are called the *elementary divisors* of A.

As far as I know, the integer s doesn't have a standard name other than the "number of elementary divisors".

Proof. Let's first note the uniqueness of the invariants. If A is expressible in this form, then r can be recovered as the unique integer such that $r = \dim_{\mathbb{F}_p} A/pA$ for all but finitely many primes p. The integer s can be recovered as $\max\{\dim_{\mathbb{F}_q} A/qA\} - r$. The integer a_s can be recovered as the unique positive annihilator of $\operatorname{Tors}(A)$, i.e., the smallest positive integer a such that $a \cdot \operatorname{Tors}(A) = 0$. The other a_i's admit a similar description; I will leave the details of these checks and the characterization of the other a_t for $t \leq s$ as an exercise on the next Exercise Sheet.

Let's now prove the existence of such r, s, and a_i for a given finitely generated abelian group A. Since A is finitely generated, we can pick finitely many, say $m \in \mathbb{Z}_{\geq 0}$, generators for A; this amounts to giving a surjection $\mathbb{Z}^{\oplus m} \twoheadrightarrow A$. If we let $K \subset \mathbb{Z}^{\oplus m}$ be the kernel of this surjection, then we can express this "presentation" of A via the short exact sequence

$$0 \to K \to \mathbb{Z}^{\oplus m} \to A \to 0.$$

(See 2.5.11 for a reminder on basic properties of exact sequences.) The next step is to prove

Lemma 1.21.5. Let $m \in \mathbb{Z}_{\geq 0}$, and let $K \subset \mathbb{Z}^{\oplus m}$ be a subgroup. Then there is an integer $n \in \mathbb{Z}_{\geq 0}$ with $n \leq m$ such that $K \cong_{\mathsf{Ab}} \mathbb{Z}^{\oplus n}$. In other words, a subgroup of a free abelian group of rank m is free abelian of rank at most n.

Proof. We induct on m, with the case of $m = 0$ being clear. The case $m = 1$ follows from the fact that any subgroup of $\mathbb{Z}$ has the form $c\mathbb{Z}$ for a unique $c \in \mathbb{Z}_{\geq 0}$, with the cases $n = 0$ and $n = 1$ corresponding to $c = 0$ and $c \geq 1$ respectively. Now suppose that $m \geq 2$, and that we have shown the result for $m - 1$, and that $K \subset \mathbb{Z}^{\oplus m}$ is a subgroup. Consider the projection of K to the last factor $\pi_m : K \hookrightarrow \mathbb{Z}^{\oplus m} \twoheadrightarrow \mathbb{Z}$. The image $\pi_m(K) \subset \mathbb{Z}$ is a subgroup, and hence by the above discussion is of the form $c\mathbb{Z}$ for some unique $c \in \mathbb{Z}_{\geq 0}$.

If $c = 0$, then K is actually a subgroup of $\mathbb{Z}^{\oplus(m-1)} \oplus 0 \subset \mathbb{Z}^{\oplus m}$, and hence by induction isomorphic to $\mathbb{Z}^{\oplus n}$ for some $n \leq m - 1$.

If $c \geq 1$, then since $c \in \pi_m(K)$, we can pick an $\alpha \in K$ such that $\pi_m(\alpha) = c$. We claim that α is part of a free basis of K. Explicitly, consider the intersection $K' := K \cap (\mathbb{Z}^{m-1} \oplus 0)$. By induction again, K' is isomorphic to $\mathbb{Z}^{\oplus n}$ for some $n \leq m-1$. Therefore, it suffices to show that the group homomorphism

$$\mathbb{Z} \oplus K' \to K, \quad (r, \beta) \mapsto r\alpha + \beta$$

is an isomorphism. Let's prove injectivity and surjectivity. If $(r, \beta) \in \mathbb{Z} \oplus K'$ are such that $r\alpha + \beta = 0$, then

$$0 = \pi_m(0) = \pi_m(r\alpha + \beta) = r\pi_m(\alpha) + \pi_m(\beta).$$

Since $\beta \in K'$, we have $\pi_m(\beta) = 0$. Since $\pi_m(\alpha) = c \neq 0$, this forces $r = 0$, and then $r\alpha + \beta = 0$ forces $\beta = 0$ as well; this proves injectivity. Conversely, if $\gamma \in K$, then $\pi_m(\gamma) \in \pi_m(K) = c\mathbb{Z}$ tells us that there is an integer $r \in \mathbb{Z}$ such that $\pi_m(\gamma) = rc$. If we let $\beta := \gamma - r\alpha$, then

$$\pi_m(\beta) = \pi_m(\gamma) - r\pi_m(\alpha) = 0,$$

so that $\beta \in K'$. Therefore, $(r, \beta) \in \mathbb{Z} \oplus K'$ has the property that $r\alpha + \beta = \gamma$, proving surjectivity as needed. ∎

At this point, we have a "presentation" of A of the form

$$0 \to \mathbb{Z}^{\oplus n} \xrightarrow{\varphi} \mathbb{Z}^{\oplus m} \to A \to 0,$$

which expresses A as the cokernel of the map φ. The map φ itself is given by an $m \times n$ integer matrix M. If we can perform row and column operations (over the integers, which correspond to choosing different bases of $\mathbb{Z}^{\oplus n}$ and $\mathbb{Z}^{\oplus m}$ respectively) to bring M into a particularly simple form, then we can hope to read off the isomorphism class of A from this simple form. One such simple form is known as the Smith Normal Form.

Lemma 1.21.6 (Smith Normal Form)**.** Let $m, n \in \mathbb{Z}_{\geq 0}$, and let M be an $m \times n$ integer matrix. We can perform row and column operations on M (over the integers!) to bring M into the *Smith Normal Form*, as follows: there is a unique integer $t \in \mathbb{Z}_{\geq 0}$ with $t \leq \min\{m, n\}$ and unique integers $b_1, \ldots, b_t \in \mathbb{Z}_{\geq 1}$ with $b_1 \mid \cdots \mid b_t$ such that upon performing these row and column operations, M transforms into a matrix of the form

$$M \sim \begin{bmatrix} D & 0_{t \times (n-t)} \\ 0_{(m-t) \times t} & 0_{(m-t) \times (n-t)} \end{bmatrix},$$

where $0_{a \times b}$ for $a, b \in \mathbb{Z}_{\geq 0}$ denotes the zero matrix of size $a \times b$, and D is the $t \times t$ diagonal matrix with entries b_i, i.e.,

$$D = \begin{bmatrix} b_1 & & & \\ & b_2 & & \\ & & \ddots & \\ & & & b_t \end{bmatrix}.$$

Proof. Again, let's note uniqueness first: t can be recovered as the rank (say over $\mathbb{Q}$) of the matrix M, and b_1 can be recovered as the unique positive generator (if $t \geq 1$) of the subgroup of $\mathbb{Z}$ generated by the entries of M. Similar descriptions exist also for the b_i for $i \geq 2$; again, I will leave this to you as an exercise on the next sheet.

Existence is best proved by exhibiting the following simple algorithm, which takes input (m, n, M) and returns output $(t; b_1, \ldots, b_t)$ as above.

Algorithm 1.21.7 (Smith Normal Form).

Step 1. Input a triple (m, n, M), where $m, n \in \mathbb{Z}_{\geq 0}$ and M is an $m \times n$ integer matrix. Are all entries of M zero? If yes, return $t = 0$ and the empty list. Else, proceed to Step 2.

Step 2. Pick a nonzero entry c_0 of M with least positive absolute value, and perform row and column switches to bring it to the top left corner. Does the entry c_0 of the top left corner divide every entry of the matrix? If not, proceed to Step 3. If yes, proceed to Step 4.

Step 3. There is some integer c_1 in some row or column that is not divisible by c_0. Is this entry in the first row or column? If yes, proceed to Step 3a; if no, proceed to Step 3b.

Step 3a. Suppose c_1 is in the first *row* (the column case is dual). Subtract a multiple of the first *column* from the column containing c_1 to obtain an entry of smaller positive absolute value than c_0. Return to Step 2, replacing c_0 by this entry.

Step 3b. Every element of the first row and column is divisible by c_0. Subtract a multiple of the first column from each column to make all entries other than c_0 in the topmost row zero. Similarly, subtract a multiple of the first row from each row to make all entries other than c_0 in the leftmost column zero. There is still some entry c_1' not divisible by c_0, but now c_1 is not in the first row or column. By adding the row containing c_1' to the first row, put c_1' in the first row, and now do Step 3a.

Step 4. Every element of M is divisible by c_0. As in Step 3b, clear the entries of the first row and column other than c_0 (i.e., make them zero). If $c_0 < 0$, replace it by $-c_0$ by multiplying the first row by -1 to ensure $c_0 > 0$. Let M' be the matrix obtained from M by deleting the first row and column. Return to Step 1 and input $(m-1, n-1, M')$. If the output is $(t'; b_1', \ldots, b_{t'}')$, then return $(t; b_1, \ldots, b_t) := (t'+1; c_0, b_1', \ldots, b_{t'}')$.

Let's illustrate this with a quick example.

Example 1.21.8. Consider $m = 3, n = 2$, and sequence of moves on the matrix M as follows:

$$M := \begin{bmatrix} 3 & 5 \\ 0 & 2 \\ -3 & 7 \end{bmatrix} \sim \begin{bmatrix} 0 & 2 \\ 3 & 5 \\ -3 & 7 \end{bmatrix} \sim \begin{bmatrix} 2 & 0 \\ 5 & 3 \\ 7 & -3 \end{bmatrix}.$$

At this point, $c_0 = 2$ and we're in Step 3a, where we take $c_1 = 5$. Using this to proceed, we get

$$M \sim \begin{bmatrix} 2 & 0 \\ 1 & 3 \\ 7 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ 7 & -3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 2 & -6 \\ 7 & -24 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & -6 \\ 7 & -24 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & -6 \\ 0 & -24 \end{bmatrix}.$$

At this point, we are in Step 4, and apply the procedure to the submatrix

$$M' := \begin{bmatrix} -6 \\ -24 \end{bmatrix} \sim \begin{bmatrix} 6 \\ -24 \end{bmatrix} \sim \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

The procedure outputs $t = 2$ and $(b_1, b_2) = (1, 6)$, i.e., the Smith Normal Form of M is

$$M \sim \begin{bmatrix} 1 & 0 \\ 0 & 6 \\ 0 & 0 \end{bmatrix}.$$

■

At this point, we are basically done. We started with a finitely generated abelian group A, and picked a presentation for it of the form

$$0 \to \mathbb{Z}^{\oplus n} \xrightarrow{\varphi} \mathbb{Z}^{\oplus m} \to A \to 0$$

for some $m, n \in \mathbb{Z}_{\geq 0}$ and map φ represented by an $m \times n$ integer matrix M. We use 1.21.6 to put the M into Smith Normal Form up to row and column operations. This expresses A as the cokernel of a matrix map in Smith Normal Form with invariants say $(t; b_1, \ldots, b_t)$. Here, it is not hard to see that we must have $t = n$ (since φ is injective, it has maximal rank), and we therefore see that A is isomorphic to

$$\mathbb{Z}/b_1\mathbb{Z} \oplus \cdots \mathbb{Z}/b_n\mathbb{Z} \oplus \mathbb{Z}^{m-n}.$$

At this point $r = m - n$. It only remains to discard all the b_is which are ones; set $s := \#\{i : b_i \neq 1\}$, and let $a_1, \ldots, a_s$ be the last s of the $b_1, \ldots, b_t$. ■

We now have a good understanding of finitely generated abelian groups. For instance, a finitely generated abelian group is finite iff it is torsion iff it has rank zero. Next time, we will use these ideas to compute the ranks of lots of elliptic curves, and derive several delicious arithmetic consequences.

1.22 26/07/17 - The Rank Formula, Examples and Applications, Mazur's Theorem and Rank Records

We have at this point shown that if $E/\mathbb{Q}$ is an elliptic curve, then $E(\mathbb{Q})$ is a finitely generated abelian group (1.15.8) and hence (1.21.4) that $E(\mathbb{Q}) \cong_{\mathsf{Ab}} \mathbb{Z}^r \oplus \operatorname{Tors} E(\mathbb{Q})$ for a unique integer $r \in \mathbb{Z}_{\geq 0}$ (depending on E) and finite group $\operatorname{Tors} E(\mathbb{Q})$. This integer r is called the *(algebraic) rank* of the elliptic curve E, and is written as $r = \operatorname{rank} E$. This rank encodes important arithmetic information about rational points on E; for instance, E has finitely many rational points iff $\operatorname{rank} E = 0$. The Nagell-Lutz Theorem (1.13.7) gives an algorithmic procedure to compute $\operatorname{Tors} E(\mathbb{Q})$, but how do we actually compute the rank of a given elliptic curve E over $\mathbb{Q}$?

Remark 1.22.1. If we look at the proofs of 1.18.1, 1.19.1, and 1.21.1 more carefully, we will note that (up to a suitable analog of 1.14.8), we have basically proven the weak Mordell-Weil theorem (1.15.9) for not just $\mathbb{Q}$ but any number field $K/\mathbb{Q}$. The proof of the full Mordell-Weil Theorem (1.8.1, the analog of 1.15.8 for number fields K) is then just one step away: namely, developing the theory of heights as in §1.16 over a general number field, for then we can invoke 1.15.11. This is done systematically, for instance, in [8, Chapter VIII].

1.22.A The Rank Formula

As in §1.20, we will henceforth assume that the elliptic curve E has a nontrivial rational 2-torsion point (but see 1.22.5(b)). This means that E can be brought into the form $E = E_{a,b} : y^2 = x(x^2 + ax + b)$ for some integers $a, b \in \mathbb{Z}$ with $b(a^2 - 4b) \neq 0$. As in 1.20, we can then let $(a', b') := (-2a, a^2 - 4b)$, and let $E' : (y')^2 = (x')^3 + a'(x')^2 + b'x'$. Then 1.20.13 gives us morphisms $\phi : E \to E'$ and $\phi' : E' \to E$ with the property that $\phi' \circ \phi = [2]_E$ and $\phi \circ \phi' = [2]_{E'}$. Finally, 1.21.1 gives us a map $\alpha : E(\mathbb{Q}) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ such that $\ker \alpha = \phi'(E'(\mathbb{Q}))$ and $\operatorname{im} \alpha \subset \mathbb{Q}(S, 2)$, where $S = \{p : p \mid b\}$; then $\operatorname{coker} \phi'(\mathbb{Q}) \xrightarrow{\sim} \operatorname{im} \alpha$. Similarly, we have a map $\alpha' : E'(\mathbb{Q}) \to \mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ with the property that $\ker \alpha' = \phi(E(\mathbb{Q}))$ and $\operatorname{im} \alpha' \subset \mathbb{Q}(S', 2)$, with $S' := \{p : p \mid b'\}$; again $\operatorname{coker} \phi(\mathbb{Q}) \xrightarrow{\sim} \operatorname{im} \alpha'$.

Now we can consider the sequence of group homomorphisms

$$E(\mathbb{Q}) \xrightarrow{\phi(\mathbb{Q})} E'(\mathbb{Q}) \xrightarrow{\phi'(\mathbb{Q})} E(\mathbb{Q})$$

which compose to the duplication map $[2] : E(\mathbb{Q}) \to E(\mathbb{Q})$. We can apply 1.20.3(b) to this to obtain the exact sequence

$$0 \to \ker \phi(\mathbb{Q}) \to E[2](\mathbb{Q}) \to \ker \phi'(\mathbb{Q}) \to \operatorname{coker} \phi(\mathbb{Q}) \to E(\mathbb{Q})/2E(\mathbb{Q}) \to \operatorname{coker} \phi'(\mathbb{Q}) \to 0.$$

We know that $\ker \phi(\mathbb{Q}) = \{\mathscr{O}, T\}$ and $\ker \phi'(\mathbb{Q}) = \{\mathscr{O}', T'\}$. Further, we know that $\operatorname{coker} \phi(\mathbb{Q}) \xrightarrow{\sim} \operatorname{im} \alpha$ and $\operatorname{coker} \phi'(\mathbb{Q}) \xrightarrow{\sim} \operatorname{im} \alpha'$. Using these facts along with 2.5.11(d) tells us immediately that

$$2 \cdot 2 \cdot \#E(\mathbb{Q})/2E(\mathbb{Q}) = \#E[2](\mathbb{Q}) \cdot \#\operatorname{im} \alpha \cdot \#\operatorname{im} \alpha',$$

which can be rearranged to be

$$\frac{\#E(\mathbb{Q})/2E(\mathbb{Q})}{\#E[2](\mathbb{Q})} = \frac{\#\operatorname{im} \alpha \cdot \#\operatorname{im} \alpha'}{4}. \tag{1.22.2}$$

Now 1.21.4 tells us that $E(\mathbb{Q}) \cong_{\mathsf{Ab}} \mathbb{Z}^r \oplus T$ for some unique integer $r \in \mathbb{Z}_{\geq 0}$ (namely $r = \operatorname{rank} E$) and some unique finite group T (up to isomorphism). In this case, we clearly have that

$$E(\mathbb{Q})/2E(\mathbb{Q}) \cong_{\mathsf{Ab}} (\mathbb{Z}/2\mathbb{Z})^r \oplus (T/2T) \text{ and } E[2](\mathbb{Q}) \cong T[2].$$

In particular,

$$\#E(\mathbb{Q})/2E(\mathbb{Q}) = 2^r \cdot \#T/2T \text{ and } \#E[2](\mathbb{Q}) = \#T[2].$$

However, it is not hard to see at all that $\#T/2T = \#T[2]$ (you could use 1.21.4, or apply 2.5.11(c) cleverly!), and so this is enough to conclude that

$$2^r = \frac{\#\operatorname{im} \alpha \cdot \#\operatorname{im} \alpha'}{4}. \tag{1.22.3}$$

The formula on the right side is completely symmetric in E and E'. We have deduced

Theorem 1.22.4 (The Rank Formula). In the above setup, we have that

$$\operatorname{rank} E = \operatorname{rank} E' = \log_2 \#\operatorname{im}\alpha + \log_2 \#\operatorname{im}\alpha' - 2.$$

Given our characterization 1.21.2 of the images of α and α' in terms of homogeneous spaces, we have a strategy for determining the ranks of elliptic curves (but see 1.22.16), which we will use in the next section to compute lots of examples.

Remark 1.22.5.

(a) The result of 1.22.4 implies that the two-isogenous curves E and E' have the same rank over $\mathbb{Q}$. It can be shown in general that if $E/\mathbb{Q}$ and $E'/\mathbb{Q}$ are two elliptic curves are isogenous over $\mathbb{Q}$ (1.20.4 and 1.20.7(b)), then $\operatorname{rank} E = \operatorname{rank} E'$. Indeed, the same is true of elliptic curves isogenous over any number field (1.8.1); this follows immediately from the fact that the kernel of a nonconstant isogney is finite and that being isogenous is an equivalence relation (1.20.7(b)).

(b) When the elliptic curve $E/\mathbb{Q}$ does not have a nontrivial rational two-torsion point, then the above strategy cannot be used to compute $\operatorname{rank} E$. This can be done either by the use of a little algebraic number theory (2.5.17), or by using the theory of binary quartics; this is explained, for instance, in Cremona's book ([17, Chapter III]).

1.22.B Examples and Applications

Let's now work through some examples.

Example 1.22.6. Let $E : y^2 = x^3 - x$, so that $(a, b, a', b') = (0, -1, 0, 4)$. In this case, $\operatorname{im}\alpha \subset \mathbb{Q}(\emptyset, 2) = \{[\pm 1]\}$, but the image certainly contains $[1] = \alpha(\mathscr{O})$ and $[-1] = \alpha(T)$. This implies that $\operatorname{im}\alpha = \{\pm 1\}$, and hence $\#\operatorname{im}\alpha = 2$.

The corresponding two-isogenous curve is $E' : y^2 = x^3 + 4x$. Here $\operatorname{im}\alpha' \subset \mathbb{Q}(\{2\}, 2) = \{[\pm 1], [\pm 2]\}$. The image again certainly contains $[1]$. By 1.21.2, we have that $[-1] \in \operatorname{im}\alpha'$ iff there is a nontrivial solution to $w^2 = -u^4 - 4v^4$, which visibly does not exist (there is an obstruction at infinity!). The same reasoning implies that $[-2] \notin \operatorname{im}\alpha'$. Finally, we have that $[2] \in \operatorname{im}\alpha$ iff there is a nontrivial solution to the equation $w^2 = 2u^4 + 2v^4$, which does in fact exist–we could take $(u, v, w) = (1, 1, 2)$. This proves that $\operatorname{im}\alpha' = \{[1], [2]\}$, and hence $\#\operatorname{im}\alpha' = 2$.

At this point, we are done and can use 1.22.4 to conclude that $\operatorname{rank} E = \operatorname{rank} E' = 0$. Since $\operatorname{rank} E = 0$, we conclude that $E(\mathbb{Q}) = \operatorname{Tors} E(\mathbb{Q})$, which we find using Nagell-Lutz (1.13.7) to be $\{\mathscr{O}, (0,0), (\pm 1, 0)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/2$. Therefore, we have determined all rational points of E: we have that

$$E(\mathbb{Q}) = \{\mathscr{O}, (0,0), (\pm 1, 0)\}.$$

Similarly, $\operatorname{rank} E' = 0$, so $E'(\mathbb{Q}) = \operatorname{Tors} E'(\mathbb{Q})$, which again by Nagell-Lutz (1.13.7) is easily seen to be $\{\mathscr{O}, (0,0), (2, \pm 4)\} \cong_{\mathsf{Ab}} \mathbb{Z}/4$. Therefore, we have determined all the rational points of E' too: we have that

$$E'(\mathbb{Q}) = \{\mathscr{O}, (0,0), (2, \pm 4)\}.$$

Let's now derive some delicious consequences.

Corollary 1.22.7. There does not exist a rational right-angled triangle with area a perfect square.

Proof. Suppose there exists such a right angled triangle. We can rescale it to have coprime integer sides; then 1.2.3 tells us that it must have sides of the form $(u^2 - v^2, 2uv, u^2 + v^2)$ for some integers $u > v > 0$. The area of this triangle is exactly $uv(u^2 - v^2)$; therefore, if this area is a square, then there is an integer $w > 0$ such that $w^2 = uv(u^2 - v^2)$. If we divide throughout by v^4, then we see immediately from this equation that $(u/v, w/v^2) \in E(\mathbb{Q})$, where $E : y^2 = x^3 - x$ as in 1.22.6. However, we saw in that example that E does not have *any* rational points with nonzero finite y-coordinate, and this is the required contradiction. ∎

You will investigate some more questions relating to the areas of rational right-angled triangles on the last Exercise Sheet. We now use the theory of elliptic curves to give a second proof of Fermat's Last Theorem for $n = 4$.

Corollary 1.22.8. Let $X, Y, Z \in \mathbb{Z}_{\geq 0}$ be such that $X^4 + Y^4 = Z^4$. Then $XYZ = 0$.

Proof. If $Z = X$, then $Y = 0$. Else, 2.1.12(b) shows (spoiler alert!) that the point

$$(x, y) := \left(2\left(\frac{Z+X}{Z-X}\right), \frac{4Y^2}{(Z-X)^2}\right)$$

is a rational point other than $\mathscr{O}$ on the curve $E' : y^2 = x^3 + 4x$ of 1.22.6. By our complete understanding of $E'(\mathbb{Q})$ in 1.22.6, either $Y = 0$, or

$$2\left(\frac{Z+X}{Z-X}\right) = 2,$$

which implies that $X = 0$. ∎

You will use a different curve to give a second proof of 1.3.1 using elliptic curve theory on the last Exercise Sheet.

Remark 1.22.9. We have shown in 1.3.1 that there does not exist an integer square which is the sum of two integral fourth powers. As a natural next question, it makes sense to relax the condition of being square to something weaker. One such weakening is the notion of being *powerful*: an integer $n \in \mathbb{Z}$ is said to be powerful iff for each prime p, if $p \mid n$, then $v_p(n) \geq 2$. Can a powerful integer be the sum of two integral fourth powers? Yes: for instance, $32 = 16 + 16$. This is silly. A better question to ask is: can a powerful integer be the sum of two *coprime* integral fourth powers? This is a much more subtle question. The answer to this particular question turns out to be "yes", but the smallest integer so expressible is

$$\begin{aligned} &308825748949336027872519696547735921 7 \\ &= 427511122^4 + 1322049209^4 \\ &= 17^3 \cdot 73993169^2 \cdot 338837713^2. \end{aligned}$$

This is a *huge* number ($\approx 3.08 \times 10^{36}$), and it was not found–nor this result proven–using a brute-force search. This number arises rather from a suitable rational point on a quadratic twist (namely $\mathcal{E}_{17} : y^2 = x^3 - 68^2 x$) of our E from 1.22.6. For a detailed study of this particular problem, we refer the reader to [16].

Let's now look at another example.

Example 1.22.10. Let $E : y^2 = x^3 + 5x^2 + 4x$, so that $(a, b, a', b') = (5, 4, -10, 9)$. In this case, $\operatorname{im}\alpha \subset \mathbb{Q}(\{2\}, 2) = \{[\pm 1], [\pm 2]\}$. As above, $\operatorname{im}\alpha \ni [1]$. By 1.21.2, for each squarefree divisor b_1 of $b = 4$, to determine whether each of $[b_1] \in \operatorname{im}\alpha$, we need to study whether a homogeneous space has a nontrivial rational point; let's do this for $b_1 = -1, 2, -2$ below.

(a) If $b_1 = -1$, we are looking at the equation $w^2 = -u^4 + 5u^2v^2 - 4v^4$, which does have a nontrivial rational solution $(1, 1, 0)$. This proves that $[-1] \in \operatorname{im}\alpha$.
(b) If $b_1 = 2$, we are looking at $w^2 = 2u^4 + 5u^2v^2 + 2v^4$, which also has a nontrivial rational solution $(1, 1, 3)$. This proves that $[2] \in \operatorname{im}\alpha$ as well.
(c) To determine if $b_1 = -2$ works, we could look at the relevant homogeneous space, or we could argue that $\operatorname{im}\alpha$ is a subgroup of $\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$, so if it contains both $[-1]$ (thanks to (a)) and $[2]$ (thanks to (b)), then it also contains $[-1] \cdot [2] = [-2]$.

In particular, we conclude that $\operatorname{im}\alpha = \{[\pm 1], [\pm 2]\}$, so that $\#\operatorname{im}\alpha = 4$.

The relevant two-isogenous curve is $E' : y^2 = x^3 - 10x^2 + 9x$. Now, $\operatorname{im}\alpha' \subset \mathbb{Q}(\{3\}, 2) = \{[\pm 1], [\pm 3]\}$. As above, let's use 1.21.2 and work through the cases:

(a) We have that $[-1] \in \operatorname{im}\alpha$ iff $w^2 = -u^4 - 10u^2v^2 - 9v^4$ has a nontrivial rational solution, but it doesn't–this equation has no nontrivial solution over $\mathbb{R}$ (i.e., is obstructed at infinity).
(b) We can exclude $[-3]$ for the same reason as in (a).

(c) Finally, $[3] \in \operatorname{im}\alpha$ iff there is a nontrivial rational solution to $w^2 = 3u^4 - 10u^2v^2 + 3v^4$. This equation is more subtle. It clearly has solutions over $\mathbb{R}$, so our technique for (a) and (b) above is not going to work. Let's try to argue differently. Suppose that (u, v, w) is a solution, and we clear out denominators so that $u, v, w \in \mathbb{Z}$. Then reducing the equation mod 3 tells us that

$$w^2 \equiv -u^2v^2 \pmod 3,$$

which implies that $3 \mid w$ and $3 \mid uv$. Write $w = 3w_1$ for some $w_1 \in \mathbb{Z}$. Since $3 \mid uv$, we have that either $3 \mid u$ or $3 \mid v$; since u and v play symmetric roles, we might as well assume that $3 \mid u$ and write $u = 3u_1$ for some $u_1 \in \mathbb{Z}$. The above equation can then be rewritten as

$$3w_1^2 = 81u_1^4 - 30u_1^2v^2 + v^4.$$

This equation implies now that $3 \mid v$, and a quick inspection then shows that $3 \mid w_1$ as well. At this point, we have shown that if (u, v, w) is an integer solution to $w^2 = 3u^4 - 10u^2v^2 + 3v^4$, then so is $(u/3, v/3, w/9)$. Infinite descent shows that this equation is therefore *not* solvable over $\mathbb{Q}$ (it is even obstructed at 3, or equivalently has no solutions over $\mathbb{Q}_3$); in particular, $[3] \notin \operatorname{im}\alpha'$.

We have now shown that $\operatorname{im}\alpha = \{[1]\}$, and so $\#\operatorname{im}\alpha' = 1$.

At this point, we are done and can use 1.22.4 to conclude that $\operatorname{rank} E = \operatorname{rank} E' = 0$. Since $\operatorname{rank} E = 0$, we conclude that $E(\mathbb{Q}) = \operatorname{Tors} E(\mathbb{Q})$, which we can find using Nagell-Lutz (1.13.7) to be

$$E(\mathbb{Q}) = \operatorname{Tors} E(\mathbb{Q}) = \{\mathscr{O}, (0,0), (-1,0), (-4,0), (2,\pm 6), (-2,\pm 2)\} \cong_{\mathsf{Ab}} \mathbb{Z}/2 \oplus \mathbb{Z}/4.$$

(The points can be found with Nagell-Lutz; the group structure can be found by a little experimentation–left as an exercise!). Similarly, all points on E' can be found by using the Nagell-Lutz Theorem.

Exercise 1.22.11. Check that the the description of the group structure of $E(\mathbb{Q}) = \operatorname{Tors} E(\mathbb{Q})$ in 1.22.10 is correct. Next, use the Nagell-Lutz Theorem (1.13.7) to find all rational points on E', and determine the group structure of $E'(\mathbb{Q})$.

Remark 1.22.12. We could have deduced the result of 1.22.10 with less computation. Indeed, it follows already from $\#\operatorname{im}\alpha' = 1$, the fact that $b = 4$, the fact that the rank is a nonnegative integer, and 1.22.4 that $\operatorname{rank} E = \operatorname{rank} E' = 0$. This illustrates the general principle that the rank of an elliptic curve E of the form $y^2 = x^3 + ax^2 + b$ as above is bounded above by some quantity depending only on the number of primes dividing b and $b' = a^2 - 4b$; this is an exercise on the next sheet.

As another delicious corollary, we can now derive the following result stated by Fermat in 1640 and first proven by Euler in 1780.

Corollary 1.22.13 (Squares in Arithmetic Progression)**.** There is no nonconstant four-term sequence of integer squares in arithmetic progression.

Proof. This is a fun consequence of the fact that for $E : y^2 = x^3 + 5x^2 + 4x$ as in 1.22.10 as above, we have $\operatorname{rank} E = 0$ (see 2.3.5). ∎

You will be asked to investigate a similar result about cubes and fourth powers in arithmetic progression on the last exercise sheet.

Let's work through one final (extended) example together.

Example 1.22.14. Let p be a prime, and let $E_p : y^2 = x^3 + px$ so that $(a, b, a', b) = (0, p, 0, -4p)$. We already have enough information to conclude that $\operatorname{rank} E_p \leq 2$ for odd p, and that $\operatorname{rank} E_2 \leq 1$ (check!).

We know first that $\operatorname{im}\alpha \subset \mathbb{Q}(\{p\}, 2) = \{[\pm 1], [\pm p]\}$. As above, we know already that $[1], [p] \in \operatorname{im}\alpha$. At this point, it only remains to determine whether $b_1 = -1$ works (check!), for which by 1.21.2 we are trying to solve $w^2 = -u^4 - pv^4$. This equation is evidently obstructed at infinity, and so we conclude that $\operatorname{im}\alpha = \{[1], [p]\}$; in particular, $\#\operatorname{im}\alpha = 2$.

The corresponding two-isogenous curve is $E'_p : y^2 = x^3 - 4px$, and here $\operatorname{im}\alpha' \subset \mathbb{Q}(\{2, p\}, 2)$. Here the analysis of the case $p = 2$ diverges from that of the case $p > 2$; I will leave the easier case of $p = 2$ as an exercise on the last sheet, and focus on the case of $p > 2$ starting now. If $p > 2$, then

$\mathbb{Q}(\{2,p\},2) = \{\pm 1, \pm 2, \pm p, \pm 2p\}$ has size eight, so it seems a priori that to use 1.21.2 we need to study eight cases, but the situation is actualy simpler than that.

We know that $[b'] = [-p] \in \operatorname{im}\alpha'$ automatically. The only subgroups of $\mathbb{Q}(\{2,p\},2) = \{\pm 1, \pm 2, \pm p, \pm 2p\}$ that contain $\{[1],[-p]\}$ are

$$\{[1],[-p]\}, \quad \{[\pm 1],[\pm p]\}, \quad \{[1],[2],[-p],[-2p]\}, \quad \{[1],[-2],[-p],[-2p]\},$$

and everything. Therefore, to determine the image of α', it suffices to study three cases: namely, the cases $b_1 = \pm 2$ and $b_1 = p$. By 1.21.2, we need to study whether the following homogeneous spaces have any points:

(a) the case $b_1 = 2$ corresponds to $w^2 = 2u^4 - 2pv^4$,
(b) the case $b_1 = -2$ corresponds to $w^2 = -2u^4 + 2pv^4$, and
(c) the case $b_1 = p$ corresponds to $w^2 = pu^4 - 4v^4$.

These equations are hard to solve in general. Let's look at specific cases; I'll do a few and leave a few more for you to study on the last sheet.

(i) Let's consider $p = 5$. The space corresponding to $b_1 = 2$ has the form $w^2 = 2u^4 - 10v^4$, but it is easy to see that this equation is locally obstructed at 5. Indeed, suppose to the contrary that there is a nontrivial solution (u,v,w), and clear denominators to assume that $u,v,w \in \mathbb{Z}$ with $\gcd(u,v) = 1$. At this point, we note that $5 \nmid u$ because if $5 \mid u$ then $5 \mid w$ and $5 \mid v$ as well, which can't happen. Since $5 \nmid u$, we conclude from $w^2 = 2u^4 - 10v^4$ that 2 is a square mod 5, which is not true. A very similar argument shows that $b_1 = -2$ won't work either. It remains to study the case $b_1 = 5$, which corresponds to the equation $w^2 = 5u^4 - 4v^4$. This equation evidenlty has the nontrivial solution $(1,1,1)$, and so we conclude from this analysis that $\operatorname{im}\alpha' = \{[\pm 1],[\pm 5]\}$. This along with 1.22.4 tells us that $\operatorname{rank} E_5 = \operatorname{rank} E_5' = 1$; in particular, we have shown that $E_5 : y^2 = x^3 + 5x$ has infinitely many solutions. We can easily use these ideas to find generators of $E_5(\mathbb{Q})/2E_5(\mathbb{Q})$, but to go from that to generators of the $E_5(\mathbb{Q})$ needs a little argument involving heights; this I will again leave for the Exercise Sheet.

(ii) More generally, if $p \equiv 5 \pmod 8$, then the same argument as in (i) using that $\left(\frac{\pm 2}{p}\right) = -1$ shows that the values $b_1 = \pm 2$ don't work. This leaves us with the case of $b_1 = p$ corresponding to $w^2 = pu^4 - 4v^4$. Note that by 1.22.4, the equation $w^2 = pu^4 - 4v^4$ has a nontrivial solution iff $\operatorname{rank} E_p = 1$. We can certainly try to solve the above equation for u,v,w–or prove there are no nontrivial solutions–for specific values of p, but doing this for all such p simultaneously seems very hard. We could start by finding local obstructions, but it is not too hard to use Hensel's Lemma to show that this equation is everywhere locally unobstructed, i.e., has solutions in $\mathbb{Q}_q$ for each prime $q \leq \infty$. This is sadly not sufficient to conclude that we always have global solutions. Indeed, this is an open problem: it is conjectured, but not yet known, that for all primes p with $p \equiv 5 \pmod 8$, the curve E_p has rank 1.

(iii) Finally, let's consider the case $p = 17$. The case $b_1 = 2$ corresponds to the equation

$$w^2 = 2u^4 - 34v^4. \tag{1.22.15}$$

This equation is very interesting; that something interesting is going on in this example was discovered essentially independently by Carl-Erik Lind, Hans Reichardt, and Gunnar Billing (see this fascinating article [18] by three authors, with two being our very own Paul Pollack and Dan Shapiro discussing the history and mathematics of this specific equation). Suppose this equation has a nontrivial solution (u,v,w), and assume again that $u,v,w \in \mathbb{Z}$ with $\gcd(u,v) = 1$. Note that $w \neq 0$.

Let q be a prime dividing w. If $q = 17$, then $17 \mid u$ too, and then dividing throughout by 17 shows that $17 \mid v$ as well, which can't happen. Therefore, $q \neq 17$. Now if $q \mid u$ or $q \mid v$, then we would conclude from 1.22.15 that q divides both u and v (check!), which is a contradiction to $\gcd(u,v) = 1$. Therefore, $q \nmid uv$, and then we conclude from taking 1.22.15 mod q that $\left(\frac{17}{q}\right) = 1$. Quadratic reciprocity then implies that $\left(\frac{q}{17}\right) = 1$ as well (check!). Since this is true of every prime $q \mid w$ (we do not exclude the cases $w = \pm 1$), and since $17 \equiv 1 \pmod 4$, this is sufficient to conclude that $\left(\frac{w}{17}\right) = 1$, i.e., that w is a quadratic residue mod 17. Finally, consider the equation 1.22.15 mod 17. If $17 \mid u$ or $17 \mid w$, then 17 divides both u and w, and then 1.22.15 implies that $17 \mid v$ as well, which is a contradiction. Since $\left(\frac{w}{17}\right) = 1$, it then follows from 1.22.15 that 2 is a *quartic* residue mod 17, but it is in fact not. This is the contradiction that proves that 1.22.15 has no

rational solution, and hence that $[2] \notin \operatorname{im}\alpha'$. I will leave the analysis of the cases $b_1 = -2$ and $b_1 = 17$ in this case to you as an exercise on the last sheet.
Let us note something very interesting going on here. It can be shown by a very easy application of Hensel's Lemma that 1.22.15 is solvable locally over $\mathbb{Q}_q$ for each prime $q \leq \infty$. Therefore, the fact that it is not solvable over $\mathbb{Q}$ is truly a global phenomenon, and indeed a counterexample to the local-to-global principle for such objects. In the language specific to elliptic curves, this principal homogeneous space (1.22.15) corresponds to a nontrivial element of the Tate-Shafarevich group $Ш(E_{17})$ of the curve $E_{17}/\mathbb{Q}$ (see [8, §X.4]).

Another counterexample of the local-to-global principle (Selmer's counterexample, 1.2.7) will be left for you to analyze on the last sheet.

Remark 1.22.16. The above example (especially the case of the general p with $p \equiv 5 \pmod 8$ and the case of $p = 17$) shows that in general it is not easy to determine which homogeneous spaces arising in the above computation have rational points, since these spaces don't follow the local-to-global principle. This is precisely what makes all current proofs of Mordell-Weil nonconstructive; we just have no algorithm guaranteed to find points on homogeneous spaces or prove that they don't exist.

1.22.C Mazur's Torsion Theorem and Rank Records

We have seen now that we can attach two important arithmetic invariants to an elliptic curve $E/\mathbb{Q}$, namely its rank $r := \operatorname{rank} E$, and its torsion subgroup $T = \operatorname{Tors} E(\mathbb{Q})$. It is a very natural question to ask what values of r and what isomorphism types of T are possible for elliptic curves over $\mathbb{Q}$.

The case of isomorphism types of torsion subgroups is well-understood; the definitive result along this direction is

Theorem 1.22.17 (Mazur). Let $E/\mathbb{Q}$ be an elliptic curve. Then $\operatorname{Tors} E(\mathbb{Q})$ is isomorphic to one of the following 15 groups:

(a) $\mathbb{Z}/n\mathbb{Z}$ for $n = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$, or 12, or
(b) $\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z}$ for $n = 2, 4, 6, 8$.

Each of these groups appears as the torsion subgroup of infinitely many elliptic curves $E/\mathbb{Q}$.

Remark 1.22.18.

(a) This theorem (1.22.17)–conjectured first by Italian mathematician Beppo Levi (1875-1961) in 1908, but known more colloquially as Ogg's conjecture after American mathematician Andrew Ogg (b. 1931)–was proven by the American mathematician Barry Mazur (b. 1937) in 1978, although several special cases (such as the nonexistence of eleven- or sixteen-torsion) were known previously by the work of Levi, Lind Billing-Mahler, and others, which used only elementary techniques.
(b) For each of the isomorphism types allowed by Mazur's Theorem, one can write down explicit families of elliptic curves with that specific torsion subgroup; one such table can be found in [13, Appendix E]. We will discuss some of these "torsion families" when we meet the Tate normal form for elliptic curves next week.
(c) It is a very natural question to ask what the analog of Mazur's Theorem is over arbitrary number fields. A detailed account of this story can be found on the Wikipedia page about Mazur's Theorem, to which I direct the interested reader.

The question of allowable ranks is much more subtle. As of this writing (2026), it is an open problem whether infinitely many values of r are possible, or equivalently if the set of allowable r is bounded. We have evidence that suggests that this should not be the case (e.g., the result is false in the "geometric case", i.e., if we replace $\mathbb{Q}$ by a function field such that $\mathbb{F}_p(t)$), but this is not proof. The curve with the current rank record (as of 2026) of $r \geq 29$ was found by Noam Elkies and Zev Klagsbrun in 2024 and has the form

$$E : y^2 + xy = x^3 + a_4 x + a_6,$$

where

$$a_4 = -270061832416309222184346521452974537847680546218363579547373 85$$

and

$$a_6 = 5525805855134237647573669959111819182152106703253507960837240477914941327771617342563672149 7.$$

The rank of this curve is expected to be precisely 29 (this follows from the generalized Riemann Hypothesis), but this fact is not yet unconditionally known. For much more on the history of rank records, rank records with specified torsion, and various other things of this sort, I direct the reader to the website maintained by Croatian mathematician Andrej Dujella, linked here.

1.23 26/07/20 - The "Fruits Meme", Singular Cubic Curves, and the Tate Normal Form

I want to spend the last two lectures discussing topics requested by the course participants. We'll start with the "fruits meme".

1.23.A The "Fruits Meme"

The following discussion has been adapted from [19] and [20]. Consider the following "meme".

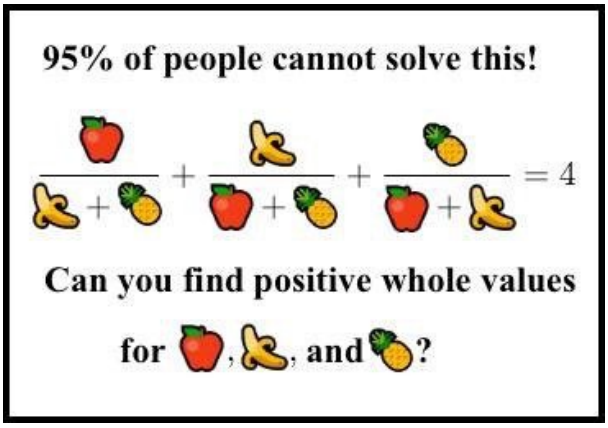


Figure 1.23.1: The "fruits meme". Credit: Sridhar Ramesh. See the discussion in [19].

For ease of typesetting, let's take the variable names to be a, b, c instead, so that we are trying to solve the equation

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} = 4.$$

One such solution is easily seen to be $(4, -1, 11)$ (check!), but this does not have positive whole a, b, c, so we are not done yet. Well, the obvious thing to do is to cross multiply to get the equation

$$a(a+b)(a+c) + b(b+c)(b+a) + c(c+a)(c+b) = 4(a+b)(b+c)(c+a). \tag{1.23.2}$$

This is a homogeneous cubic equation. Therefore, it defines a cubic curve $C \subset \mathbb{P}^2_{\mathbb{Q}}$, which is easily seen to be smooth (2.6.5(a)).

At this point, the general theory (1.9.5(a)) tells us that C can be used to define an elliptic curve over $\mathbb{Q}$. In fact, it is not hard to see that the point $[a : b : c] = [1 : -1 : 0] \in C(\mathbb{Q})$ is an inflection point (2.6.5(b)), and so we can use it to put C into (short) Weierstrass form by following the strategy of 1.9.6. Doing this, we get via the change of variables

$$(X, Y, Z) = (-28(a+b+2c), 364(a-b), 6a+6b-c) \tag{1.23.3}$$

and $(x, y) := (X/Z, Y/Z)$ the equation

$$E : y^2 = x^3 + 109x^2 + 224x.$$

The details are left as 2.6.5(b).

We see that this curve has a rational 2-torsion point $T = (0, 0)$ and hence has the form we are familiar with from §1.20; it has $(a, b) = (109, 224)$. Now, we can use Nagell-Lutz (1.13.7) to easily determine all torsion points and see that $\operatorname{Tors} E(\mathbb{Q}) \cong_{\mathsf{Ab}} \mathbb{Z}/6\mathbb{Z}$ (see 2.6.5(c)). If we're feeling exceptionally energetic, we can then try to use descent via two-isogeny (1.21.1) along with the rank formula (1.22.4) to determine $\operatorname{rank} E$ (2.6.5(d)); then we can use the technique involving heights described in 2.6.4 to determine explicit Mordell-Weil generators. If we're feeling less energetic, we can get Sage to do this for us, from which we obtain $\operatorname{rank} E = 1$ and the Mordell-Weil generator $P = (-4, 28)$. At this point, given

our understanding of the finite group $\operatorname{Tors} E(\mathbb{Q})$ and 1.21.4, we can completely describe all points on $E(\mathbb{Q})$ as

$$E(\mathbb{Q}) = \{nP + Q : n \in \mathbb{Z}, Q \in \operatorname{Tors} E(\mathbb{Q})\}.$$

Undoing the transformation 1.23.3 then gives us a complete description of all points on 1.23.2, and hence all solutions to the original equation. To finish the discussion, it only remains to determine which (x, y) correspond to positive whole values of a, b, c. This is summarized in

Proposition 1.23.4. A point $(x, y) \in E(\mathbb{Q})$ corresponds to positive (whole) values of a, b, c iff

$$x \in \left(-\frac{109 + 13\sqrt{65}}{2}, -28(2 + \sqrt{3})\right) \cup \left(-28(2 - \sqrt{3}), -\frac{14}{3}\right).$$

Proof. This follows from studying the inverse transformation to 1.23.3; the details are left as 2.6.5(e). A full solution can be found at [20, Theorem 4.1]. ∎

In the real picture of E, this region corresponds to all of the left "oval" or "egg" (except for the left most two-torsion point) up to x-coordinate $x < -28(2+\sqrt{3})$, and then two more "pieces" in the range $-28(2 - \sqrt{3}) < x < -14/3$. Now, the resolution of the problem is straightforward: for $n = \pm 1, \pm 2, \ldots$, compute each of the points $nP+Q$ for $Q \in \operatorname{Tors} E(\mathbb{Q})$, and check if the x-coordinate of this point satisfies 1.23.4. As it turns out, the smallest n for which this happens is $n = 9$ (see 2.6.5(f)); this then gives us positive whole solutions (a, b, c) to the original equation with

$$a = 154476802108746166441951315019919837485664325669565431700026634898253202035277999,$$
$$b = 368751317941299998271978115652254748254929799689719709962831374716372246340555 79, \text{ and}$$
$$c = 437361267792869725786125260237139015281653755816161361862143799337842346777 2036,$$

as you can check! This is the origin of the huge solutions to question posed in the "fruits meme". Note that a priori it is possible that no value of n works; this does indeed happen for some similar equations (for instance, if we replace 4 by 5; see [20, Theorem 5.1]), but in this case we get lucky that we don't have to search very far.

In closing, we remark that this procedure–or indeed similar ones involving elliptic curves–don't always produce solutions to such Diophantine equations in increasing order of size, and it is often trickier to prove that any given candidate solution is *the smallest* solution. In this present case, we fortunately have a very simple tool to establish this: the height of points on E. Indeed, we have:

Proposition 1.23.5. Suppose $R = (x, y) \in E(\mathbb{Q})$ corresponds to positive whole values a, b, c with $\gcd(a, b, c) = 1$. Then

$$\max\{\log(a), \log(b)\} > \frac{3}{2}\hat{h}(R) - (10 + 12\log 2).$$

Proof. This is [20, Theorem 7.1]. ∎

In particular, since for each $n \in \mathbb{Z}$ we have $\hat{h}(nP + Q) = n^2\hat{h}(P)$ (see 1.17.6 and 2.5.7); this tells us that as n increases, so does the size of $R = nP + Q$, and hence the size of the corresponding a, b, c. Using this, it can be then verified that the (roughly) 80-digit solutions for a, b, c obtianed above are indeed the "smallest" positive solutions to the "fruits meme". What is even more impressive is that if we replace 4 by 896, then the resulting seemingly innocent Diophantine equation has positive whole solutions, but the smallest solution–which corresponds to $n = 161477$–has a, b, c with *trillions* of digits! For another example where determining the smallest solution to a particular Diophantine equation is very difficult indeed, we direct the reader to the example of, and paper cited in, 1.22.9.

For a complete discussion of the "fruits meme" problem–for arbitrary N instead of $N = 4$–we direct the reader to the highly readable article [20] (popularized by the Quora answer [19]). For many other uses of elliptic curves in "recreational number theory", we direct the reader to [21].

1.23.B Singular Cubic Curves and Why to Study Them

We have spent a lot of time discussing the case of smooth cubic curves. Let us now spend a few minutes talking about what happens in the singular case. Why might we want to do that? Besides the fact that the theory is beautiful in its own right, it turns out that even if we are only interested in smooth cubics, we will have to deal with singular ones. For instance, given an elliptic curve over $K = \mathbb{Q}$, we can try to find a model of the curve with integral coordinates and try to reduce it modulo various primes p.

Example 1.23.6. Consider the curve $E : y^2 = x^3 + 1$ over $K = \mathbb{Q}$. This has discriminant $\Delta = -432$, and so for any prime $p \geq 5$, it makes sense to reduce E mod p to get the curve $\widetilde{E}_p : y^2 = x^3 + 1$ over $\mathbb{F}_p$. For $p = 2, 3$, we the same equation defines a singular cubic curve.

It turns out that there is a deep and mysterious connection between the set of rational points $E(\mathbb{Q})$ of an elliptic curve $E/\mathbb{Q}$ and the number of points $\#\widetilde{E}_p(\mathbb{F}_p)$ on the reductions $\widetilde{E}_p$ of E modulo various primes p. To formulate this precisely is beyond our means at present, but one deep conjecture along these lines is the Birch and Swinnerton-Dyer Conjecture, about which you can find more in [8, §C.16], and solving which will get you $1,000,000 (and much more!).

Now when you start with a given elliptic curve $E/\mathbb{Q}$ and a specific prime p, you may or may not be able to find an integral model for it which you can reduce mod p to get an elliptic curve over $\mathbb{F}_p$.

Example 1.23.7. Consider the curve $E : y^2 = x^3 + 16$ over $\mathbb{Q}$. A priori, reducing this integral equation mod 2 gives us the singular curve $y^2 = x^3$. However, if we first do the admissible change of coordinates $(x, y) = (4x', 8y' + 4)$ given by $(u; r, s, t) = (2; 0, 1, 0)$, we get instead the model $E' : (y')^2 + y' = (x')^3$ of E, which *does* reduce to an elliptic curve mod 2.

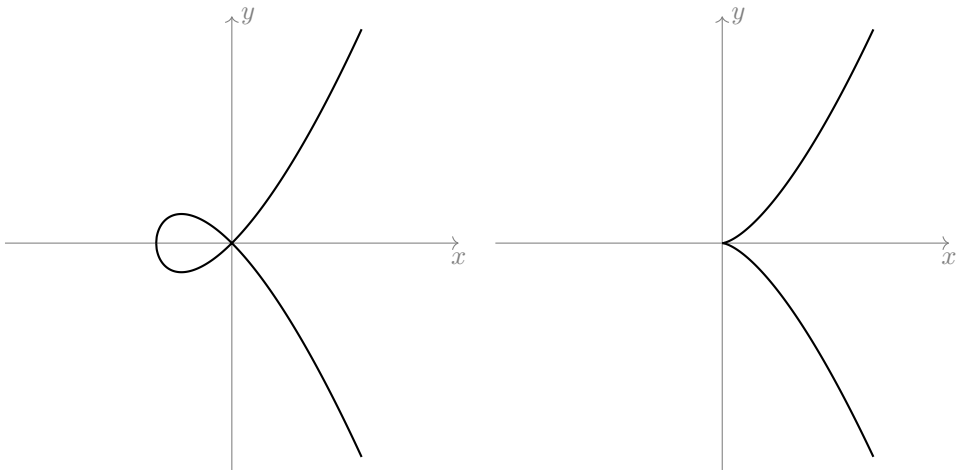
Example 1.23.8. Consider again the curve $E : y^2 = x^3 + 1$ of 1.23.6 over $\mathbb{Q}$. We claim that *there is no* admissible change of coordinates that brings E into a model that has integral coefficients and can be reduced mod 2 to give an elliptic curve over $\mathbb{F}_2$. The reason for this is simple: we know from 1.10.4 that an admissible change of coordinates can only change $v_2(\Delta)$ by a multiple of 12. Since $v_2(\Delta)$ for us is 4, there is no change of coordinates that makes $v_2(\Delta) = 0$, as would be needed for a 2-integral model which reduces mod 2 to an elliptic curve.

Given a curve $E/\mathbb{Q}$ and a prime p, we say that E has *good reduction* at a prime p if there is an admissible change of coordinates that brings E into a form with $v_p(\Delta) = 0$; it is said to have bad reduction at p otherwise. For instance, the curve $E : y^2 = x^3 + 16$ has good reduction at 2 (1.23.7), while the curve $E : y^2 = x^3 + 1$ has bad reduction at 2 (1.23.8). Further, we have seen that a necessary condition for E to have good reduction at p is $v_p(\Delta) \equiv 0 \pmod{12}$. A famous theorem due to Tate ([12, §5.8])–analog of a famous theorem of Minkowski for number fields ([22, Theorem 5.37, Corollary 3])–asserts that for each elliptic curve E, there is a prime p such that E has bad reduction mod p. For more on reductions on elliptic curves mod p, see [1, §10].

Now, the theory of arbitrary singular cubics (over arbitrary fields!) is somewhat fiddly (see, for instance, 2.2.7and 2.6.10), but the basic idea is that, at least for geometrically integral cubics, there is only one singular point, and this singularity is either a node or a cusp (see 1.23.9). The distinction between these two comes from the structure of the "tangent cone" at the singularity–a node arises when there are two distinct "tangent directions" at the singularity, whereas a cusp arises when there is only one. You may find a precise statement of results along this direction in ([6, Exercises 5.10-13]), or attempt to solve 2.6.10. In this section, we will content ourselves by noting that these singular cubic curves have the property that–unlike smooth cubic curves–they admit simple parametrizations. The basic idea is very similar to our parametrization of plane conics with rational points in §1.2.A. We start with the singularity (which we'll assume for simplicity is K-rational, which is not always the case (1.5.5), but which *is* true if K is perfect and the cubic geometrically integral; see 2.2.7), and consider the lines passing through it. By 2.2.7(a) each such line line intersects the cubic to multiplicity two at the singularity, and hence by the case of Bézout's Theorem already proven (1.7.1) intersects the curve in a unique third point, giving us the required parametrization. Let's illustrate this procedure by considering an example.

Consider the nodal cubic $C : y^2 = x^3 + x^2$ with singularity at the origin $P = (0, 0)$. A line through P has the form $y = mx$ for some m. Substituting $y = mx$ into $y^2 = x^3 + x^2$ gives us the cubic equation

$$x^2(x + 1 - m^2) = 0.$$



(a) The nodal cubic $y^2 = x^3 + x^2$. (b) The cuspidal cubic $y^2 = x^3$.

Figure 1.23.9: Examples of singular plane cubics.

As expected, this has two roots at $x = 0$; the third root is then $x = m^2 - 1$, which gives us $y = m(m^2 - 1)$. We obtain the parametrization $(m^2 - 1, m(m^2 - 1))$ of points on the singular cubic (note that this works over any field!). Doing the same for the cuspidal cubic $y^2 = x^3$ gives the parametrization (m^2, m^3) instead.

Finally, we saw in 1.7.B that if C/K is a smooth cubic and $\mathscr{O} \in C(K)$, then the operation $*$ on C gives us a way of giving a group structure on $C(K)$; this group structure is further characterized by 1.12.18, at least when $\mathscr{O}$ is a flex point. Does the same work for singular cubics? Well, something clearly goes wrong if we include the singularity, so a better question to ask is: does the same work if we exclude the singularity? Surprisingly, the answer is "yes".

Let's work this out in the case of the cuspidal cubic $y^2 = x^3$, which has projective completion $C := \mathbb{V}(Y^2Z - X^3) \subset \mathbb{P}^2_K$. Let $\mathscr{O} := [0 : 1 : 0] \in C(K)$ as before, and let $P_0 := [0 : 0 : 1] \in C(K)$ be the singularity. Suppose we are interested in a group law which still satisfies 1.12.18; for this we should study when three points, say $P_i \in C(K) \smallsetminus \{P_0\}$ for $i = 1, 2, 3$ are collinear. To do this, it is easiest to switch to $(t, s) = (X/Y, Z/Y)$ coordinates as in 1.3.11. This affine patch misses precisely one point on C, namely P_0. The rest of the curve, $C \smallsetminus \{P_0\}$ is given by the equation

$$s = t^3.$$

This evidently admits a simple parametrization $(t, s) = (t, t^3)$; this is related to the previous parametrization by $t = 1/m$. Now suppose that the points $P_i = (t_i, t_i^3)$ for $i = 1, 2, 3$ are collinear; this means that the line joining them is not vertical (check!), and hence has equation $\ell : s = \lambda t + \nu$ for some $\lambda, \nu \in K$. It follows that the t_i are the roots of the cubic equation

$$t^3 - \lambda t - \nu = 0,$$

and so Vieta's formulae tell us that these satisfy

$$t_1 + t_2 + t_3 = 0.$$

By 1.15.5, this suggests that the map

$$t : C(K) \smallsetminus \{P_0\} \to \mathbb{G}_a(K) = K$$

might be a group homomorphism, where on the right side $\mathbb{G}_a$ just means "the additive group of". (Similarly, the notation $\mathbb{G}_m$ just means the "multiplicative group of", so that $\mathbb{G}_m(K) = K^\times$.) This is indeed the case, as we leave for the reader to check (2.6.10(c)), and we conclude that the group law one might obtain from the operation $*$ on the set of smooth points of the singular cubic $y^2 = x^3$ is simply the additive group law on the field K itself. Something similar can be said in the nodal case, where we end up with the group $\mathbb{G}_m(K) = K^\times$ instead (2.6.10(d)). We thus conclude that singular cubics have very simple group laws (at least as compared to those on elliptic curves).

Unimportant Remark 1.23.10. If $E/\mathbb{Q}$ is an elliptic curve and p a prime, then there is a notion of a *global minimal model* of E which has integral coefficients, and can be reduced modulo every prime p to get the "best possible" model over $\mathbb{F}_p$. (The same can be done over any number field K with class number $h(K) = 1$; see [8, §VIII.8].) In particular, a given prime p is a prime of good reduction of E iff the minimal model reduces to an elliptic curve over $\mathbb{F}_p$, i.e., defines a smooth cubic over $\mathbb{F}_p$. When this model reduces to a nodal cubic over $\mathbb{F}_p$, then we say that E has *multiplicative (bad) reduction* at p, and when it reduces to a cuspidal cubic over $\mathbb{F}_p$, we say that E has *additive (bad) reduction* at p. Ultimately, the reason for this terminology is precisely the structure of the "degenerate" group laws discussed above. For more on this, we refer the reader to [8].

We will meet more examples of such parametrizations next time (§1.24.A). Let us now turn to a slightly different topic.

1.23.C Normal Forms

Suppose now that we are interested in the "general form" of a curve with a torsion point of a specified order. How might we describe such a curve? In particular, can we construct curves with torsion points of any given order?

Let K be a field and $a_1, \dots, a_6 \in K$ as usual such that $\Delta \neq 0$, so that

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

defines an elliptic curve E/K. If $E(K) = \{\mathscr{O}\}$, there is nothing more to say.

Example 1.23.11. When $K = \mathbb{Q}$ and $E/\mathbb{Q}$ is given by the Weierstrass equation

$$E : y^2 = x^3 + x^2 - 18x - 43,$$

then $E(\mathbb{Q}) = \{\mathscr{O}\}$, according to the LMFDB. The reader wishing to obtain this result by hand can use Nagell-Lutz (1.13.7) to show first that $\operatorname{Tors} E(\mathbb{Q}) = \{\mathscr{O}\}$. At this point, a procedure similar to (but quite a bit harder than) the one found in 2.5.17 can be used to show that $\operatorname{rank} E = 0$, which along with 1.15.8 and 1.21.4 is enough to conclude that $E(\mathbb{Q}) = \{\mathscr{O}\}$.

If $E(K) \neq \{\mathscr{O}\}$, then we may pick a $P \in E(K) \smallsetminus \{\mathscr{O}\}$. By an admissible change of coordinates, we can move P to $(0, 0)$, the effect of which is to kill of a_6; in other words, we have $P = (0, 0) \in E(K)$ iff $a_6 = 0$. At this point, §1.12.C tells us that $-P = (0, -a_3)$. Therefore, we conclude that $2P = \mathscr{O}$ iff $P = -P$ iff $a_3 = 0$. We have shown

Proposition 1.23.12. Let K be a field and let E/K be an elliptic curve with a point $P \in E[2](K) \smallsetminus \{\mathscr{O}\}$. Then there is a(n) (admissible) change of coordinates bringing E into the form

$$E : y^2 + a_1xy = x^3 + a_2x^2 + a_4x$$

for some $a_1, a_2, a_4 \in K$ such that

$$\Delta := a_4^2(a_1^4 + 8a_1^2a_2 + 16a_2^2 - 64a_4) \neq 0,$$

and taking P to the point $(0, 0) \in E[2](K) \smallsetminus \{\mathscr{O}\}$.

Note that this works in any characteristic, and so gives us way to construct curve in characteristic two with 2-torsion points.

Example 1.23.13. If K is a field of characteristic two, then in 1.23.12, the expression for Δ simplifies tremendously to $\Delta = a_1^4a_4^2$. In particular, the equation $E : y^2 + a_1xy = x^3 + a_2x^2 + a_4x$ defines an elliptic curve iff a_1 and a_4 are both nonzero. For instance, $E : y^2 + xy = x^3 + x$ defines an elliptic curve over $K = \mathbb{F}_2$ with two-torsion point $P = (0, 0) \in E[2](\mathbb{F}_2) \smallsetminus \{\mathscr{O}\}$.

Now returning to before 1.23.12, suppose that $P \notin E[2]$, so that $a_3 \neq 0$. In this case, the admissible change of coordinates given by replacing y by $y - a_4a_3^{-1}x$ has the effect of getting rid of a_4, and we have shown

Proposition 1.23.14. Let K be a field and E/K an elliptic curve with a point $P \in E(K) \smallsetminus E[2](K)$. There is a(n) (admissible) change of coordinates bringing E into the form

$$E : y^2 + a_1xy + a_3y = x^3 + a_2x^2$$

for some $a_1, a_2, a_3 \in K$ such that

$$\Delta = -a_3^2(a_1^4a_2 - a_1^3a_3 + 8a_1^2a_2^2 - 36a_1a_2a_3 + 16a_2^3 + 27a_3^2) \neq 0,$$

and taking P to $(0,0) \in E(K) \smallsetminus E[2](K)$.

Geometrically, the condition $a_3 \neq 0$ means that the tangent line to P at $(0,0)$ is not vertical, and in this case the above transformation "flattens it out" to be horizontal instead. Explicitly, we see that form the form of the curve in 1.23.14, the tangent line T_PE is given by simply $y = 0$.

Example 1.23.15. Taking $K = \mathbb{F}_2$ and $a_1 = a_2 = 0$ and $a_3 = 1$ gives us the elliptic curve $E : y^2 + y = x^3$ over $\mathbb{F}_2$ with the point $P = (0,0) \in E(\mathbb{F}_2) \smallsetminus E[2](\mathbb{F}_2)$. It is easy to check that in fact $E(\mathbb{F}_2) = \{\mathscr{O}, (0,0), (0,1)\} \cong_{\mathsf{Ab}} \mathbb{Z}/3$. We met this curve already in 1.12.9(b).

Now suppose we are in the situation of 1.23.14. We know from §1.12.E that $3P = \mathscr{O}$ iff P is an inflection point on E. Since $\mathrm{T}_PE = \mathbb{V}(y)$, this happens iff the equation $x^3 + a_2x^2 = 0$ has a triple root at $x = 0$, i.e., if $a_2 = 0$. We have shown

Proposition 1.23.16. Let K be a field and E/K an elliptic curve with a point $P \in E[3](K) \smallsetminus \{\mathscr{O}\}$. There is a(n) (admissible) change of coordinates bringing E into the form

$$E : y^2 + a_1xy + a_3y = x^3$$

for some $a_1, a_3 \in K$ such that

$$\Delta = a_3^3(a_1^3 - 27a_3) \neq 0,$$

and taking P to $(0,0) \in E[3](K) \smallsetminus \{\mathscr{O}\}$. This curve E has $c_4(E) = a_1(a_1^3 - 24a_3)$, and hence j-invariant

$$j(E) = \frac{a_1^3(a_1^3 - 24a_3)^3}{a_3^3(a_1^3 - 27a_3)}.$$

In fact, we can do slightly better. It is now helpful to split into two cases depending on whether or not $\operatorname{ch} K = 3$. First suppose that $\operatorname{ch} K = 3$. Then Δ futher simplifies to $\Delta = a_1^3a_3^3$, so for this to be an elliptic curve, we must have $a_1, a_3 \neq 0$. The admissible change of coordinates $(u; r, s, t) = (a_1^{-1}; 0, 0, 0)$ further sets $a_1 = 1$, and we obtain

Proposition 1.23.17. Let K be a field with $\operatorname{ch} K = 3$ and let E/K be an ordinary elliptic curve with $P \in E[3](K) \smallsetminus \{\mathscr{O}\}$. There is a(n) (admissible) change of coordinates bringing E into the form

$$E_\beta : y^2 + xy + \beta y = x^3$$

for some unique $\beta \in K^\times$, and taking P to $(0,0) \in E_\beta[3](K) \smallsetminus \{\mathscr{O}\}$. This curve has discriminant $\Delta(E_\beta) = \beta^3$ and j-invariant $j(E_\beta) = \beta^{-3}$.

The uniqueness of β follows immediately from the expression for $j(E_\beta)$ and 1.11.16; this β is none other than the quantity b_2 of the curve E, i.e., $\beta = b_2(E_\beta)$. Thanks to the discussion following 1.12.23, we have thus obtained the general form of an ordinary elliptic curve over a perfect field of characteristic three. The reader preferring short Weierstrass forms in characteristic not two can check that this curve can be written via a simple admissible change of coordinates as

$$E'_\beta : y^2 = x^3 + x^2 - \beta x + \beta^2$$

instead.

Finally, suppose that $\operatorname{ch} K \neq 3$; for simplicity, assume further that K is algebraically closed as well, but see 1.23.20 and 1.23.21. In this case, from $\Delta \neq 0$ we conclude that $a_3 \neq 0$. Since K is algebraically closed, we can rescale by a suitable u to achieve $a_3 = 1$. From this, we get

Proposition 1.23.18 (Deuring Normal Form)**.** Let K be an algebraically closed field with $\operatorname{ch} K \neq 3$, and let E/K be an elliptic curve and $P \in E[3](K) \smallsetminus \{\mathscr{O}\}$. There is a(n) (admissible) change of coordinates bringing E into the form

$$E_\alpha : y^2 + \alpha xy + y = x^3$$

for some $\alpha \in K^\times$ with $\alpha^3 \neq 27$, and taking P to $(0,0) \in E_\alpha[3](K) \smallsetminus \{\mathscr{O}\}$. This curve has discriminant $\Delta(E_\alpha) = \alpha^3 - 27$ and j-invariant

$$j(E_\alpha) = \frac{\alpha^3(\alpha - 24)^3}{\alpha^3 - 27}.$$

Exercise 1.23.19. Suppose K is a field with $\operatorname{ch} K \neq 3$, and let $E : y^2 + y = x^3$. Let $T := (0,0) \in E(K)$. Figure out an algebraic formula for the operation $\tau_T : E(K) \to E(K)$ given by $\tau_T(P) = P + T$ for $P \in E(K)$, and verify directly that $\tau_T^3 = \mathrm{id}$.

Exercise 1.23.20. Suppose $K = \mathbb{F}_2(t)$. Show that the curves $E_i : y^2 + t^i y = x^3$ for $t = 0, 1, 2$ are cubic twists of each other, i.e., pairwise nonisomorphic over K, but isomorphic over $\overline{K}$.

Exercise 1.23.21. Analyze more generally what happens when K is an arbitrary (i.e., not necessarily algebraically closed field) with $\operatorname{ch} K \neq 3$. You should get a classification using $K^\times/(K^\times)^3$.

We now return to the case of general (not necessarily algebraically closed) field K, and in which $a_2 a_3 \neq 0$, which corresponds to a point P that is not n-torsion for $n = 1, 2, 3$. At this point, we can certainly find a $u \in K^\times$ such that $u^2 a_2 = u^3 a_3$. Calling this quantity $-b$ (and calling $ua_1 = (1 - c)$) immediately gives us

Proposition 1.23.22 (Tate Normal Form)**.** Let K be a field, E/K an elliptic curve, and $P \in E(K)$ such that $P, 2P, 3P \neq \mathscr{O}$. Then there is a(n) (admissible) change of coordinates bringing E into the form

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

for some $b, c \in K$ such that

$$\Delta = b^3(c^4 - 8bc^2 - 3c^3 + 16b^2 - 20bc + 3c^2 + b - c) \neq 0,$$

and taking P to $(0,0) \in E_{b,c}(K)$.

Note that this proposition The Tate normal form (1.23.22) is true in any characteristic. It is an incredibly useful result, and we will see next time how to use it to give parametrizations of various cyclic torsion structures.

1.24 26/07/21 - Parametrization of Torsion Structures by Modular Curves, Lenstra's Algorithm, and Outlook

Today, I want to finish the story involving the Tate normal form (1.23.22), talking about one relationship the theory of elliptic curves has with cryptography, and end with an outlook on what comes next.

1.24.A Parametrizations of Torsion Structures and the Modular Curves $X_1(N)$

Let K be a field and $b, c \in K$ with

$$\Delta := b^3(c^4 - 8bc^2 - 3c^3 + 16b^2 - 20bc + 3c^2 + b - c) \neq 0.$$

Then

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

defines an elliptic curve over K, with the point $P = (0,0) \in E(K)$ satisfying $P, 2P, 3P \neq \mathscr{O}$; conversely, we showed in 1.23.22 that any elliptic curve E/K with a point $P \in E(K)$ such that $P, 2P, 3P \neq 0$ can be brought into this form, with P sent to $(0,0)$. The point of renaming the coefficients in this particular way is that the coefficients of the multiples of $P \in E_{b,c}(K)$ are very easy to write out. Explicitly, we have

n	nP	$-n$	$-nP$
1	$(0,0)$	-1	$(0,b)$
2	(b,bc)	-2	$(b,0)$
3	$(c,b-c)$	-3	(c,c^2)
4	$\left(\frac{b(b-c)}{c^2}, \frac{b^2(c^2+c-b)}{c^3}\right)$	-4	$\left(\frac{b(b-c)}{c^2}, \frac{b(b-c)^2}{c^3}\right)$

Now suppose that we ask for $4P = \mathscr{O}$, which is equivalent to asking for $2P = -2P$. From the above table, and using that $b \neq 0$, we see that this happens iff

$$X_1(4) : c = 0. \tag{1.24.1}$$

The reason for the name $X_1(4)$ for the simple equation $c = 0$ will soon be clear. Note that the fact that $c = 0$ is equivalent to $4P = \mathscr{O}$ is the reason that c has to appear in the "denominators" of the formulae for $\pm 4P$ in the above table. In any case, we have proven

Proposition 1.24.2. Let K be a field, E/K an elliptic curve, and $P \in E[4](K) \smallsetminus E[2](K)$. Then there is a(n) (admissible) change of coordinates bringing E into the form

$$E_b : y^2 + xy - by = x^3 - bx^2$$

for some $b \in K$ with

$$\Delta = b^4(16b + 1) \neq 0,$$

and taking P to $(0,0) \in E_b(K)$.

Exercise 1.24.3. How does 1.24.2 compare with 2.2.8?

Example 1.24.4. Taking $K = \mathbb{F}_2$ and $b = 1$ gives us the curve $E : y^2 + xy + y = x^3 + x^2$ over $\mathbb{F}_2$ which has $P \in E[4](\mathbb{F}_2) \smallsetminus E[2](\mathbb{F}_2)$. In fact, $E(\mathbb{F}_2) \cong_{\mathsf{Ab}} \mathbb{Z}/4$ with generators $(0,0)$ and $(0,1)$. Indeed, we may take $b = 1$ for any field K with $\operatorname{ch} K \neq 17$; when $\operatorname{ch} K = 17$, we may take $b = -1$ instead.

Now suppose we asked for $5P = \mathscr{O}$ instead, or equivalently $3P = -2P$. From the above table, we see that this happens iff

$$X_1(5) : b = c. \tag{1.24.5}$$

We have proven

Proposition 1.24.6. Let K be a field, E/K an elliptic curve, and $P \in E[5](K) \smallsetminus \{\mathscr{O}\}$. Then there is a(n) (admissible) change of coordinates bringing E into the form

$$E_b : y^2 + (1-b)xy - by = x^3 - bx^2$$

for some $b \in K$ with

$$\Delta = b^5(b^2 - 11b - 1) \neq 0,$$

and taking P to $(0,0) \in E_b(K)$.

Example 1.24.7. For any field K with $\operatorname{ch} K \neq 11$, we may take $b = 1$, which gives us the curve $E : y^2 - y = x^3 - x^2$ having $P = (0,0) \in E[5](K) \smallsetminus \{\mathscr{O}\}$. For instance, taking $K = \mathbb{Q}$ and transforming to short Weierstrass form yields the curve $E/\mathbb{Q}$ given by

$$E : y^2 = x^3 - 432x + 8208$$

with $P = (-12, -108) \in E[5](\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$.

Similarly, asking for $6P = \mathscr{O}$ amounts to asking for $3P = -3P$, which according to the above table happens iff

$$X_1(6) : b = c^2 + c. \tag{1.24.8}$$

We have proven

Proposition 1.24.9. Let K be a field, E/K an elliptic curve, and $P \in E(K)$ a point of exact order 6. Then there is a(n) (admissible) change of coordinates bringing E into the form

$$E_c : y^2 + (1-c)xy - (c^2+c)y = x^3 - (c^2+c)x^2$$

for some $c \in K$ with

$$\Delta = c^6(c+1)^3(9c+1) \neq 0$$

and taking P to $(0,0) \in E_c(K)$.

Example 1.24.10. For any field K with $\operatorname{ch} K \neq 2, 5$, we may take $c = 1$, which gives us the curve

$$E : y^2 - 2y = x^3 - 2x^2$$

having $P = (0,0) \in E(K)$ of exact order 6. If we like, we can complete the square on the left side to transform this curve into the one with equation

$$E' : y^2 = x^3 - 2x^2 + 1,$$

which has the point $(0,1)$ of exact order 6.

Let's do two more examples. What happens if we ask for $7P = \mathscr{O}$, which is equivalent to $3P = -4P$? Well, from $3P \neq \mathscr{O}$, we would then conclude that $4P \neq \mathscr{O}$, so certainly $c \neq 0$ is necessary. Next, from the above table, we see that $3P = -4P$ is apparently *two* conditions on b and c, namely

$$c = \frac{b(b-c)}{c^2} \text{ and } b - c = \frac{b(b-c)^2}{c^3},$$

but indeed these two conditions are one and the same, namely

$$X_1(7) : b^2 - bc - c^3 = 0. \tag{1.24.11}$$

This is more interesting. The equation $X_1(7)$ defines a singular cubic in the (b,c)-plane, with a nodal singularity at $(0,0)$ (check!). Therefore, we may use the technique of 1.23.B to parametrize it. Explicitly, we write $b = tc$ for some parameter t and plug it into the equation for $X_1(7)$ from 1.24.11 to arrive at

$$c^2(t^2 - t - c) = 0.$$

Two of the roots of this cubic in c are at $c = 0$ as expected; the third intersection point corresponds to $c = t^2 - t$ and hence $b = t^3 - t^2$. This gives us

Proposition 1.24.12. Let K be a field, E/K an elliptic curve, and $P \in E[7](K) \smallsetminus \{\mathscr{O}\}$. Then there is a(n) (admissible) change of coordinates bringing E into the form

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

for some $t \in K$ with

$$\Delta = t^7(t-1)^7(t^3 - 8t^2 + 5t + 1) \neq 0,$$

and taking P to $(0,0) \in E_t(K)$.

Example 1.24.13. Let K be a field with $\operatorname{ch} K \neq 2, 13$. Then we may take $t = -1$ in the above to get the curve

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

with the 7-torsion point $P = (0,0) \in E[7](K) \smallsetminus \{\mathscr{O}\}$. Since $\operatorname{ch} K \neq 2$, we can use an admissible change of coordinates (check!) to bring E into the form

$$E' : y^2 = x^3 - 43x + 166,$$

on which P gets sent to the 7-torsion point $(3,8) \in E'[7](K) \smallsetminus \{\mathscr{O}\}$.

Finally, the condition $8P = \mathscr{O}$ is equivalent to $4P = -4P$. Looking at the table tells us that again we should ask for $c \neq 0$ and also

$$\frac{b^2(c^2 + c - b)}{c^3} = \frac{b(b-c)^2}{c^3},$$

which in light of $b \neq 0$ amounts to the cubic equation

$$X_1(8) : bc^2 + 3bc - 2b^2 - c^2 = 0. \tag{1.24.14}$$

Again, this defines a nodal cubic, and we can proceed as above to find a parametrization of $X_1(8)$.

Exercise 1.24.15. Finish the above calculation, i.e., find a parametrization of the nodal cubic 1.24.14 and use it to deduce the general form of an elliptic curve with an 8-torsion point as in 1.24.12. Use this to construct an elliptic curve over $\mathbb{Q}$ with a rational point of exact order 8.

Exercise 1.24.16. Push the above calculations a little further. Figure out the coordinates of $\pm 5P$, and use this to determine the equation defining $X_1(9)$ as we did for $X_1(N)$ for $N \leq 8$ above. What do you notice? How much further can you go?

At this point, the general pattern should be clear. For each integer $N \in \mathbb{Z}_{\geq 4}$, there is a *single* equation $f_N(b,c)$ in b and c defining the *modular curve* $X_1(N) : f_N(b,c) = 0$, such that the points $(b,c) \in X_1(N)(K)$ over any field K satisfying $\Delta(b,c) \neq 0$ give rise to elliptic curves $E_{b,c}/K$ on which the point $P = (0,0) \in E(K)$ has order dividing N (and order exactly N if we impose some other algebraic nonvanishing conditions involving b and c, such as $c \neq 0$). (This can be proven by a careful study of division polynomials as in 2.2.10; there is also a way to modify the above theory to work for $N \leq 1, 2, 3$ as well, but let's not bother.) This has the amazing consequence that the existence and nonexistence of points over a given field K on the modular curve $X_1(N)$ directly determines the existence or nonexistence of elliptic curves over K with specified torsion. In general, the equations $f_N(b,c)$ obtained by the above technique quickly become quite unwieldy, and mechanisms are needed to bring them into simpler forms; one such algorithm is discussed in detail in the wonderful paper [23]. Let's conclude this discussion

Example 1.24.17. After a suitable transformation, the modular curve $X_1(11)$ can be brought into the form $v^2 - v = u^3 - u^2$ (see [23, Table 3])–but this is precisely the curve of 1.24.7 ! This means that the existence or non-existence of an elliptic curve with 11-torsion point on it is determined by the set of rational points of the elliptic curve $E : y^2 = x^3 - 432x + 8208$ over $\mathbb{Q}$. As it turns out, is that $\operatorname{rank} E = 0$, and that all rational points on E give rise to singular curves (i.e., basically satisfy $\Delta(b,c) = 0$). This is enough to conclude that there does not exist an elliptic curve $E/\mathbb{Q}$ with a nontrivial rational 11-torsion point. This is one flavor of the kind of argument that goes into proving some "elementary" cases of Mazur's torsion theorem (1.22.17); see 1.22.18(a).

These modular curves $X_1(N)$ have many fantastic properties, and are connected to many different things in mathematics (e.g., modular forms). It is beyond our means at the present to go into details here, but I direct the interested reader to [12, Chapter 11]. Similarly, the reader interested in parametrizations of noncylic torsion structures can find more details in [12, §4.6].

1.24.B Lenstra's Factoring Algorithm

I want to end this course by briefly mentioning how elliptic curve theory connects to cryptography. This is a vast subject, and we will not have time to say much here; I want to highlight just one specific topic, namely Lenstra's factoring algortihm. We will loosely follow [2, §4.4] and [8, §X.2].

It is well-known that the strength of many modern public key cryptosystems relies on our inability to factor very large numbers. Probably the most famous of these cryptosystems is the RSA cryptosystem, devised in 1977 by Rivest, Shamir, and Adleman (see 2.6.6 for a reminder on how the RSA cryptosystem works). For this reason, we are continually in the search of quicker and more efficient factoring algorithms. There are many such algorithms in place today; probably the fastest of these is the *general number field sieve* (see [24] for a quick introduction with several references). For the purposes of illustration, we will highlight one algorithm–called Lenstra's factoring algorithm–which uses elliptic curves.

To do this, we first need to talk about elliptic curves over more general rings than just fields. To do this properly would need us to use the language of schemes; we will content ourselves with

Definition 1.24.18. Let R be a ring. An *elliptic curve E over R* is defined by an equation of the form

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

for $a_1, a_2, a_3, a_4, a_6 \in R$ such that the discriminant Δ (given in 1.9.2) is a unit in R.

Note that if K is a field, this basically recovers the familiar definition (1.9.2). Again, given a ring R, an elliptic curve E/R, and an "extension" (i.e., an algebra) L/R, it makes sense to talk about the L-points $E(L)$ of R; these are precisely the pairs (x, y) of elements of L which satisfy the above Weierstrass equation (along with a suitable analog of the point $\mathscr{O}$ at infinity). A lot of the theory from before carries over, but not everything does, as we shall see below.

Remark 1.24.19.

(a) For those who know a little scheme theory, the above definition amounts to

$$E := \operatorname{Spec} R[x, y]/(y^2 + a_1xy + a_3y - x^3 - a_2x^2 - a_4x - a_6).$$

It is a highly nontrivial fact (see [25, Theorem 2.1.2]) that such an object (or more precisely its projective completion $\operatorname{Proj} R[X, Y, Z]/(Y^2Z + a_1XYZ + a_3YZ^2 - X^3 - a_2X^2Z - a_4XZ^2 - a_6Z^3)$) is automatically a commutative group scheme over R, i.e., for each extension L/R, the set of points $E(L)$ can be made into a group in essentially the same way as when R is a field.

(b) When $R = \mathbb{Z}/N\mathbb{Z}$ for some squarefree integer N, an elliptic curve over R is specified by coefficients $a_1, \ldots, a_6 \in \mathbb{Z}/N\mathbb{Z}$ such that $\Delta \in (\mathbb{Z}/N\mathbb{Z})^\times$. A version of the Chinese Remainder Theorem then says that the theory of such objects is totally equivalent to the "simultaneous" theory of elliptic curves $E_p/\mathbb{F}_p$ for each prime divisor p of N. We invite the reader to make this as precise as they want.

(c) Sage handles elliptic curves over more general rings very well. For instance, to construct $R = \mathbb{Z}/35\mathbb{Z}$, we can type `R = Zmod(35)`, having done which we can construct a curve over R the same way as over a field, e.g., by `E = EllipticCurve(R, [0, 1])`, which returns the curve $y^2 = x^3 + 1$ over $\mathbb{Z}/35\mathbb{Z}$.

Let us now use this terminology to explain Lenstra's algorithm. Suppose that we are given a large integer N that we are interested in factoring. To do this, we take $R = \mathbb{Z}/N\mathbb{Z}$, and look at a suitable elliptic curve E/R, such as

$$E_b : y^2 = x^3 + bx - b$$

for a small integer b, which has discriminant $\Delta = -16(4b+27)b^4$. Note that for this to be an elliptic curve over R, we need $\Delta \in R^\times$, or equivalently that $\gcd(\Delta, N) = 1$. (If this is not the case, and $b \ll N$, then we have already found a nontrivial divisor of N and we are done.) This choice of elliptic curve has the advantage that it has a readily visible point on it, namely $P = (1,1) \in E(R)$. Now either by pretending R is a field, or using the highly nontrivial 1.24.19(a), we can try to find multiples of points on P on $E(R)$, say $(20!) \cdot P$ for concreteness. (Here, a number like 20! is useful because it is divisible by lots of small primes, or more precisely is *smooth*; we won't say more about this, but the interested reader can find more on this in [1, §26] or [8, §XI.2].) To do a computation of this sort naively would take very long, but we can use more efficient algorithms, such as the familiar (to Ross students!) double-and-add algorithm (as outlined in [2, §4.4]).

Let us try to review what point addition on elliptic curves looks like, using the explicit formula described in §1.12.D. For this, suppose we are given points $P_i = (x_i, y_i) \in E(R)$ for $i = 1, 2$, with $x_i, y_i \in R$, and we are interested in the addition $P_1 + P_2$. Suppose for simplicity that $x_1 \neq x_2 \pmod N$. If R were a field (equivalently if N were prime), then the procedure described in §1.12.D would ask us to first find

$$\lambda = \frac{y_2 - y_1}{x_2 - x_1},$$

and then use it to express the sum $P_3 = P_1 + P_2$ as

$$P_3 = (x_3, y_3) = (\lambda^2 - x_1 - x_2, -\lambda x_3 - (y_1 - \lambda x_1)).$$

In general–for any N–having found λ, the second step is entirely straightforward. The interesting step is the first one, to do which we need to find the inverse of $x_2 - x_1$ in $R = \mathbb{Z}/N\mathbb{Z}$. Here essentially two things can happen:

(a) If $\gcd(x_2 - x_1, N) = 1$, then we can efficiently find the inverse of $x_2 - x_1 \bmod N$ by using the Euclidean algorithm, use this to compute λ, and then move on.
(b) If $1 < \gcd(x_2 - x_1, N) < N$, then we have succeeded in finding a nontrivial factor of N.

(The third case described in [2, §4.4] cannot happen for us, since $x_1, x_2 \in \mathbb{Z}/N\mathbb{Z}$, i.e., we are working mod N to begin with, and we assumed that $x_1 \neq x_2$.) If instead we have $x_1 = x_2$ and $y_1 = y_2$, we can try to use the duplication formula as described in §1.12.D (see 1.12.13), for which the denominator of λ is $2y_1$; then, we are interested in $\gcd(2y_1, N)$, and we do an analysis similar to the one above. Next, if $x_1 = x_2$ but $y_1 \neq y_2$, something very interesting is going on: under the Chinese Remainder Theorem decomposition of 1.24.19(b), we have $P_1 = P_2$ in some "component" $E_p(\mathbb{F}_p)$'s and not in others, and in this case, we necessarily have $1 < \gcd(y_2 - y_1, N) < N$, and we win.

Exercise 1.24.20. Figure out what the previous line means precisely, particularly when N is not known to be squarefree.

The only case in which we get stuck is when $x_1 = x_2$ and $y_1 = -y_2$, in which case the point sum $P_1 + P_2$ is just the point at infinity on $E(R)$; this makes the above computation of $(20!)P$ loop without actually giving us a factor of N. Summarizing this discussion, we see that there are two possibilities:

(a) At some point of time, the computation gets "stuck" and we recover a factor g of N with $1 < g < N$, at which point we are done.
(b) Either we successfully finish the computation of $(20!)P$, or we end up "going to infinity" in one of the computations; either way, we don't get a factor of N. If this happens, we just got unlucky; we pick a different b and try again.

The basic idea behind Lenstra's algorithm is very simple: we just keep incrementing b, until we get to a point where some computation fails, and this (hopefully) gives us a factor g of N; in practice, this is more than sufficient to stumble upon a factor of N, if N is not too big and indeed composite. Since the above algorithm is extremely enjoyable to implement by oneself, I leave the computer implementation of this naive algorithm to the reader as 2.6.7.

In reality, there are several efficiency upgrades we can perform on this basic algorithm. For starters, we don't have to use a fixed "type" of elliptic curve; we might as well try to vary our parameters b, c, and hence choice of curve $E : y^2 = x^3 + bx + c$, and our seed point P. (It turns out that it is advantageous to select these essentially randomly.) For other upgrades, we refer the reader to [1, §26], [2, §4.4], and [8, §XI.2]. Of course, we are only able to scratch the surface of the algorithmic and

cryptographic aspects to the theory of elliptic curves. The curious reader can find the much more that remains to be said in several places such as [2, §4.4-5], [8, Chapter XI], and [12, Appendix II].

1.24.C Outlook: What Next?

It goes without saying that there is much, much more to the theory of elliptic curves than we have been able to cover in this course. However, I hope that this course was able to whet your appetite towards this kind of mathematics (elliptic curve theory specifically, but also algebraic and arithmetic geometry more broadly), and that you are curious to learn more; I end this set of notes by providing a list of suggested further topics along with references.

(a) One aspect of the theory of cubic curves we did not get to talk about at all is the theory of integer points. This is a beautiful subfield of Diophantine approximation, with several results essentially due to Thue. For a quick overview, see [2, Chapter 5], and for a more in-depth study, see [8, Chapter IX].

(b) Another important topic we did not get to talk about (except very briefly in §1.23.B) is the idea of reducing algebraic curves modulo various primes. This particular idea has several applications to elliptic curve theory, such as to the computation of torsion subgroups. The interested reader can field an approachable exposition in in [1, §10] and [2, §4.3].

(c) We saw in §1.13.A and §1.20.C how the complex theory of elliptic curves motivates and supplements the algebraic theory. For a gentle introduction to the complex-analytic theory of elliptic functions, I recommend [15, Chapter 7], and for a slightly more modern, but still accessible, treatment of the theory of Riemann surfaces in general, I would recommend [26]. A motivated exposition, focusing only on the case of elliptic curves, can also be found at [8, Chapter VI]. For more advanced topics such as connections to modular forms and L-functions, I would suggest [13, Chapters 3-5], and eventually [27].

(d) As noted in §1.24.B, the reader interested in more algorithmic and cryptographic aspects to the theory of elliptic curves in places such as [2, §4.4-5], [8, Chapter XI], and [12, Appendix II].

(e) There are many ways to generalize the theory of elliptic curves to higher dimensional varieties (e.g., abelian varieties, Calabi-Yau varieties such as K3 surfaces, and cubic hypersurfaces). Entire books are devoted to these topics, although a detailed study of these requires more algebraic geometry. An obvious candidate for what to study next in algebraic geometry is the theory of general algebraic curves (or indeed algebraic varieties of any dimension); for an accessible introduction, I would recommend [6] and [28], although the definitive reference for an introduction to more serious algebraic geometry is the masterpiece [9].

(f) The reader specifically interested in elliptic curve theory can find more sophisticated treatments of several topics that we covered, as well as many natural further topics such as the theory over local and global fields, Galois cohomology, and complex multiplication, in textbooks such as [8] and [12].

Chapter 2

Exercise Sheets

2.1 Exercise Sheet 1

2.1.A Numerical and Exploration

Exercise 2.1.1. Let $b, c \in \mathbb{R}$. When does the equation $x^3 + bx + c = 0$ have exactly one (resp. two, resp. three) real root(s)? (Hint: Play around on Desmos. A little calculus might come in handy.)

Exercise 2.1.2.

(a) Let R be a UFD with fraction field K. State and prove the analog of the Rational Root Theorem (1.1.2) with $\mathbb{Z}$ replaced by R, and conclude that there is an algorithmic procedure (given the ability to find all factors of an element of R[1]) for finding roots in K of single-variable polynomials with coefficients in K.

(b) Using (a), or otherwise, determine all solutions in the field $\mathbb{Q}(i)$ to

$$2ix^3 + (2 - 4i)x^2 - (2 - 5i)x - (2 - i) = 0.$$

Exercise 2.1.3.

(a) Find all integer solutions (x, y) to $3x + 5y + 4 = 0$ (and prove you have found all of them!).

(b) Do the same for $4x + 6y + 5 = 0$.

(c) For $a, b, c \in \mathbb{Z}$, give a necessary and sufficient criterion for the equation $ax + by + c = 0$ to have (i) infinitely many solutions, (ii) a finite nonzero number of solutions, and (iii) no solutions (x, y) in the integers.

(d) Generalize (c) to an arbitrary number of variables. More precisely, given any $n \in \mathbb{Z}_{\geq 1}$ and integers $a_1, \ldots, a_n, c \in \mathbb{Z}$, find necessary and sufficient conditions for the equation

$$a_1x_1 + \cdots + a_nx_n + c = 0$$

to have integer solutions $(x_1, \ldots, x_n)$ as in (i), (ii), and (iii).

Exercise 2.1.4.

(a) Find all rational solutions (x, y) to $x^2 + 3y^2 = 2$.

(b) Let p be a prime. Find all rational solutions (x, y) to the equation $x^3 + py^3 = p^2$.

Exercise 2.1.5.

(a) Finish the computation started in class (1.1.6) to verify that the described procedure ("chord and tangent process") yields Bachet's formula 1.1.8.

(b) In the case $c = 1$, why does Bachet's formula get "stuck" at the points $(0, \pm 1)$?

Exercise 2.1.6.

(a) Find all rational solutions (x, y) to $3x^2 - 2y^2 = 1$.

(b) $(**)$ Find all integer solutions (x, y) to $3x^2 - 2y^2 = 1$.

Exercise 2.1.7. (For those who know about the fields $\mathbb{Q}_p$ and Hensel's Lemma.)

(a) Find a prime at which the curve $x^2 + y^2 = 3$ is locally obstructed.

(b) $(*)$ Find *all* primes at which the curve $x^2 + y^2 = 3$ is locally obstructed.

Exercise 2.1.8. (Adapted from [2, Exercise 2.4].)

(a) Consider the plane curve $y^2 = x^3 + 1$. For each prime p, let n_p be the number of solutions (x, y) to this equation modulo p (i.e., the number of $\mathbb{F}_p$-points on this affine curve). Using your favorite computer software, compute the values n_p for all primes $p \leq 100$. Do you see any patterns? Make and prove conjectures. (Hint: Consider the cases $p \equiv \pm 1 \pmod 3$ separately.)

(b) Repeat the same procedure for the curve $y^2 = x^3 + x$. (Hint: Consider the cases $p \equiv \pm 1 \pmod 4$ separately.)

Exercise 2.1.9 (Fisher)**.** Let E be the elliptic curve over $\mathbb{Q}$ defined by the equation $y^2 + y = x^3 - x$.

(a) Draw a graph of the real points of E.

[1] This is the same as having the ability to find *one* factorization of any element of R, along with the knowledge of all units of R.

(b) Let $P := (0,0) \in E(\mathbb{Q})$. Get Sage to compute nP for $n = 0, 1, \ldots, 20$. What do you notice about the denominators? Make and prove conjectures. (Hint: Can you express the operation of adding P formulaically?)

2.1.B PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.1.10 (Hasse-Minkowski for Quadratic Equations)**.** Let $b, c \in \mathbb{Q}$. The quadratic equation $x^2 + bx + c = 0$ has a root in $\mathbb{Q}$ iff it has a root in $\mathbb{Q}_p$ for all primes $p \leq \infty$. (For those not too familiar with $\mathbb{Q}_p$ yet, the only fact you will need is: if a $d \in \mathbb{Q}^\times$ is a square in $\mathbb{Q}_p$, then $v_p(d)$ is even. You may assume this for now.)

Exercise 2.1.11 (Homogeneous Polynomials)**.** Let K be a field, $d \in \mathbb{Z}_{\geq 0}$, and $F \in K[X, Y, Z]$. The following conditions are equivalent:

(a) Every monomial in F has total degree d.
(b) We have the equality $F(WX, WY, WZ) = W^d F(X, Y, Z)$ in the polynomial ring $K[W, X, Y, Z]$.
(c) For all $\lambda \in K$, we have $F(\lambda X, \lambda Y, \lambda Z) = \lambda^d F(X, Y, Z)$.
(d) We have that
$$X\frac{\partial F}{\partial X} + Y\frac{\partial F}{\partial Y} + Z\frac{\partial F}{\partial Z} = d \cdot F.$$

Exercise 2.1.12. (Adapted from [2, Exercise 1.15].) Let K be a field, $g(u) \in K[u]$ be a monic quartic polynomial, and consider the curve
$$v^2 = g(u).$$

(a) Suppose we are given a root $\alpha \in K$ of g, and a $\beta \in K^\times$. The transformation
$$(x, y) := \left(\frac{\beta}{u - \alpha}, \frac{\beta^2 v}{(u - \alpha)^2}\right)$$
transforms $v^2 = g(u)$ to a curve of the form
$$y^2 = f(x)$$
for some cubic polynomial $f(x) \in K[x]$. This transformation is birational (i.e., a one-to-one correspondence between points of the two curves, except possibly for a finite number of points). Note that this birational transformation does not arise from a linear change of coordinates on (projective closures of) these curves.

(b) If $X, Y, Z \in K$ satisfy $X^4 + Y^4 = Z^4$ with $Z \neq X$, then
$$(x, y) := \left(2\left(\frac{Z + X}{Z - X}\right), \frac{4Y^2}{(Z - X)^2}\right)$$
lies on the curve
$$y^2 = x^3 + bx$$
for some $b \in K$. This b is unique. (What is it?)

(Hint: You will need some condition on g for (a) to work. What is this condition? What implications does this condition have for the roots of f? What do parts (a) and (b) have to do with each other? Part (b) is definitely not correct as stated.)

Exercise 2.1.13. Let K be a field.

(a) Any two lines in $\mathbb{P}^2(K)$ meet in exactly one point. Any two points in $\mathbb{P}^2(K)$ simultaneously lie on exactly one line.
(b) Given any two ordered quadruples (P_1, P_2, P_3, P_4) and (Q_1, Q_2, Q_3, Q_4) of points of $\mathbb{P}^2(K)$, there is a unique linear change of coordinates φ such that $\varphi(P_i) = Q_i$ for $i = 1, 2, 3, 4$. The same result holds for ordered pentuples.

Exercise 2.1.14 (Fisher)**.** Let $C \subset \mathbb{P}^2_{\bar{\mathbb{Q}}}$ be a smooth cubic curve. If $K/\mathbb{Q}$ is a quadratic or cubic number field, then $C(K) \neq \emptyset$ implies that $C(\mathbb{Q}) \neq \emptyset$.

2.2 Exercise Sheet 2

Make sure to finish all the exercises from Sheet 1 that you did not get to in the first week (e.g., 2.1.9) too!

2.2.A Numerical and Exploration

Exercise 2.2.1. For any $n \in \mathbb{Z}_{\geq 0}$ and prime p, how many points does $\mathbb{P}^n(\mathbb{F}_p)$ have? Draw a picture of $\mathbb{P}^1(\mathbb{F}_3)$. Draw a picture of $\mathbb{P}^1(\mathbb{F}_p)$ for any $p \geq 2$. Draw a picture of $\mathbb{P}^2(\mathbb{F}_2)$, including all lines on it, and capturing all the incidence relations between points and lines. (Does this picture remind you of something magical?)

Exercise 2.2.2. When does a (projective linear) change of coordinates fix the line at infinity ($Z = 0$) as a set? When does it fix it pointwise? When does it fix the line $Z = 0$ as a set, and fix the point $[0 : 1 : 0]$ on it?

Exercise 2.2.3. Let $K = \mathbb{R}$, let $b, c \in \mathbb{R}$ with $\Delta = -16(4b^3 + 27c^2) \neq 0$, and let E be the elliptic curve with Weierstrass equation $y^2 = x^3 + bx + c$.

(a) When does $E(\mathbb{R})$ consist of one connected component? When does it consist of two connected components?[2] (Hint: What does this have to do with Exercise 2.1.1?)
(b) Consider the curve with $(b, c) = (-21, -20)$. How many connected components does $E(\mathbb{R})$ have? Take the point $P := (-3, 4) \in E(\mathbb{R})$. Get Sage to compute (and maybe even plot!) the points nP for $n = 0, 1, \ldots, 10$. Which component does each of these points lie on? Observe patterns, and make and prove conjectures.
(c) Repeat the same procedure as in (b) for more elliptic curves over $\mathbb{R}$ with two components.

Exercise 2.2.4. Let K be a field with $\operatorname{ch} K \neq 3$, and let $\alpha \in K$, and let

$$E_\alpha = \mathbb{V}(X^3 + Y^3 - \alpha Z^3) \subset \mathbb{P}^2_K.$$

(a) Show that if $\alpha \neq 0$, then E_α is smooth. (What happens when $\operatorname{ch} K = 3$?)

Hence suppose that $\alpha \neq 0$.

(b) Given a point $P_0 = (x_0, y_0) \in E_\alpha(K)$, find a formula for the point $P_1 = (x_1, y_1)$ obtained by the same "chord and tangent process" as in the Bachet example (1.1.6): draw the tangent line, and then take the third point of intersection with E_α. This procedure gives a way to produce an infinite sequence of points $P_n \in E_\alpha(K)$ as before. (Hint: Your final formula should be quite elegant.)
(c) A point $P_0 \in E_\alpha(K)$ is said to be *exceptional* if the set $\{P_0, P_1, \ldots\}$ is finite; in other words, when the "chord and tangent process" only gets you finitely many points. Find at least one exceptional point on E_α for any α. When $K = \mathbb{Q}$ and $\alpha = 1, 2$, find some more exceptional points.
(d) $(*)$ Suppose now that $K = \mathbb{Q}$ and that $\alpha \geq 1$ is a cube-free integer. Use your answer to (b) to show that the only exceptional points on E_α are the ones you found in (c).
(e) Using the discussion from 1.2.B or otherwise, compute the j-invariant $j(E_\alpha)$ as a function of α.
(f) Finally, suppose now that $K = \mathbb{F}_p$ for primes $p \neq 3$. Using Sage, compute $\#E_1(\mathbb{F}_p)$ for all primes p with $5 \leq p \leq 1000$. What patterns do you observe? Make and prove conjectures. (Hint: Consider the cases $p \equiv 1, 2 \pmod 3$ separately.)

2.2.B PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.2.5. For any $b \in \mathbb{Z}$, the polynomial $x^3 + bx + 1 \in \mathbb{Q}[x]$ is irreducible, i.e., does not factor into two polynomials of smaller degree. The same is true for $x^3 + bx - 1$. (Hint: What does this have to do with this course? A result we talked about very early on may help.)

Exercise 2.2.6. Let K be a field, $F \in K[X, Y, Z]$ a homogeneous polynomial, $D = \mathbb{V}(F) \subset \mathbb{P}^2_K$. Let L/K be a field extension and $P = (x_0, y_0) = [x_0 : y_0 : 1] \in D(L)$.

[2] What's a connected component?

(a) Let $f(x,y) := F(x,y,1) \in K[x,y]$ denote the dehomogenization of F with respect to Z, and $C = D^\circ = \mathbb{V}(f) \subset \mathbb{A}^2_K$ the corresponding affine curve. Then $(x_0, y_0) \in C(L)$ is a smooth point of the affine curve C iff $[x_0 : y_0 : 1] \in D(L)$ is a smooth point of the projective curve D. Further, the projective tangent line $\mathbb{T}_P(D)$ is the projective closure of the affine tangent line $\mathrm{T}_P(C)$, i.e., $\mathbb{T}_P(D) = \overline{\mathrm{T}_P(C)}$.
(b) Suppose that $\varphi : \mathbb{P}^2_K \to \mathbb{P}^2_K$ is a projective linear change of coordinates. Then P is a smooth point of D iff $\varphi(P)$ is a smooth point of $\varphi(D)$, in which case we have $\varphi(\mathbb{T}_P D) = \mathbb{T}_{\varphi(P)}\varphi(D)$. (What is $\varphi(D)$?)

Exercise 2.2.7.

(a) Let K be a field and $D \subset \mathbb{P}^2_K$ a curve. Suppose that $P \in D(K)$ is a singular point. Then for any line L passing through P, we have that $m_P(D, L) \geq 2$, i.e., L intersects D in multiplicity at least 2 at P.
(b) (**) (This might require a little Galois theory, so if you haven't seen that, you can skip this part.) Suppose $D \subset \mathbb{P}^2_K$ is an integral cubic curve. Suppose that L/K is a field extension, and that $P \in D(L)$ is a singular point. Then in fact $P \in D(K)$. (Hint: See 1.5.5, which might suggest you look only at perfect K. Next, consider the cubic $\mathbb{V}(X^3+XY^2+Y^3+X^2Z+XYZ+Y^2Z+Z^3) \subset \mathbb{P}^2_{\mathbb{F}_2}$ or the cubic $\mathbb{V}(X^3-3XY^2-Y^3-3X^2Z-3XYZ-3Y^2Z+3Z^3) \subset \mathbb{P}^2_{\mathbb{Q}}$. Finally, for a very strong salvage, consider the case of *geometrically* integral D (e.g., a D in Weierstrass form) over a perfect field K. Here "geometrically integral" means that D remains integral over an algebraic closure $\overline{K}$ of K. If it makes things easier for you, you may suppose that L/K is finite.)

Exercise 2.2.8. (Adapted from [2, Exercise 2.13].) Let K be a field of characteristic other than 2, let $t \in K$, and consider the cubic $E_t \subset \mathbb{P}^2_K$ defined as the projective closure of the affine curve

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

(a) The curve E_t is nonsingular iff $t \notin \{0, 1/4\}$, in which case E_t is an elliptic curve over K in Weierstrass form. What is $j(E_t)$ as a function of t?
(b) In the situation in (a), the point $(t,t) \in E(K)$ has order 4.
(c) If $E \subset \mathbb{P}^2_K$ is any elliptic curve over a field K of characteristic other than 2 such that there is a point $P \in E(K)$ of order 4, then there is a projective linear change of coordinates $\varphi : \mathbb{P}^2_K \to \mathbb{P}^2_K$ and a $t \in K \smallsetminus \{0, 1/4\}$ such that $\varphi(E) = E_t$ and $\varphi(P) = (t,t)$.
(d) For a given (E, P) as in (c), the t found in (c) is unique.

Exercise 2.2.9. Let $R := \mathbb{Z}[a,b,c]$ be the polynomial ring in the variables a, b, c. Take the polynomial $f = f(x) = x^3 + ax^2 + bx + c \in R[x]$, and let $f'(x) = 3x^2 + 2ax + b$ and $f''(x) = 6x + 2a$ be the first and second formal derivatives of f with respect to x. The equation $y^2 = f(x)$ defines an elliptic curve E in Weierstrass form over $K = \mathbb{Q}(a,b,c)$, or over any field K of characteristic other than 2 when given specific $a, b, c \in K$ such that $\Delta \neq 0$ in K, which happens iff f and f' have no common roots over an algebraic closure.

(a) Let $f_3(x) := 2f(x)f''(x) - f'(x)^2$. Compute $f_3(x)$ explicitly as an element of $R[x]$. What is the degree of $f_3(x)$? What is its leading coefficient?
(b) PODASIP: For any field K and $P = (x,y) \in E(K) \smallsetminus \{\mathscr{O}\}$, the point P has order 3 (i.e., $3P = \mathscr{O}$, i.e., $P \in E[3](K)$) iff $f_3(x) = 0$. (Possible hints: $3P = \mathscr{O}$ is equivalent to $2P = -P$. Alternatively, think about inflection points. Can you check that

$$\frac{\mathrm{d}^2 y}{\mathrm{d}x^2} = \frac{f_3(x)}{4yf(x)}?$$

What does the second derivative have to do with inflection points?)
(c) PODASIP: For any field K, given $a, b, c \in K$ with $\Delta \neq 0$, the polynomial $f_3(x) \in K[x]$ never has repeated roots. (Hint: What is $f_3'(x)$? How many roots can this share with $f_3(x)$?)
(d) How many roots can it have in K when (i) $\operatorname{ch} K = 3$, (ii) when $K = \mathbb{R}$, (iii), when K is algebraically closed of characteristic other than 3? What implications does this have for the group $E[3](K)$ for various K as above? Specifically, for $K = \mathbb{R}$:
(e) Suppose $K = \mathbb{R}$, and $a, b, c \in \mathbb{R}$ are such that $\Delta \neq 0$. Then $f_3(x) \in \mathbb{R}[x]$ has exactly two real roots, say $\alpha < \beta$. We have $f(\alpha) < 0$ and $f(\beta) > 0$. In particular, $E[3](\mathbb{R})$ is cyclic of order 3.

Here's a real challenge.

Exercise 2.2.10 (Division Polynomials). (∗∗) (Adapted from [5, Exercise 2.6.10].) Let's generalize 2.2.9. For simplicity, we'll do it assuming $a = 0$. Let $R := \mathbb{Z}[b, c]$ be the polynomial ring in two variables b, c. Take the polynomial $f = f(x) = x^3 + bx + c \in R[x]$, and let $f'(x) = 3x^2 + b$ and $f''(x) = 6x$ be the first and second formal derivatives of f with respect to x.

(a) Define the sequence $(f_n)_{n\geq 0}$ of polynomials in $R[x]$ recursively by $f_0 = 0, f_1 = f_2 = 1$,

$$\begin{aligned} f_3 &:= 2f \cdot f'' - (f')^2, \\ f_4 &:= -16f^2 + 4f \cdot f' \cdot f'' - 2(f')^3, \\ f_{2n+1} &:= f_{n+2} \cdot f_n^3 - 16f^2 \cdot f_{n-1} \cdot f_{n+1}^3 \quad \text{for } n \geq 2 \text{ odd}, \\ f_{2n+1} &:= 16f^2 \cdot f_{n+2} \cdot f_n^3 - f_{n-1} \cdot f_{n+1}^3 \quad \text{for } n \geq 2 \text{ even, and} \\ f_{2n} &:= f_n(f_{n+2} \cdot f_{n-1}^2 - f_{n-2} \cdot f_{n+1}^2) \quad \text{for } n \geq 3. \end{aligned}$$

For $n \geq 1$, we have

$$f_n = \begin{cases} nx^{(n^2-1)/2} + \cdots, & \text{for } n \text{ odd, and} \\ (n/2)x^{(n^2-4)/2} + \cdots, & \text{for } n \text{ even,} \end{cases}$$

where $\cdots$ denotes terms of lower degree. Get Sage to write down the polynomials f_n for $n \leq 5$ explicitly.

(b) The equation $y^2 = f(x)$ defines an elliptic curve E in Weierstrass form (over $K = \mathbb{Q}(b, c)$ or over any field K of characteristic other than 2 when given specific $b, c \in K$ such that $4b^3+27c^2 \neq 0 \in K$). In this case,

$$\gcd(f_n, f \cdot f_{n+1} \cdot f_{n-1}) = (1)$$

when n is odd and

$$\gcd(f \cdot f_n, f_{n+1} \cdot f_{n-1}) = (1)$$

when $n \geq 2$ is even.

(c) If $P = (x, y) \in E(K)$, then the coordinates of $nP \in E(K)$ are given as

$$nP = \left(x - \frac{4 \cdot f \cdot f_{n+1} \cdot f_{n-1}}{f_n^2}, y \cdot \frac{f_{2n}}{f_n^4}\right)$$

when n is odd and

$$nP = \left(x - \frac{f_{n+1} \cdot f_{n-1}}{4f \cdot f_n^2}, y \cdot \frac{f_{2n}}{16f^2 \cdot f_n^4}\right)$$

when n is even. (Hint: WOP, of course.)

(d) Now fix an $n \geq 1$, and suppose that K is a field with $\operatorname{ch} K \nmid 6n$.

(1) For $P = (x, y) \in E(K)$, we have $nP = \mathscr{O}$ iff the x-coordinate $x(P)$ of P satisfies $f_n(x) = 0$ when n is odd or satisfies $f(x) \cdot f_n(x) = 0$ when n is even.

(2) When n is odd, the polynomial f_n is separable (i.e., has nonzero discriminant, i.e., has distinct roots over an algebraic closure of K), and when n is even, the polynomial $f \cdot f_n$ is separable.

(3) There are exactly n^2 points of order dividing n in $E(K)$, and, in fact, we have

$$E[n](K) \cong \mathbb{Z}/n \oplus \mathbb{Z}/n.$$

(Hint: If A is an abelian group of order n^2 for some $n \geq 1$ such that for each divisor $d \mid n$ we have $\#A[d] = d^2$, where $A[d] \subset A$ is the subgroup of all points of order dividing d, then $A \cong \mathbb{Z}/n \oplus \mathbb{Z}/n$. This follows from the classification of finite abelian groups, which we will do at the end of the course. You may either assume this for now, or try to prove it yourself!)[3]

[3] This result is also true if we only assume $\operatorname{ch} K \nmid n$, but the proof is harder. We might do it in class if time permits.

2.3 Exercise Sheet 3

2.3.A Numerical and Exploration

Exercise 2.3.1. For each prime $p \neq 2, 3$, and $b, c = 0, \ldots, p-1$ such that $4b^3 + 27c^2 \not\equiv 0 \pmod{p}$, let $E_{p,b,c}$ be the elliptic curve $y^2 = x^3 + bx + c$ over $\mathbb{F}_p$.

(a) Using, say, a simple `for` loop and the `E.abelian_group()` function in Sage, study the abelian group structure on $E_{p,b,c}$ for each $p \leq 19$ and each allowable b, c. Find patterns and make conjectures. For instance, what can you say about the ratio $\#E_{p,b,c}(\mathbb{F}_p)/p$ in general? Can you find (p, b, c) such that $E_{p,b,c}(\mathbb{F}_p)$ is cyclic of order 11? 19? When is this group cyclic? When it is not, does it ever need more than 2 generators?

(b) Modify your `for` loop in (a) to remove all results in which the group $E_{p,b,c}(\mathbb{F}_p)$ is cyclic, so that we can focus on the non-cyclic case. (In Sage, you can test whether a group is cyclic using the `is_cyclic()` function.) Allowing $p \leq 100$ now, find patterns and make conjectures. Does this group ever need more than 2 generators? Can you find a (p, b, c) such that the group $E_{p,b,c}(\mathbb{F}_p)$ is isomorphic to $\mathbb{Z}/3 \oplus \mathbb{Z}/3$? $\mathbb{Z}/m \oplus \mathbb{Z}/m$ for $m = 4, 5, \ldots, 9, 10$? $\mathbb{Z}/33 \oplus \mathbb{Z}/3$? $\mathbb{Z}/18 \oplus \mathbb{Z}/6$? When the group $E_{p,b,c}(\mathbb{F}_p)$ is $\mathbb{Z}/m \oplus \mathbb{Z}/m$, what do you notice about $p \mod m$?

(c) Can you use your calculations in (a) and (b) to give a lower bound on the number N_p of isomorphism classes of elliptic curves over $\mathbb{F}_p$ for each p with say $5 \leq p \leq 100$?

Exercise 2.3.2. Given a prime $p \neq 2, 3$, write Sage code to compute the number N_p of isomorphism classes of elliptic curves over $\mathbb{F}_p$ (up to isomorphism over $\mathbb{F}_p$). For $p \leq 200$, compute N_p/p. What patterns do you notice? (Possible hints: You might try something similar to what you did in 2.3.1. In Sage, you can test whether a curve `E1` defined over a field say `GF(p)` is isomorphic to another curve `E2` defined over the same field by using the boolean function `E1.is_isomorphic(E2)`.)

Exercise 2.3.3.

(a) Classify all elliptic curves over $\mathbb{F}_2$ (up to isomorphism over $\mathbb{F}_2$); in particular, compute N_2. (Hint: Note that each of the quantities $a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}_2$ could give rise to a curve, which means a priori that there are $2^5 = 32$ possibilities, but it is easy to eliminate a lot of these possibilities by inspection. For instance, it is clear from the formula for Δ in characteristic two (1.9.13) that not both a_1 and a_3 can be zero. You may use Sage to speed up calculations if that is helpful.)

(b) For each (isomorphism class of a) curve E as in (a),
  (i) identify the group $E(\mathbb{F}_2)$,
  (ii) find out if the curve E is supersingular,
  (iii) figure out the automorphism groups $\mathrm{Aut}_{\mathbb{F}_2}(E)$ and $\mathrm{Aut}_{\overline{\mathbb{F}}_2}(E)$, and
  (iv) calculate the j-invariant $j(E)$.

(c) $(*)$ For each pair of curves that are not isomorphic over $\mathbb{F}_2$ but have the same j-invariant, figure out the smallest field extension $L/\mathbb{F}_2$ over which those two curves *are* isomorphic. (Hint: This L is one of $\mathbb{F}_4, \mathbb{F}_{16}, \mathbb{F}_{256}$.)

(d) $(*)$ Repeat the same for $\mathbb{F}_3, \mathbb{F}_4$, $\mathbb{F}_5$, and $\mathbb{F}_7$. (Hint: Again, you may use Sage. This problem is not as difficult as it may look, since over $\mathbb{F}_3$ you can put elliptic curves in the short Weierstrass form $y^2 = x^3 + ax^2 + bx + c$, and over $\mathbb{F}_5$ and $\mathbb{F}_7$ you can put them in the even shorter Weierstrass form $y^2 = x^3 + bx + c$, as above. Recall that $\mathbb{F}_4 \cong \mathbb{F}_2[\omega] \cong \mathbb{F}_2[x]/(x^2 + x + 1)$. The way to work with $\mathbb{F}_4$ in Sage is to use `K.<a> = GF(4)`; this returns a field K isomorphic to $\mathbb{F}_4$ with a generator named a satisfying $a^2 + a + 1 = 0$.)

If you haven't seen the fields $\mathbb{F}_{p^n}$ for $n \geq 2$ (note that $\mathbb{F}_{p^n} \ncong \mathbb{Z}/(p^n)$ as rings!), you can skip those parts of this exercise for now and return to them later when you do.

Exercise 2.3.4.

(a) Find all rational solutions to $x^2 + y^2 = 2$ by projecting from $(-1, -1)$ and using t as the slope of the line, i.e., by mapping t to the other intersection of $y + 1 = t(x + 1)$ with $x^2 + y^2 = 2$.

(b) Using your solution to (a) or otherwise, prove that if p, q, r are positive integers with $p < q < r$ such that p^2, q^2 and r^2 are in arithmetic progression, then there are coprime $m, n \in \mathbb{Z}$ of different parity with $m > n > 0$ such that (p, q, r) is one of

$$\left(\pm(2mn - (m^2 - n^2)), m^2 + n^2, 2mn + m^2 - n^2\right).$$

Figure out exactly when each sign holds. (Hint: Your answer may have something to do with $\sqrt{2} \pm 1$.)

Exercise 2.3.5. (Adapted from [29].) For simplicitly, let's work over $K = \mathbb{Q}$. Let $C \subset \mathbb{P}^2_K$ be the cubic curve with equation

$$XY^2 - X^2Y + 2X^2Z - 2XZ^2 - 3Y^2Z + 3YZ^2 = 0.$$

(a) $(*)$ Suppose $p, q, r, s \in \mathbb{Z}_{>0}$ are such that p^2, q^2, r^2, s^2 are in arithmetic progression. Suppose that $p \neq s$, so the sequence is not the constant sequence. Show that

$$[p - s : q - s : r - s] \in C(K).$$

Therefore, to get a handle on all four-term sequences of squares in arithmetic progressions, it suffices to get a handle on $C(K)$. (Hint: There are many ways to do this. One is to just plug and check. Another, possibly more illuminating, way might be to express the given condition in terms of two quadratic conditions on p, q, r, s; indeed, what *does* it mean to be in arithmetic progression? Next, let $(X, Y, Z) = (p - s, q - s, r - s)$, and write these quadratic conditions in terms of X, Y, Z, and s. Finally, try to eliminate s from the equations.)[4]

(b) Is C smooth? Find a rational inflection point on C. (Hint: Get Desmos to draw a picture of the affine patch $Z \neq 0$ of C. Then try an intersection with a coordinate axis.)

(c) Using your solution to (b) or otherwise, find the linear change of coordinates that transforms C into the elliptic curve $E \subset \mathbb{P}^2_K$ in short Weierstrass form

$$y^2 = x(x + 1)(x + 4).$$

We met this curve in 1.8.4.

(d) Using the Nagell-Lutz Theorem, find all torsion points in $E(K)$. How many of them are there? Describe the structure of the torsion subgroup $\operatorname{Tors} E(K)$. (Hint: Up to isomorphism, there are three groups of size $\#\operatorname{Tors} E(K)$. Part of your task is to figure out which of these is the isomorphism type of $\operatorname{Tors} E(K)$. What distinguishes these three types?)

(e) Later in the course, we will prove that $\operatorname{rank} E = 0$. Assuming this result, prove the theorem stated by Fermat in 1640 and first proven by Euler in 1780 which states that there are no nonconstant four-term sequence of squares in arithmetic progression in the integers.

(f) How much of the above discussion carries over when we replace K by a number field or a finite field? For which characteristics does the above discussion still apply, and for which does it not? When it does apply, what do we learn about four squares in arithmetic progressions over these fields?

2.3.B PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.3.6. Let K be a field.

(a) Let $C \subset \mathbb{A}^2_K$ be an affine curve. If C is integral, then so its projective closure $\overline{C} \subset \mathbb{P}^2_K$.

(b) Let $D \subset \mathbb{P}^2_K$ be a projective curve. If D is integral, then so is each affine patch $D^\circ \subset \mathbb{A}^2_K$.

The same statements are true with "integral" replaced by "geometrically integral".

Exercise 2.3.7. Let K be a field.

(a) Let $C \subset \mathbb{A}^2_K$ be an integral affine curve with defining equation $f \in K[x, y]$. The element $f \in K[x, y]$ is prime.

We define the *function field* of C to be $K(C) := \operatorname{Frac} K[x, y]/(f)$. (Make sure you see why this is independent of the choice of f!)

(b) If $C \subset \mathbb{A}^2_K$ is an integral curve with projective closure $\overline{C}$, then $K(C) \cong K(\overline{C})$.

(c) If $D \subset \mathbb{P}^2_K$ is an integral curve with affine patch say $D^\circ \subset \mathbb{A}^2_K$, then $K(D) \cong K(D^\circ)$.

[4] For the cognoscenti: The curve $\mathbb{V}(p^2 + r^2 - 2q^2, q^2 + s^2 - 2r^2) \subset \mathbb{P}^3_K$ is a smooth projective geometrically integral curve of genus one, and hence an elliptic curve when equipped with the rational point $[1 : -1 : -1 : 1]$. We are trying to find a Weierstrass equation for it, which can be done by projecting from the point $[1 : 1 : 1 : 1]$ to embed the curve as a cubic $C \subset \mathbb{P}^2_K$, and then applying a linear change of coordinates.

Exercise 2.3.8. Let K be a field.

(a) Let $C_1, C_2 \subset \mathbb{A}^2_K$ be affine curves, say $C_i = \mathbb{V}(f_i)$ for nonzero $f_1, f_2 \in K[x,y]$. Let $C = \mathbb{V}(f_1 \cdot f_2)$, so that for all fields L/K we have $C(L) = C_1(L) \cup C_2(L)$. (Note that C is not the "union" of C_1 and C_2 in the naive sense; for instance we could have $f_1 = f_2 = x$, in which case C would be $C = \mathbb{V}(x^2)$.) If L/K is a field extension, and $P \in C_1(L) \cap C_2(L)$, then $P \in C(L)$ is a singular point. (This might help explain why curves like $\mathbb{V}(x^2)$ are everywhere singular.) A similar result is true for projective curves.

(b) If $C \subset \mathbb{A}^2_K$ is a smooth affine curve, then C is (geometrically) integral. The same is true for projective curves. (Hint: You may use the full strength of Bézout's Theorem (1.6.9), which we have admittedly not proven.)

Exercise 2.3.9. Let K be a field, $C \subset \mathbb{P}^2_K$ a smooth cubic curve, and $\mathscr{O}_1, \mathscr{O}_2 \in C(K)$. For $i = 1, 2$, let $\oplus_i$ denote the operation $P \oplus_i Q := \mathscr{O}_i * (P * Q)$ for $P, Q \in C(K)$, so that $(C(K), \oplus_i)$ is an abelian group by 1.7.7. The map

$$(C(K), \oplus_1) \to (C(K), \oplus_2), \quad P \mapsto \mathscr{O}_1 * (\mathscr{O}_2 * P)$$

is a group isomorphism. In particular, the group law on $C(K)$ does not depend on the choice of $\mathscr{O}$, up to isomorphism. (Hint: There are many ways to do this. In one possible direction, it might be helpful to show that the associativity of the $\oplus$ operation corresponding to $\mathscr{O}$ is equivalent to the statement that for all $P, Q, R \in C(K)$ we have $R * (\mathscr{O} * (P * Q)) = P * (\mathscr{O} * (Q * R))$.)

Exercise 2.3.10. (**)

(a) Let $b \in \mathbb{Z}$ be fourth-power-free, and let $E_b/\mathbb{Q}$ be the elliptic curve with Weierstrass equation

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

Then $\operatorname{Tors} E_b(\mathbb{Q})$ is either $\mathbb{Z}/2$ or $\mathbb{Z}/4$. When does each case hold?

(b) Let $c \in \mathbb{Z}$ be sixth-power-free, and let $E_c/\mathbb{Q}$ be the elliptic curve with Weierstrass equation

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

Then $\operatorname{Tors} E_c(\mathbb{Q})$ is either $\mathbb{Z}/2$ or $\mathbb{Z}/3$. When does each case hold? Compare your result to your answer to 2.2.4(e).

(Hint: First, try lots of examples, using Sage. Then, if needed, refer to [1, §12] for a detailed hint.)

2.4 Exercise Sheet 4

2.4.A Numerical and Exploration

Exercise 2.4.1. Here we list some pairs (E, P), where $E/\mathbb{Q}$ is an elliptic curve and $P \in E(\mathbb{Q})$. Using Sage, find the order of P in $E(\mathbb{Q})$.

(a) $E : y^2 = x^3 - 4x + 4$ and $P = (2, 2)$.
(b) $E : y^2 = x^3 - x + 1$ and $P = (1, 1)$.
(c) $E : y^2 = x^3 - 219x + 1654$ and $P = (11, 24)$.
(d) $E : y^2 + xy = x^3 - x + 137$ and $P = (2, 11)$.
(e) $E : y^2 = x^3 - 189x + 1269$ and $P = (3, 27)$. Also do $P = (21, 81)$ on the same curve.

Find other interesting examples on the LMFDB.

Exercise 2.4.2. (Adapted from [2, Exercise 2.12]) Using the Nagell-Lutz Theorem or otherwise, for each elliptic curve $E/\mathbb{Q}$ below, find all rational points of finite order on E and determine the structure of $\operatorname{Tors} E(\mathbb{Q})$:

(a) $y^2 = x^3 - 2$
(b) $y^2 + xy + y = x^3$.
(c) $y^2 = x^3 + 1$
(d) $y^2 = x^3 - 43x + 166$
(e) $y^2 + xy = x^3 - 45x + 81$

(Hint: You may have to complete the square to bring the given curve into shortest Weierstrass form to apply Nagell-Lutz. At least try to do some of these computations by hand (i.e., without using Sage) to convince yourself that they can be done in a reasonable amount of time!)

Exercise 2.4.3. Let p be a prime and let $E/\mathbb{Q}$ be the elliptic curve with equation $y^2 = x^3 + p^2$. What is $\operatorname{Tors} E(\mathbb{Q})$? Experiment in Sage, make conjectures, and then prove your conjectures.

Exercise 2.4.4.

(a) Fix a prime p. For each integer $m \in \mathbb{Z}_{\geq 0}$, let $S_m(p) := \sum_{x=0}^{p-1} x^m$. What can you say about the function $m \mapsto S_m(p) \pmod p$? Observe patterns, and make and prove conjectures. What does the value of 0^0 need to be in order to obtain a very clean and uniform result?
(b) (Only for those who know about finite fields other than $\mathbb{F}_p$ for primes p.) Fix a finite field $\mathbb{F}_q$. Repeat the same as in (a) for the function $m \mapsto \sum_{x \in \mathbb{F}_q} x^m$.

Exercise 2.4.5 (Fisher)**.** Let p be an odd prime, and let E be an elliptic curve over $\mathbb{F}_p$.

(a) Show that there exists and elliptic curve E' over $\mathbb{F}_p$ such that $\#E(\mathbb{F}_p) + \#E'(\mathbb{F}_p) = 2(p+1)$. (Hint: Modulo the right definitions of these objects, you could probably have solved this problem before you knew anything about elliptic curves.)
(b) Given the curves E and E' as in (a), what do you notice about the sizes of $E(\mathbb{F}_p) \times E'(\mathbb{F}_p)$ and of $E(\mathbb{F}_{p^2})$ and $E'(\mathbb{F}_{p^2})$? Observe patterns and make conjectures.
(c) (∗∗) Prove your conjectures from (b). (Hint: It might be helpful to consider the "conjugation" map $\mathbb{F}_{p^2} \to \mathbb{F}_{p^2}$ given by $a \mapsto a^2$. If you don't know much about the fields $\mathbb{F}_{p^2}$ in general, try to attempt this problem for small special cases such as $p = 3$ where $\mathbb{F}_9 \cong \mathbb{F}_3[i] := \mathbb{F}_3[x]/(x^2 + 1)$ or $p = 5$ where $\mathbb{F}_{25} \cong \mathbb{F}_5[\omega] := \mathbb{F}_5[x]/(x^2 + x + 1)$.)

Exercise 2.4.6. Let K be a field and $E \subset \mathbb{P}^2_K$ be an elliptic curve. In §1.12.E, we studied what the group $E[3](\overline{K})$ can look like, where $\overline{K}$ denotes an algebraic closure of K. Perform an analysis similar to the one in §1.12.C to figure out what happens when K is not necessarily algebraically closed. For instance:

(a) Give an example of a field K and an elliptic curve E/K such that $(\#E[3](\overline{K}), \#E[3](K))$ is (i) $(9, 1)$ (ii) $(9, 3)$, (iii) $(9, 9)$, (iv) $(3, 1)$, (v) $(3, 3)$, and (vi) $(1, 1)$.
(b) (∗∗) As done in §1.12.C for two-torsion, classify exactly when each of the possibilities in (a) occurs.
(c) (∗) Show that for any field K, there is an extension L/K of degree $[L : K] \leq 8$ such that there is a $P \in E[3](L) \smallsetminus \{\mathcal{O}\}$. Show that if $\operatorname{ch} K = 3$ and E is not supersingular, this can bound can be improved to $[L : K] \leq 3$. Give examples to show that both of these bounds are sharp. (Hint: Look at the proof of 1.12.23 carefully, and use that $8 = 4 \cdot 2$.)

2.4.B PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.4.7. Let K be a field, and for $\lambda \in K \smallsetminus \{0,1\}$, let

$$j(\lambda) := 2^8 \frac{(\lambda(1-\lambda)+1)^3}{\lambda^2(1-\lambda)^2}.$$

(a) For $\lambda, \lambda' \in K \smallsetminus \{0,1\}$, we have that $j(\lambda) = j(\lambda')$ iff $\lambda' \in S(\lambda)$, where

$$S(\lambda) := \left\{\lambda, \frac{1}{\lambda}, 1-\lambda, \frac{1}{1-\lambda}, 1-\frac{1}{\lambda}, 1-\frac{1}{1-\lambda}\right\}.$$

(b) For any $\lambda \in K \smallsetminus \{0,1\}$, the set $S(\lambda)$ of (a) has cardinality exactly six.

Exercise 2.4.8. Let K be a field.

(a) For any elliptic curve E over K, there is a(n) (admissible) change of coordinates which brings E into the form $E_\lambda : y^2 = x(x-1)(x-\lambda)$ for some $\lambda \in K \smallsetminus \{0,1\}$. The j-invariant of E_λ is precisely the $j(\lambda)$ of 2.4.7, i.e., $j(E_\lambda) = j(\lambda)$.

(b) For $\lambda, \lambda' \in K \smallsetminus \{0,1\}$, there is an admissible change of coordinates taking E_λ to $E_{\lambda'}$ iff $\lambda' \in S(\lambda)$ iff $j(\lambda) = j(\lambda')$. In particular, the j-invariant classifies elliptic curves up to admissible changes of coordinates (and hence by 1.10.6 up to isomorphism). (Hint: The "change in the x-coordinate alone" would have to send the unordered triple $\{0,1,\lambda\}$ to $\{0,1,\lambda'\}$.)

Exercise 2.4.9. Let K be a field (of any characteristic!). Given any $j \in K$, there is a unique elliptic curve E over K (up to admissible changes of coordinates over K) such that $j(E) = j$.

Exercise 2.4.10. Let K be a field. Given finitely many points $P_i \in \mathbb{P}^2(K)$, there is a line $L \subset \mathbb{P}^2_K$ that doesn't pass through any of the P_i. In particular, up to a linear change of coordinates, we can assume that each P_i lies in $\mathbb{P}^2_K \smallsetminus \mathbb{V}(Z) \cong \mathbb{A}^2_K$.

Exercise 2.4.11. Let K be a field and $E \subset \mathbb{P}^2_K$ an elliptic curve in Weierstrass form.

(a) If $P, Q \in E(K)$ are inflection points, then so is $P * Q$.

(b) If $S \subset E(K)$ is the set of inflection points, then no four points in S are collinear, but any line passing through two points in S passes through a third point in S.

Exercise 2.4.12.

(a) $(*)$ Suppose $S \subset \mathbb{A}^2(\mathbb{R})$ is a finite set of points with the property that a line passing through any two of them passes through a third. Then S is contained in a line. (Hint: This is just plane geometry; it has nothing to do with elliptic curves.)

(b) If $E \subset \mathbb{P}^2_{\mathbb{R}}$ is an elliptic curve, then exactly three of its nine complex inflection points are real, i.e., even though $\#E[3](\mathbb{C}) = 9$, the subgroup $E[3](\mathbb{R}) \subset E[3](\mathbb{C})$ has size $\#E[3](\mathbb{R}) = 3$. In particular, $E[3](\mathbb{R}) \cong_{\mathsf{Ab}} \mathbb{Z}/3$. (See also 2.2.9.)

Exercise 2.4.13 (Fisher)**.** Let $K = \mathbb{Q}$ and let E be the elliptic curve with Weierstrass equation $y^2 = x^3 + ax^2 + bx$ with $a, b \in \mathbb{Z}$.

(a) If $P = (x,y) \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{0\}$, then either $x = 0$ or $x \mid b$ and $x + a + (b/x)$ is a perfect square. (Hint: Review the proof of the Nagell-Lutz Theorem.)

(b) Repeat the same as in 2.4.3 for the curve $E : y^2 = x^3 + px$ for a prime p.

Exercise 2.4.14. Let $E/\mathbb{Q}$ be an elliptic curve given in long Weierstrass form with all coefficients $a_1, \ldots, a_6 \in \mathbb{Z}$.

(a) If $P = (x,y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then when expressed as fractions in lowest terms, x and y look like

$$(x,y) = \left(\frac{m}{e^2}, \frac{n}{e^3}\right)$$

for some $m, n, e \in \mathbb{Z}$ with $e \geq 1$ and $\gcd(m,e) = \gcd(n,e) = 1$.

(b) For each prime p and integer $r \in \mathbb{Z}_{\geq 1}$, if we define

$$E_{p,r}(\mathbb{Q}) := \{(x,y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\} : v_p(x) \leq -2r\} \cup \{\mathscr{O}\},$$

then $E_{p,r}(\mathbb{Q}) \subset E(\mathbb{Q})$ is a subgroup.

In other words, (a) and (b) are saying that the results shown in class are for the short Weierstrass form are also true for the long Weierstrass form, as long as the a_i are integers.

(c) (∗∗) If $P = (x, y) \in \operatorname{Tors} E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then $x, y \in \mathbb{Z}$. (Hint: Consider $E : y^2 + xy = x^3 + 4x + 1$. For a strong salvage, show that the only thing that can go wrong involves some very small powers of 2. Show also that if $2 \mid a_1$ or if $2P \neq \mathscr{O}$, then $x, y \in \mathbb{Z}$.)

For the next few exercises, we use the following notation: if p is a prime, and $E/\mathbb{F}_p$ an elliptic curve, we let $a_p(E) := p + 1 - \#E(\mathbb{F}_p) \in \mathbb{Z}$. the The integer $a_p(E)$ is called the *trace of Frobenius* associated to the elliptic curve $E/\mathbb{F}_p$. Note that $\#E(\mathbb{F}_p) = p + 1$ iff $a_p(E) = 0$.

Exercise 2.4.15. Let p be a prime other than 2, and let $E : y^2 = f(x)$ be an elliptic curve over $K = \mathbb{F}_p$ in short Weierstrass form.

(a) The integer $a_p(E)$ is congruent mod p to the coefficient of x^{p-1} in $f(x)^{(p-1)/2}$. (Hint: The Legendre symbol and 2.4.4.)

(b) Suppose that $f(x) = x(x-1)(x-\lambda)$ for $\lambda \in \mathbb{F}_p \smallsetminus \{0, 1\}$. The coefficient of x^{p-1} in $f(x)^{(p-1)/2}$ is $h_p(\lambda)$, where $h_p(x) \in \mathbb{F}_p[x]$ is the polynomial defined by

$$h_p(x) := \sum_{i=0}^{(p-1)/2} \binom{(p-1)/2}{i}^2 x^i \in \mathbb{F}_p[x].$$

We have that $h_p(0) \cdot h_p(1) \neq 0$.

Exercise 2.4.16. Fix an integer $b \in \mathbb{Z}$.

(a) For each prime p with $p \nmid 2b$, the equation $y^2 = x^3 + bx$ defines an elliptic curve E_b over $\mathbb{F}_p$. Further,

$$a_p(E_b) \equiv \binom{(p-1)/2}{(p-1)/4} b^{(p-1)/4} \pmod{p}.$$

(b) (∗∗) Further, if $p \geq 5$ and $p \equiv 3 \pmod 4$, then $a_p(E_b) = 0$.

Exercise 2.4.17. Fix an integer $c \in \mathbb{Z}$.

(a) For each prime p, the equation $y^2 = x^3 + c$ defines an elliptic curve E_c over $\mathbb{F}_p$. Further,

$$a_p(E_c) \equiv \binom{(p-1)/2}{(p-1)/3} c^{(p-1)/6} \pmod{p}.$$

(b) (∗∗) Further, if $p \geq 5$ and $p \equiv 2 \pmod 3$, then $a_p(E_c) = 0$.

How do 2.4.17 and 2.4.16 relate to 2.1.8? A possible hint for both 2.4.17(a) and 2.4.16(a): experiment in Sage! A possible hint for 2.4.17(b) and 2.4.16(b): combine your solution to (a) with [8, Theorem V.1.1]. Is there an easier way?

Here are a few more exercises, just for fun.

Exercise 2.4.18. Let K be a field. For $h(x) \in K[x]$ and $n \in \mathbb{Z}_{\geq 0}$, let $h^{(n)}(x)$ denote the n^{th} (formal) derivative of h with respect to x, i.e.,

$$h^{(n)}(x) = \frac{\mathrm{d}^n}{\mathrm{d}x^n} h(x) \in K[x].$$

Let $\overline{K}$ be an algebraic closure of K, and suppose there is an $\alpha \in \overline{K}$ such that $h^{(n)}(\alpha) = 0$ for all $n \geq 0$. Then h is the zero polynomial, i.e., $h(x) = 0 \in K[x]$. In other words, if $h(x) \in K[x]$ is not the zero polynomial, then for each $\alpha \in \overline{K}$, there is some $n \in \mathbb{Z}_{\geq 1}$ such that $h^{(n)}(\alpha) \neq 0$. (Hint: For a strong salvage, consider the case in which either $\operatorname{ch} K = 0$ or $\operatorname{ch} K = p > \deg h$. It may help to "translate" so that $\alpha = 0$.)

Exercise 2.4.19. Let K be a field. A *linear differential operator* $\mathscr{D}$ *of degree* $n \in \mathbb{Z}_{\geq 1}$ is an expression of the form

$$\mathscr{D} = c_0(x)\frac{\mathrm{d}^n}{\mathrm{d}x^n} + c_1(x)\frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} + \cdots + c_{n-1}(x)\frac{\mathrm{d}}{\mathrm{d}x} + c_n(x),$$

where $c_0(x), c_1(x), \ldots, c_n(x) \in K[x]$ and $c_0(x) \neq 0$.

(a) Given a differential operator $\mathscr{D}$ and a polynomial $h(x) \in K[x]$, describe what it means to "apply $\mathscr{D}$ to h" to get the polynomial $(\mathscr{D}h)(x) \in K[x]$. The resulting map $\mathscr{D} : K[x] \to K[x]$ is K-linear.
(b) Suppose $\mathscr{D}$ as above is a differential operator of order $n \in \mathbb{Z}_{\geq 1}$ and suppose $h(x) \in K[x]$ is a nonzero polynomial killed by $\mathscr{D}$, i.e., such that $(\mathscr{D}h)(x) = 0 \in K[x]$. If $\overline{K}$ is an algebraic closure of K and $\alpha \in \overline{K}$ a root of $h(x)$ of multiplicity at least n, then $c_0(\alpha) = 0$. (Hint: 2.4.18.)

Exercise 2.4.20. Let p be a prime other than 2, and let $h_p(x) \in \mathbb{F}_p[x]$ be the polynomial defined in 2.4.15.

(a) Let $\mathscr{D}$ be the *Picard-Fuchs differential operator*

$$\mathscr{D} := 4x(1-x)\frac{\mathrm{d}^2}{\mathrm{d}x^2} + 4(1-2x)\frac{\mathrm{d}}{\mathrm{d}x} - 1.$$

We have that $(\mathscr{D}h_p)(x) = 0 \in \mathbb{F}_p[x]$.
(b) The polynomial $h_p(x) \in \mathbb{F}_p[x]$ is separable, i.e., has no repeated roots in $\overline{\mathbb{F}}_p$. In particular, there are exactly $(p-1)/2$ elements $\lambda \in \overline{\mathbb{F}}_p \smallsetminus \{0, 1\}$ such that $h_p(\lambda) = 0$.

The following result might explain why this problem is worth considering and potentially motivate you to attempt it: the curve $E : y^2 = x(x-1)(x-\lambda)$ over $\overline{\mathbb{F}}_p$ is supersingular iff $h_p(\lambda) = 0$ ([8, Theorem V.4.1(b)]). This in turn allows us to count the number of supersingular elliptic curves over $\overline{\mathbb{F}}_p$ up to isomorphism.

2.5 Exercise Sheet 5

2.5.A Numerical and Exploration

Exercise 2.5.1. Let $E/\mathbb{Q}$ be the elliptic curve $E : y^2 = x^3 + x$.

(a) For each $n = 1, 2, 3, 4$, explicitly write down all the elements of, and determine the groups structures of, $E[n](\mathbb{Q})$ and $E[n](\overline{\mathbb{Q}})$. How many points of each order do we have in each of these groups? Does 1.12.28 hold?

(b) (For those who have seen some Galois theory.) For each $n = 1, 2, 3, 4$, let $K_n/\mathbb{Q}$ denote the number field generated by the x- and y-coordinates of all the points $P \in E[n](\overline{\mathbb{Q}})$. Determine the degree $[K_n : \mathbb{Q}]$. Is $K_n/\mathbb{Q}$ Galois? If so, what is $\mathrm{Gal}(K_n/\mathbb{Q})$? For which n does K_n contain $\mathbb{Q}[i]$? When it does, what can you say about $\mathrm{Gal}(K_n/\mathbb{Q}[i])$? When it doesn't, what can you say about $\mathrm{Gal}(K_n \cdot \mathbb{Q}[i]/\mathbb{Q}[i])$? (Hint: It might be helpful to get Sage to factor a sextic polynomial.)

(c) Repeat the same as in (a) and (b) for the curve $E : y^2 = x^3 + 1$. Note that you may wish to replace $\mathbb{Q}[i]$ in (b) by some other imaginary quadratic number field (which one?). (Hint: The cases $n = 1, 2, 3$ can be done easily by hand. You may use Sage for $n = 4$.)

(d) $(**)$ Make and prove conjectures generalizing your results from (a), (b), and (c). (Hint: Can you generalize these results to curves of the form $y^2 = x^3 + bx$? To curves of the form $y^2 = x^3 + c$? To curves with complex multiplication?)

Exercise 2.5.2.

(a) Does every finitely generated abelian group A admit a height function, i.e., a function $h : A \to [0, \infty)$ satisfying conditions (a), (b), (c) of 1.15.11?

(b) Is the map $\mathbb{Q} \to [0, \infty)$ given by $q \mapsto h(q)$ a height function on the additive group $\mathbb{Q}$? Is the same map a height function on the multiplicative group $\mathbb{Q}^\times$?

(c) Determine, for each of the following fields K,
   (i) if $K/2K$ is finite,
   (ii) if K admits a height function,
   (iii) if $K^\times/(K^\times)^2$ is finite, and
   (iv) if $K^\times$ admits a height function.
   Consider $K = \mathbb{Q}, \mathbb{F}_q, \mathbb{R}, \mathbb{C}, \mathbb{F}_2(t)$. If you know about the p-adic numbers, also consider $K = \mathbb{Q}_p$.

Exercise 2.5.3. Let $E/\mathbb{Q}$ be the elliptic curve $E : y^2 = x^3 - 7x + 10$. Consider the points $P = (1, 2)$ and $Q := (3, 4)$ on E.

(a) Use Sage to compute $mP + nQ$ for all $m, n \in \mathbb{Z}$ with $-10 \leq m, n \leq 10$, and check if $mP + nQ = \mathscr{O}$ for any of these values of m and n.

(b) Use Sage to compute the Néron-Tate canonical heights of P, Q, and $P+Q$. (Hint: You can get Sage to return the canonical height of a point `P` on an elliptic curve over $\mathbb{Q}$ by asking for `P.height()`.)

(c) Are P and Q $\mathbb{Z}$-linearly dependent in the group law on E? In other words, are there integers $m, n \in \mathbb{Z}$, not both zero, such that $mP + nQ = \mathscr{O}$? (Hint: Use (b) to make a suitable 2×2 matrix, and then compute its determinant. You can either trust Sage's floating-point arithmetic (highly recommended!), or very tediously verify it yourself to sufficient accuracy.)

(d) Repeat the same as in (a), (b), and (c) for a few other curves $E/\mathbb{Q}$ and pairs of points $P, Q \in E(\mathbb{Q})$ of your choosing (obtained either by just experimentation in Sage, or from the LMFDB). Formulate a general criterion in terms of the Néron-Tate height pairing to determine whether or not a set of points $P_1, \ldots, P_n \in E(\mathbb{Q})$ is $\mathbb{Z}$-linearly dependent.

Exercise 2.5.4. Using the method explained in class (descent via 2-isogeny), compute the ranks of the following elliptic curves $E/\mathbb{Q}$ by hand, and then check your answer with Sage. Further, in each case, determine a full set of coset representatives for $E(\mathbb{Q})/2E(\mathbb{Q})$.

(a) $y^2 = x^3 + 7x$
(b) $y^2 = x^3 - 12x^2 + 20x$
(c) $y^2 = x^3 - 3x^2 + 10x$
(d) $y^2 = x^3 - 377x$

(Hint: This is NOT a random list–the numbers are specifically chosen so that the computations are easily doable by hand! Remark: To get an idea of the sort of thing Sage is doing when you call `rank()`, you can either do `E.rank?` or `print(E.mwrank())` for a given curve `E`.)

2.5.B PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.5.5. (Adapted from [2, Exercise 3.2]) Let $E/\mathbb{Q}$ be an elliptic curve in short Weierstrass form $E : y^2 = x^3 + ax^2 + bx + c$ with $a, b, c \in \mathbb{Z}$. There is a constant $C_5 \in \mathbb{R}_{\geq 0}$ depending only on a, b, c such that for all $P, Q \in E(\mathbb{Q})$, we have

$$|h(P+Q) + h(P-Q) - 2(h(P) + h(Q))| \leq C_5.$$

(Hint: Express $x(P+Q)+x(P-Q)$ and $x(P+Q)\cdot x(P-Q)$ in terms of $x(P)$ and $x(Q)$ alone. Then try something similar to what we did in class to get one direction of the inequality. To get the other direction, use the direction already proven along with the inequality $h(2P) \geq 4h(P) - C_3$ from class.)

Exercise 2.5.6. Let E_1, E_2 be isomorphic elliptic curves over $\mathbb{Q}$ in Weierstrass form, and let $\varphi : E_1 \to E_2$ be an isomorphism, i.e., an admissible change of coordinates. Then for any point $P \in E_1(\mathbb{Q})$, we have that

$$\hat{h}_1(P) := \lim_{N\to\infty} \frac{1}{4^N} h(x(2^N P)) = \lim_{N\to\infty} \frac{1}{4^N} h(x(\varphi(2^N P))) =: \hat{h}_2(P).$$

In particular, the Néron-Tate canonical height on an elliptic curve E is independent of the choice of Weierstrass equation for E, and only depends on its isomorphism type over $\mathbb{Q}$.

Exercise 2.5.7. Let $E/\mathbb{Q}$ be an elliptic curve in Weierstrass form, and let $\hat{h} : E(\mathbb{Q}) \to [0, \infty)$ be the Néron-Tate canonical height on E. For any $P \in E(\mathbb{Q})$ and $T \in \operatorname{Tors} E(\mathbb{Q})$, we have that $\hat{h}(P+T) = \hat{h}(P)$. In other words, $\hat{h}$ is insensitive to the addition of torsion points on $E(\mathbb{Q})$.

Exercise 2.5.8. Let $E/\mathbb{Q}$ be an elliptic curve.

(a) For all $P \in E(\mathbb{Q})$, we have

$$\hat{h}(P) = \lim_{m\to\infty} \frac{1}{m^2} h(mP).$$

(b) For all $P \in E(\mathbb{Q})$, we have

$$\lim_{N\to\infty} \frac{1}{4^N} h(y(2^N P)) = \frac{3}{2} \cdot \hat{h}(P).$$

More generally, for any $f \in \mathbb{Q}(E)$, there is a constant $d \in \mathbb{Q}$ depending only on f such that for all $P \in E(\mathbb{Q})$,

$$\lim_{N\to\infty} \frac{1}{4^N} h(f(2^N P)) = \frac{d}{2} \cdot \hat{h}(P).$$

(c) $(*)$ Suppose $p \in [0, \infty]$. Define the ℓ_p height H_p (resp. *logarithmic ℓ_p height* h_p) of a rational number $q = u/v$ in lowest terms to be

$$H_p(q) := \left(\frac{|u|^p + |v|^p}{2}\right)^{1/p} \quad (\text{resp. } h_p(q) := \log H_p(q)).$$

(What does this mean for $p \in \{0, \infty\}$?) For any $p \in [0, \infty]$ and any $P \in E(\mathbb{Q})$, we have

$$\lim_{N\to\infty} \frac{1}{4^N} h_p((2^N P)) = \hat{h}(P).$$

In other words, we could have equally well chosen h_p as our definition of the naive height, and we would have obtained the same definition of the canonical height.

Exercise 2.5.9. Let A be an abelian group and R a ring (the two cases we'll be most interested in are $R = \mathbb{Z}$ and $R = \mathbb{R}$). A map $q : A \to R$ is said to be an *R-valued quadratic form* iff the following two conditions hold:

(a) For all $a \in A$ and all $n \in \mathbb{Z}$, we have $q(na) = n^2 q(a)$.
(b) The map $\langle \cdot, \cdot \rangle : A \times A \to R$ defined by

$$\langle a, b \rangle := q(a+b) - q(a) - q(b)$$

for $a, b \in A$ is $\mathbb{Z}$-bilinear, i.e., satisfies for all $a_1, a_2, b \in A$ that

$$\langle a_1 + a_2, b\rangle = \langle a_1, b\rangle + \langle a_2, b\rangle,$$

and similarly for the other side.

This is the abelian group analog of an inner product on vector spaces. PODASIP: A map $q : A \to R$ of sets is a(n) (R-valued) quadratic form iff the parallelogram law holds, i.e., for all $a, b \in A$, we have

$$q(a + b) + q(a - b) = 2q(a) + 2q(b).$$

Exercise 2.5.10. (For those who have taken a course on linear algebra.) Let $r \in \mathbb{Z}_{\geq 0}$, let $A := \mathbb{Z}^{\oplus r}$, and let $V := \mathbb{R} \otimes_{\mathbb{Z}} A \cong \mathbb{R}^r$, so that $A \subset V$ as a full-rank lattice.

(a) Suppose that $q : A \to \mathbb{R}$ is a quadratic form with the following two properties.
   (i) For $a \in A$, we have that $q(a) = 0$ iff $a = 0$.
   (ii) For every constant $C_1 \in \mathbb{R}_{\geq 1}$, the set $\{a \in A : q(a) \leq C_1\}$ is finite.
   Then q extends to a positive definite quadratic form on V. (Hint: Pick a suitable basis for V in which the extended $q : V \to \mathbb{R}$ is diagonal (what does that mean?). If q is *not* positive definite, then a clever application of Minkowski's Theorem finds a lattice point (i.e., an $a \in A$) with q-size less than $\inf_{a \in A \smallsetminus \{0\}} q(a)$.)

(b) (∗∗) Is the condition (ii) necessary in (a) above? Is it necessary if we allow A to have infinite rank? Can you make sense of [8, Remark VIII.9.4]?

Exercise 2.5.11. A sequence of homomorphisms of abelian groups

$$\cdots \xrightarrow{\partial^{i-1}} A^i \xrightarrow{\partial^i} A^{i+1} \xrightarrow{\partial^{i+1}} \cdots$$

is said to be a *complex* if $\partial^2 = 0$ (i.e., for all i we have $\partial^{i+1}\partial^i = 0 : A^i \to A^{i+2}$). It is said to be an *exact complex* or an *exact sequence* iff $\ker \partial^{i+1} = \operatorname{im} \partial^i$ for all i.

(a) A sequence of the form $0 \to A \xrightarrow{f} B$ is exact iff f is injective. A sequence of the form $B \xrightarrow{g} C \to 0$ is exact iff g is surjective. A sequence of the form $0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$ is exact iff f is injective, g is surjective, and f takes A isomorphically to $\ker g$, i.e., iff f is injective and $B/f(A) \xrightarrow{\sim} C$.

(b) If $0 \to A \to B \to C \to 0$ is an exact sequence of abelian groups, then if two of A, B, C are finite, then so is the third. If this holds, then the sizes of A, B, and C are related by $\#B = \#A \cdot \#C$. (Hint: The First Isomorphism Theorem.)

(c) Generalize (b) to exact sequences of any (finite) length. Specifically, if $n \in \mathbb{Z}_{\geq 0}$ and

$$0 \to A^0 \xrightarrow{\partial^0} A^1 \xrightarrow{\partial^1} \cdots \to A^n \to 0$$

is an exact sequence of finite abelian groups, and all but one of the A^i are finite, then all are finite. In this case, the sizes of the A^i are related by

$$\prod_{i=0}^{n} (\#A^i)^{(-1)^i} = 1.$$

(Hint: Break the "long exact sequence" into several "short exact sequences" of the form $0 \to \ker \partial^i \to A^i \to \ker \partial^{i+1} \to 0$, and then apply (a).)

(d) If $A \xrightarrow{f} B \xrightarrow{g} C$ is a sequence of homomorphisms of abelian groups, then we have an exact sequence of the form

$$0 \to \ker(f) \to \ker(gf) \xrightarrow{f} \ker(g) \to \operatorname{coker}(f) \xrightarrow{g} \operatorname{coker}(gf) \to \operatorname{coker}(g) \to 0.$$

In particular, if $\operatorname{coker}(f)$, and $\operatorname{coker}(g)$ are finite, then so is $\operatorname{coker}(gf)$. If further $\ker(f), \ker(gf)$ and $\ker(g)$ are also finite, then the sizes of these groups are related by

$$\#\ker(f) \cdot \#\ker(g) \cdot \#\operatorname{coker}(gf) = \#\ker(gf) \cdot \#\operatorname{coker}(f) \cdot \#\operatorname{coker}(g).$$

Exercise 2.5.12. If A is a finitely generated abelian group, then for any $n \in \mathbb{Z}_{\geq 1}$, we have that

$$\#A[n] = \#A/nA.$$

Exercise 2.5.13. Let K be a field of characteristic other than 3, and suppose further that K contains a primitive cube root of unity ω. (For example we could take $K = \mathbb{C}$ or $K = \mathbb{F}_p$ for a prime p such that $p \equiv 1 \pmod 3$.) Let $c \in K^\times$ and let E/K be the elliptic curve $E : y^2 = x^3 + c$. Let $[\omega] : E \to E$ be the automorphism given by $[\omega](x, y) := (\omega x, y)$. Then $[\omega]^2 + [\omega] + 1 = 0$ as maps from E to E, i.e., for all extensions L/K and $P \in E(L)$ we have that $[\omega]([\omega]P) + [\omega]P + P = \mathscr{O}$ in $E(L)$.

Exercise 2.5.14. Let $E/\mathbb{F}_3$ be the elliptic curve $y^2 = x^3 + x^2 - x$. The formula

$$(x, y) \mapsto \left(\frac{y^2}{x^2}, \frac{y(x^2+1)}{x^2}\right)$$

defines a self-map $\phi : E \to E$. There are unique $t, n \in \mathbb{Z}_{\geq 1}$ such that $\phi^2 - [t] \circ \phi + [n] = 0$ as maps from E to E, i.e., such that for all extensions $L/\mathbb{F}_3$ and $P \in E(L)$ we have that $\phi(\phi(P)) - \phi(tP) + nP = \mathscr{O}$ in the group $E(L)$. (This is an example of an endoisogeny of degree 2.)

Exercise 2.5.15. (For those who have taken a course on number fields.) Let K be a number field with ring of integers $\mathscr{O}_K$, and let S be a finite set of primes (or equivalently non-archimedean places) of $\mathscr{O}_K$. For each integer $n \in \mathbb{Z}_{\geq 1}$, define the subgroup $K(S, n) \subset K^\times/(K^\times)^n$ by

$$K(S, n) := \{[a] \in K^\times/(K^\times)^n : \text{ for all } \mathfrak{p} \notin S \text{ we have } v_\mathfrak{p}(a) \equiv 0 \pmod n\}.$$

(a) For instance, if $K = \mathbb{Q}$, then S is a finite set of prime numbers, and $\mathbb{Q}(S, n) \subset \mathbb{Q}^\times/(\mathbb{Q}^\times)^n$ is the subgroup generated by ± 1 and the images of the elements of S. In this case, $\#\mathbb{Q}(S, n) \leq 2 \cdot n^{\#S}$. (You may wish to improve this estimate depending on the parity of n.)

(b) In general, for any number field K and S and n, the group $K(S, n)$ is finite. Can you give an upper bound on the size of $\#K(S, n)$? (Hint: Describe a diagram of the form

$$\begin{array}{ccccccc} & & & & 0 & & \\ & & & & \downarrow & & \\ & & & & \mathscr{O}_K^\times/(\mathscr{O}_K^\times)^n & & \\ & & & & \downarrow & & \\ 0 & \longrightarrow & K(\emptyset, n) & \longrightarrow & K(S, n) & \longrightarrow & (\mathbb{Z}/n)^{\#S} \\ & & \downarrow & & & & \\ & & \mathrm{Cl}(K), & & & & \end{array}$$

where the row and column are exact, and use two fundamental finiteness results from the theory of number fields. Bonus question: what is the image of $K(\emptyset, n) \to \mathrm{Cl}(K)$?)

Exercise 2.5.16 (A 3-Isogeny). (Adapted from [1, Exercise 14.5].) (∗∗) Let $K = \mathbb{Q}$ (for simplicity). Let $c \in K^\times$, and consider the elliptic curve $E : y^2 = x^3 + c$. We will show that $E(K)/3E(K)$ is finite.

For simplicity, suppose also that c is not a square in K, and let $\gamma \in \overline{K}$ be a square root of c, i.e., so that $\gamma^2 = c$. Finally, let $L := K[\gamma]$, so that L is a quadratic extension of K.

(a) We have $T^\pm = (0, \pm\gamma) \in E[3](L)$, and that $2T^+ = T^-$ and $2T^- = T^+$.

(b) Find a formula for the operation $\tau_{T^\pm} : E \to E$ given by $P \mapsto P + T^\pm$. In other words, given $P = (x, y) \in E(L)$, express the x- and y-coordinates of $P + T^\pm$ in terms of those of P and γ.

(c) Let $P = (x, y) \in E(K)$. Express the following quantities as rational functions in x and y:

$$\xi := x(P) + x(P + T^+) + x(P + T^-) \text{ and } \eta := y(P) + y(P + T^+) + y(P + T^-).$$

Simplify until ξ and $\eta \cdot y^{-1}$ are functions of x alone.

(d) The rational functions ξ and η as in (c) satisfy an equation of the form $\eta^2 = \xi^3 + c'$ for some $c' \in K^\times$. Express c' in terms of c. If $E' : y^2 = x^3 + c'$, then $(x, y) \mapsto (\xi, \eta)$ defines a map $\phi : E \to E'$.

(e) As done in class, define a map $\phi' : E' \to E$ by essentially the same formula as for ϕ, except for c' replacing c, and some factors of 3 (which you should figure out). In this case, $\phi' \circ \phi : E \to E$ is the triplication map $[3] : E \to E$, and the same is true of the map $\phi \circ \phi' : E' \to E'$. (Hint: You may use Sage to speed up computations if this is helpful. See also 2.2.10(c).)

(f) The induced maps $\phi : E(K) \to E'(K)$ and $\phi' : E(K) \to E'(K)$ are group homomorphisms. (Hint: By symmetry, it suffices to show the result for ϕ. Use 1.15.5 and 1.12.18 to reduce to showing that if $P_1, P_2, P_3 \in E(K)$ are collinear, then so are $\phi(P_1), \phi(P_2)$ and $\phi(P_3)$. Now suppose that P_1, P_2, P_3 lie on the line $y = \lambda x + \nu$. Show that $\phi(P_1), \phi(P_2)$, and $\phi(P_3)$ lie on the line $y = \lambda' x + \nu'$, where

$$(\lambda', \nu') = \left(\frac{\lambda(\nu^2 + 3c)}{\nu^2 - c}, \frac{\nu^3 - 4c\lambda^3 - 9c\nu}{\nu^2 - c} \right).$$

In proving this, you may need to make use of the identity

$$(\lambda x + \nu)(x^3 - 8c) - \lambda' x(x^3 + 4c) - \nu' x^3 + \left(\frac{4c(\lambda x - 2\nu)}{\nu^2 - c} \right) (x^3 - (\lambda x + \nu)^2 + c) = 0,$$

which you can check on a computer algebra software like Sage. Finally, you will have to deal with some edge cases separately.)

(g) Suppose that $P = (x, y) \in E(K)$ has the property that $y + \gamma$ is a cube in L. Then there is a $P' \in E'(K)$ such that $P = \phi'(P')$. (Hint: Write $y + \gamma = (u + v\gamma)^3$, expand, and compare coefficients. Then use the formula for ϕ' to figure out what P' should be in terms of u and v.)

(h) The map $\alpha : E(K) \to L^\times/(L^\times)^3$ given by $(x, y) \mapsto [y + \gamma]$ is a group homomorphism with kernel precisely $\phi'(E'(K))$. In particular, α induces an injective homomorphism $E(K)/\phi'(E'(K)) \hookrightarrow L^\times/(L^\times)^3$.

(i) The image $\operatorname{Im}\alpha \subset L^\times/(L^\times)^3$ is finite. In particular, $E(K)/3E(K)$ is finite. (Hint: $\operatorname{Im}\alpha \subset L(S, 3)$ for a suitably chosen S.)

(j) Try to figure out what to do when c *is* a square in K. Which parts of the above argument still work, and which need to be modified? Can they be modified? (Hint: Let L still be $K[x]/(x^2 - c)$, even though this is no longer a field.)

(k) To what extent was the above analysis dependent on the fact that $K = \mathbb{Q}$? Over what other fields K do the above results still hold?

Exercise 2.5.17. (For those who have taken a course on number fields.) $(*)$ Let us finish the proof started in class of Fermat's Last Theorem for $n = 3$. Recall that in 1.2.B, we reduced the problem to showing that if $E/\mathbb{Q}$ is the elliptic curve $E : y^2 = x^3 - 432$, then $E(\mathbb{Q}) = \{\mathscr{O}, (12, \pm 36)\}$. As in class, let L be the cubic number field $L := \mathbb{Q}[\gamma]$ where $\gamma^3 = 2$, and let $\beta := 6\gamma$, so that $x^3 - 432 = (x - \beta)(x^2 + \beta x + \beta^2) \in L[x]$. Let $\alpha : E(\mathbb{Q}) \to L^\times/(L^\times)^2$ be the morphism defined in class, so that $\ker\alpha = 2E(\mathbb{Q})$.

(a) We have $\operatorname{Tors} E(\mathbb{Q}) = \{\mathscr{O}, (12, \pm 36)\}$. In particular, it suffices to show that $\operatorname{Im}\alpha = \{1\}$.

(b) We have that $\mathscr{O}_L = \mathbb{Z}[\gamma]$, and this is a UFD. The only rational primes that ramify in L are 2 and 3, and both are totally ramified: $(2) = (\gamma)^3$ and $(3) = (\gamma + 1)^3$. Further, L has fundamental unit $\gamma - 1$, so that $\mathscr{O}_L^\times = \{\pm(\gamma - 1)^{\mathbb{Z}}\}$. (Hint: See Exercises 2.41, 5.19, and 5.36 of [22].)

(c) In the notation from class, we have that $q(x) = x^2 + \beta x + \beta^2$, so that $q(\beta) = 108\gamma^2$. This factors in $\mathscr{O}_L$ as $(q(\beta)) = (\gamma)^8(\gamma + 1)^9$. In particular, if $S = \{(\gamma), (\gamma + 1)\}$, then $\operatorname{Im}\alpha \subset L(S, 2)$.

(d) By 2.5.15, the subgroup $L(S, 2) \subset L^\times/(L^\times)^2$ has size exactly 16. Explicitly, $\mathscr{O}_L^\times/(\mathscr{O}_L^\times)^2 \cong (\mathbb{Z}/2)^{\oplus 2}$ with elements represented by $\pm 1, \pm(\gamma - 1)$, and so $L(S, 2)$ has exactly 16 elements, represented by $\pm 1, \pm(\gamma - 1), \pm\gamma, \pm(\gamma^2 - \gamma), \pm(\gamma + 1), \pm(\gamma^2 - 1), \pm(\gamma^2 + \gamma), \pm(\gamma^3 - \gamma)$. (This shows already that $\operatorname{rank} E \leq 4$.)

(e) If $(x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, then under the real embedding $L \hookrightarrow \mathbb{R}$, we have $x - 6\gamma > 0$. In particular, no element with negative sign in the list in (d) is an element of $\operatorname{Im}\alpha$, i.e., $\operatorname{Im}\alpha \subset \{1, \gamma, \gamma \pm 1, \gamma^2 \pm \gamma, \gamma^2 - 1, 2 - \gamma\}$.

(f) For any $(x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, we have that $v_\gamma(x - \beta) \equiv 0 \pmod 2$. In particular, $\operatorname{Im}\alpha \subset \{1, \gamma \pm 1, \gamma^2 - 1\}$. (Hint: Write $(x, y) = (m/e^2, n/e^3)$ as in 1.14.8, so that

$$n^2 = (m - 6e^2\gamma)(m^2 + 6me^2\gamma + 36e^4\gamma^2).$$

If $v_\gamma(x - \beta) \equiv 1 \pmod 2$, then $\gamma \mid m$ and so m and n are even and e is odd. Write $m = 2m_1$ and $n = 2n_1$ for $m_1, n_1 \in \mathbb{Z}$; then in fact $2 \mid n_1$ and so $2 \mid m_1$. In particular, $v_\gamma(m_1 - 3e^2\gamma) = 1$, so that $v_\gamma(m - 6e^2\gamma) = 4$, a contradiction.)

(g) For any $(x, y) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}$, we have that $v_{\gamma+1}(x - \beta) \equiv 0 \pmod 2$. In particular, $\operatorname{Im}\alpha \subset \{1, \gamma - 1\}$. (Hint: If $\gamma + 1 \mid m$, then m and n are divisible by 3 and e is not. Write $m = 3m_1$ and $n = 3n_1$. Then $\gamma + 1$ divides either $m_1^2 - 2e^2\gamma$ or $m_1^2 + 2m_1e^2\gamma + 4e^4\gamma^2$. In the former case, $3 \mid m_1 + 2e^2$ and $v_{\gamma+1}(m_1 - 2e^2\gamma) = 1$, so that $v_{\gamma+1}(m - 6e^2\gamma) = 4$, a contradiction. In the latter case (and when *not* in the former), we get a contradiction to the fact that 3 totally ramifies in L.)

(h) It only remains to exclude the possibility that $\gamma - 1 \in \operatorname{Im}\alpha$, so suppose this happens. Then $x - 6\gamma = (\gamma - 1)(u\gamma^2 + v\gamma + w)^2$ for some $u, v, w \in \mathbb{Q}$, not all zero. Expanding, this gives us the system of equations

$$\begin{cases} x & = -4uv + 2v^2 + 4uw - w^2, \\ -6 & = -2u^2 + 4uv - 2vw + w^2, \text{ and} \\ 0 & = 2u^2 - v^2 - 2uw + 2vw. \end{cases}$$

The last conic $\mathbb{V}(2u^2 - v^2 - 2uw + 2vw) \subset \mathbb{P}^2_{\mathbb{Q}}$ admits the rational parametrization $[u : v : w] = [t(t+2s) : 2t(t+s) : t^2 - 2s^2]$. Plugging this into the second equation gives us a nontrivial rational point on

$$-6r^2 = 4s^4 + 8s^3t + 12s^2t^2 + 12st^3 + 3t^4,$$

but such a point does not exist. (Hint: Find a local obstruction to the existence of a rational point on this hyperelliptic curve. This example explains why the theory of binary quartics shows up in the investigation of rational points on elliptic curves.)

This finishes the proof of Fermat's Last Theorem for $n = 3$.

2.6 Exercise Sheet 6

2.6.A Numerical and Exploration

Exercise 2.6.1.

(a) In 1.22.13, we showed that there are no four-term sequences in arithmetic progression comprised of integer squares (although there are infinitely many such three-term sequences which are even coprime, by 2.3.4). What can you say about sequences of integer cubes in arithmetic progression? What about integer fourth powers?

(b) Here's an open-ended follow-up: what can you say about squares in arithmetic progression over other fields? Consider the case of finite fields $K = \mathbb{F}_q$ for $q = 2, 3, 4, 5, 7, 9, 11$, and the case of the number fields $K = \mathbb{Q}[\sqrt{409}]$ and $K = \mathbb{Q}[\beta]$ where $\beta^3 - \beta^2 + 1 = 0$. Suggested further reading: this article [30] by Xarles for an original investigation, and this article [31] by González-Jiménez (and the references cited therein) for the current state of knowledge on this topic (as of this writing in 2026).

Exercise 2.6.2. Let $m, n \in \mathbb{Z}_{\geq 0}$, and let M be an $m \times n$ integer matrix. For an $r \in \mathbb{Z}_{\geq 0}$ with $r \leq \min\{m, n\}$, by *an* $r \times r$*-minor* of M we mean the determinant of *a* square submatrix of M of size $r \times r$. For each such r, let $B_r(M) \subset \mathbb{Z}$ be the subgroup generated by *all* the $r \times r$-minors of M.

(a) For a fixed matrix M, what is the containment relationship between the various $B_r(M)$s as r varies? Make and prove conjectures.

(b) Suppose we put M into Smith Normal Form to obtain $(t; b_1, \ldots, b_t)$ as in 1.21.6. For each r, what is the relationship between the integer b_r and the subgroup $B_r(M) \subset \mathbb{Z}$? Work out lots of examples, and make and prove conjectures. Show in particular that these invariants $(t; b_1, \ldots, b_t)$ are uniquely determined by M.

Exercise 2.6.3. For a prime p, let $E_p : y^2 = x^3 + px$ be the elliptic curve over $\mathbb{Q}$ as in 1.22.14.

(a) Show that for each prime p, we have that $\operatorname{rank} E_p \in \{0, 1, 2\}$. Do all three possibilities occur?

(b) Compute the rank of E_2 (by hand, as done in 1.22.14).

(c) $(*)$ Finish the computation of the rank of E_{17} started in class by determining whether the homogeneous spaces $w^2 = -2u^4 + 34v^4$ and $w^2 = 17u^4 - 4v^4$ have any nontrivial rational points.

(d) Using Sage, compute the ranks of E_p for all primes $p \leq 10000$. (Hint: You may have to use `E.rank(only_use_mwrank = False)`.)

(e) What patterns do you notice in this table? For instance, what can you say about the relationship between $\operatorname{rank} E_p$ and the congruence class of $p \mod 16$? (Hint: Study the groups of congruence classes $\{7, 11\}$ and $\{3, 5, 13, 15\}$ and $\{1, 9\}$ separately.) Make conjectures.

(f) $(*)$ Prove some of your conjectures from part (e). For instance, formulate a conjecture in the case $p \equiv 7, 11 \pmod{16}$ and prove it. Similarly, show that if $p \equiv 3, 5, 13, 15 \pmod{16}$, then $\operatorname{rank} E_p \leq 1$.

Exercise 2.6.4. We saw in 1.22.14 that the elliptic curve $E_5 : y^2 = x^3 + 5x$ over $\mathbb{Q}$ has rank 1. In this exercise, we will figure out how to determine explicit generators of the Mordell-Weil group $E_5(\mathbb{Q})$.

(a) Use Nagell-Lutz (1.13.7) to determine $\operatorname{Tors} E_5(\mathbb{Q})$.

In the notation of 1.22.14, we had found that the point $(u, v, w) = (1, 1, 1)$ lies on the homogeneous space $w^2 = 5u^4 - 4v^4$ corresponding to the $b_1 = 5$ on the two-isogenous curve $E_5' : y^2 = x^3 - 20x$. Using 1.21.2, this corresponds to the rational point $(5, 5) \in E_5'(\mathbb{Q})$. Applying ϕ' to this gives us a point $(1/4, 9/8) \in E_5(\mathbb{Q})$, which after translation by $T = (0, 0)$ (see 1.20.10) gives us the points $\pm P = (20, \pm 90) \in E_5(\mathbb{Q})$. For the sake of concreteness, let $P := (20, 90)$. Note that, according to your answer from (a), we have $P \in E_5(\mathbb{Q}) \smallsetminus \operatorname{Tors} E_5(\mathbb{Q})$ (this also follows from $P + T = (1/4, -9/8)$ thanks to 1.13.7), so this means that P is a multiple of a generator of the free part of $E_5(\mathbb{Q})$. We will prove that P is a generator of the free part of $E_5(\mathbb{Q})$, which we combined with (a) will give us a complete description of $E_5(\mathbb{Q})$.

(b) Make the bound in 1.17.9(a) explicit. In other words, find an explicit constant $C \in \mathbb{R}_{\geq 0}$ such that for all $Q \in E_5(\mathbb{Q})$, we have that $|\hat{h}(Q) - h(Q)| \leq C$. (Hint: Trace through the proof to see how to produce such a C. You may not have to work through the entirety of 2.5.5; you can just try to use the bound from 1.16.6(c) along with a very naive upper boud. It should suffice to take C to be a very small positive integer.)

(c) Suppose that P is not a generator, so that there is a $Q \in E_5(\mathbb{Q})$ and an integer $m \in \mathbb{Z}_{\geq 2}$ such that

$P = mQ$. Using the fact that $\hat{h}$ grows quadratically and using your explicit bound from (b), show that the *naive* heights of P and Q are related by

$$h(Q) \leq \frac{1}{4}h(P) + \frac{5C}{4}.$$

(d) Write a simple program in Sage to manually search for all points Q on $E_5(\mathbb{Q})$ of naive height at most the bound found in (c). Use this to prove that P is indeed the generator of the Mordell-Weil group.

A similar technique (with various efficiency enhancements) can used in general to verify whether a given point is a Mordell-Weil generator; combined with the result of 2.5.3(d), this technique is often successful in practice at finding a complete basis of the Mordell-Weil lattice after having performed two-descent.

Exercise 2.6.5. For $N \in \mathbb{Z}_{\geq 0}$, let $C \subset \mathbb{P}^2_{\mathbb{Q}}$ be the cubic curve given by 1.23.2; recall that this came from the "fruits meme" after clearing denominators in the equation

$$\frac{a}{b+c} + \frac{b}{a+c} + \frac{c}{a+b} = 4$$

as in lecture.

(a) Check that C is smooth.
(b) Show that $[1 : -1 : 0] \in C(\mathbb{Q})$ is a rational inflection point. Check that the transformation 1.23.3 is invertible and brings C into Weierstrass form $E : y^2 = x^3 + 109x^2 + 224x$. Figure out the inverse of 1.23.3.
(c) Use Nagell-Lutz (1.13.7) to determine $\operatorname{Tors} E(\mathbb{Q})$.
(d) If you're feeling energetic, use two-descent as in 1.22.B to compute $\operatorname{rank} E$, and then use the argument of 2.6.4 to find explicit generators of $E(\mathbb{Q})$. If you're feeling less energetic (but still enthusiastic about this question), ask Sage to do this for you. In any case, get that $P = (-4, 28)$ is a generator of $E(\mathbb{Q})$.
(e) Prove 1.23.4, which determines which $(x, y) \in E(\mathbb{Q})$ correspond to positive values of a, b, c.
(f) Show that one of the points $9P + Q$ for $Q \in \operatorname{Tors} E(\mathbb{Q})$ gives rise to a solution a, b, c of the above equation with $a, b, c \in \mathbb{Z}_{\geq 1}$. Show also that the points $nP + Q$ for $1 \leq |n| \leq 8$ and $Q \in \operatorname{Tors} E(\mathbb{Q})$ do *not* give rise to positive whole solutions a, b, c. (Hint: These computations are very difficult (impossible?) to do by hand, but Sage handles them quite well.)
(g) Here's an open-ended follow-up: can you repeat the same procedure with 4 replaced by an arbitrary $N \in \mathbb{Z}_{\geq 1}$? See how far you can get. Suggested further reading: [20] for a detailed discussion of this problem, and and [21] for many such similiar "recreational" examples.

Exercise 2.6.6 (RSA Cryptosystem)**.** (Adapted from [2, Exercise 4.25]) This exercise explains the basic idea behind the RSA cryptosystem.

Suppose that Alex would like to send Billie messages over a channel that can be invaded by Charlie. To do this, Billie first selects two large primes p and q; these primes are kept secret. Consider the integers $N = pq$ and e, where $1 < e < (p-1)(q-1)$ and $\gcd(e, (p-1)(q-1)) = 1$; the integers N and e are released to the public.

(a) Alex encrypts the desired message, expressed via a predetermined padding scheme to be an element $m \in \mathbb{Z}/N\mathbb{Z}$, by computing the value of the ciphertext $c \in \mathbb{Z}/N\mathbb{Z}$ given by $c \equiv m^e \pmod{N}$. Then Alex sends c to Billie along the possibly insecure channel. Show that, since Billie knows the values of p and q, she can efficiently find an integer f such that for all $m \in \mathbb{Z}/N\mathbb{Z}$ we have $m^{ef} \equiv m \pmod{N}$, and use this to recover the message m from c. The idea is that it will be very hard for Charlie to find the right value of f which "undoes the encryption" unless Charlie knows p and q; in other words, the security of the RSA cryptosystem relies on the difficulty of the factoring problem for very large integers N.
(b) Show that the knowledge of p and q is equivalent to the knowledge of $N = pq$ and $(p-1)(q-1)$ in the sense that each pair of integers can be recovered from the other. In particular, if Charlie knows $(p-1)(q-1)$, then Charlie can easily decrypt the message.

Exercise 2.6.7. (Adapted from [2, Exercise 4.21]) Suppose you are given a big integer $N \in \mathbb{Z}_{\geq 1}$ that you would like to factor using Lenstra's algorithm. As explained in class, the basic idea is to consider the point $P = (1, 1)$ on the "elliptic curve" over $\mathbb{Z}/n\mathbb{Z}$ given by $E_b : y^2 = x^3 + bx - b$ for various $b = 1, 2, \ldots$.

Using the extremely efficient double-and-add algorithm described in class, we can then compute multiples of points on E_b over $\mathbb{Z}/N\mathbb{Z}$, so suppose we try to compute say $(20!) \cdot P$ in E_b. If the computations go through, we pick the next b and repeat. If the computation breaks down for some step, then we have (hopefully!) found a factor of N.

Write simple Sage code to implement this algorithm, and use it to factor the integers $N = 199843247$ and $N = 4686706420729$. (Of course, Sage can just factor these numbers very quickly. The point of this exercise is to see how Lenstra's algorithm might be implemented in practice.)

2.6.B PODASIPs

Prove or disprove and salvage if possible the following statements.

Exercise 2.6.8. Let K be a field and $C \subset \mathbb{P}^2_K$ be an integral plane curve. Then C is *generically smooth*, i.e., for every field extension L/K, all but finitely many points of $C(L)$ are smooth points. (Hint: Consider $\mathbb{V}(X^2 + tY^2) \subset \mathbb{P}^2_{\mathbb{F}_2(t)}$. For a strong salvage, consider the case of perfect K or geometrically integral C.)

Exercise 2.6.9. Let K be a field and let $a_1, \ldots, a_6 \in K$ be any elements (in particular, we don't require that $\Delta \neq 0$). Let $E \subset \mathbb{P}^2_K$ be the curve given by 1.9.3. Then E is geometrically integral. (Hint: Use 2.3.6 to reduce to showing that a polynomial $f(x, y) \in K[x, y]$ is irreducible. Note that $f(x, y)$ is quadratic in y. Find general criteria for when such a polynomial is irreducible, possibly by invoking Gauss's Lemma to pass between $K[x, y]$ and $K(x)[y]$.)

Exercise 2.6.10. Let K be a perfect field and $C \subset \mathbb{P}^2_K$ a geometrically integral cubic curve. We investigate what singularities of C can look like. Suppose C is singular. By 2.2.7, there is a unique singular point $P_0 \in C(K)$. By a projective change of coordinates, we can take P_0 to $[0 : 0 : 1]$.

(a) $(**)$ There is a projective change of coordinates which brings C into exactly one of the following two forms

$$C_a : Y^2Z = X^3 \quad \text{or} \quad C_{m,\alpha} : Y^2Z = X^3 + \alpha X^2 Z$$

for some $\alpha \in K^\times$. (Hint: This result is incorrect as stated in characteristics 2 and 3. You may also need to consider the "Artin-Schreier cubic" $Y^2Z + XYZ + X^2Z = X^3$ in characteristic 2 and the "strange cusp" $Y^2Z = X^3 + X^2Y$ in characteristic 3.)

A curve which can be brought into the form C_a is said to be *cuspidal* and one which can be brought into the form $C_{m,\alpha}$ *nodal*; if we can further make $\alpha = 1$, then the curve is called *split nodal*, and if we cannot then it is called *nonsplit nodal*.

(b) Whether or not the nodal curve $Y^2Z = X^3 + \alpha X^2 Z$ is split depends only on the class of α in $K^\times/(K^\times)^2$. Every nodal curve over an algebraically closed field is split. The curve $Y^2Z = X^3 - X^2Z$ is a nonsplit nodal curve over $\mathbb{Q}$. (Hint: Again, something funny is going on when $\operatorname{ch} K = 2$.)

(c) Consider the cuspidal curve $C_a = \mathbb{V}(Y^2Z - X^3) \subset \mathbb{P}^2_K$. Look at the chart $Y \neq 0$ with coordinates $(t, s) = (X/Y, Z/Y)$ in which C_a looks like $s = t^3$. The map $t : C_a(K) \smallsetminus \{P_0\} \to \mathbb{G}_a(K) = K$ is an isomorphism of abelian groups when the left hand side is given the group law discussed in the lecture.

(d) Consider the split nodal curve $C_m = C_{m,1} = \mathbb{V}(Y^2Z - X^3 - X^2Z) \subset \mathbb{P}^2_K$ when $\operatorname{ch} K \neq 2$. As in (b), we get the affine model $s = t^3 + t^2 s$. The map $C_m(K) \smallsetminus \{P_0\} \to \mathbb{G}_m(K) = K^\times$ given by

$$(t, s) \mapsto \frac{1 - t}{1 + t}$$

is a group isomorphism.

(e) $(**)$ Finally, consider the non-split nodal curve $C_{m,\alpha} = \mathbb{V}(Y^2Z - X^3 + \alpha X^2 Z) \subset \mathbb{P}^2_K$ when $\operatorname{ch} K \neq 2$ and $\alpha \neq [1] \in K^\times/(K^\times)^2$. Let $L := K[\sqrt{\alpha}]$ be the quadratic extension of K obtained by adjoining a root of α. The group $C_{m,\alpha}(K) \smallsetminus \{P_0\}$ is isomorphic to the subgroup $\{\beta \in L^\times : \mathrm{N}^L_K(\beta) = 1\} \subset L^\times$. For instance, when $K = \mathbb{R}$ and $\alpha = -1$, we have that $C_{m,-1}(\mathbb{R}) \smallsetminus \{P_0\} \cong_{\mathsf{Ab}} S^1$, where $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ is the circle group (draw a picture!). What happens in characteristic two?

(f) How far can you generalize the above discussion when we don't require K to be perfect or C to be geometrically integral? What works still, and what breaks down?

Exercise 2.6.11. Let K be a field, let $a_1, \dots, a_6 \in K$, let $E \subset \mathbb{P}^2_K$ be the curve defined by 1.9.3, let Δ be as in 1.9.2 and let c_4 be as in 1.11.14.

(a) The curve E is smooth iff $\Delta \neq 0$. (This was basically shown in 1.9.9; the point of this exercise is to ask you to recall this result and fill in all the gaps in the proof left to the reader there.)
(b) The curve E is nodal iff $\Delta = 0$ and $c_4 \neq 0$. If $\operatorname{ch} K \neq 2$, then E is split nodal iff $\Delta = 0$ and c_4 is a nonzero square. What happens if $\operatorname{ch} K = 2$? (Hint: How does the invariant c_4 transform under admissible changes of coordinates? When $\operatorname{ch} K = 2$, you may have to split into cases depending on whether K is perfect.)
(c) The curve E is cuspidal iff $\Delta = c_4 = 0$.

This is the geometric meaning of the invariant c_4 corresponding to a Weierstrass equation, at least when the curve it defines is singular.

Exercise 2.6.12. A sequence (b_k) of integers is said to be a *divisibility sequence* if we have that for all $d, n \in \mathbb{Z}_{\geq 1}$ that if $d \mid n$ then $b_d \mid b_n$. It is said to be a *strong divisibility sequence* (also called a *GCD sequence* on the Ross Sets) if for all $m, n \in \mathbb{Z}_{\geq 1}$ we have that $\gcd(b_m, b_n) = b_{\gcd(m,n)}$.

(a) A GCD sequence is a divisibility sequence.
(b) For any $b \in \mathbb{Z}_{\geq 1}$, the sequence defined by $b_k := b^k - 1$ is a GCD sequence. The Fibonacci numbers $b_k = F_k$ form a GCD sequence.
(c) Let $E/\mathbb{Q}$ be an elliptic curve, say $E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ with all $a_i \in \mathbb{Z}$. Let $P \in E(\mathbb{Q})$. Define the sequence (b_k) of positive integers by

$$b_k := \begin{cases} 1 & \text{if } kP = \mathscr{O}, \text{ and} \\ e & \text{if } kP = \left(\dfrac{m}{e^2}, \dfrac{n}{e^3}\right) \in E(\mathbb{Q}) \smallsetminus \{\mathscr{O}\}, \end{cases}$$

where as in 2.4.14 we have $m, n, e \in \mathbb{Z}$ with $e \geq 1$ and $\gcd(m, e) = \gcd(n, e) = 1$. The sequence (b_k) is a divisibility sequence. Is it a GCD sequence in general?

See also 2.6.20.

Exercise 2.6.13. Let A be a finitely generated abelian group, so that thanks to 1.21.4, A has the form

$$\mathbb{Z}^r \oplus \mathbb{Z}/a_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/a_s\mathbb{Z}$$

for integers $r, s \in \mathbb{Z}_{\geq 0}$ and $a_1, \dots, a_s \in \mathbb{Z}_{\geq 2}$ such that $a_1 \mid a_2 \mid \cdots \mid a_s$.

(a) We have that $r = \dim_{\mathbb{F}_p} A/pA$ for all but finitely many primes p. In particular, r is uniquely determined by A.
(b) The integer $r + s$ is the maximum value of $\dim_{\mathbb{F}_p} A/pA$ as p ranges over all primes. In particular, s is uniquely determined by A.
(c) Suppose now that $s \geq 1$. The integer a_s is the least positive integer a such that $a \cdot \operatorname{Tors}(A) = 0$. (In other words, a_s is the *annihilator* of $\operatorname{Tors}(A)$.) In particular, a_s is uniquely determined by A.
(d) $(*)$ More generally, for each $t \in \mathbb{Z}_{\geq 1}$ with $1 \leq t \leq s$, we have that a_t is the least positive integer a such that the subgroup $a \cdot \operatorname{Tors}(A)$ of A can be generated as an abelian group by at most $s - t$ elements. In particular, a_t is uniquely determined by A.

Exercise 2.6.14. Let K be a field and E/K an elliptic curve such that $E(K)$ is finitely generated. This happens, for instance, when K is a finite field, or when $K = \mathbb{Q}$ (1.15.8), or more generally when K is a number field (1.15.8, which we have admittedly not proven). Let $s(E/K)$ denote the number of elementary divisors of the finitely generated abelian group $E(K)$ (1.21.4). Then we necessarily have $s(E/K) \in \{0, 1, 2\}$, and all three cases occur in general. (Hint: You. may use 1.12.28 and 1.12.29, which we also have not fully proven.)

Exercise 2.6.15. For an $n \in \mathbb{Z} \smallsetminus \{0\}$, let $\omega(n)$ denote the number of distinct prime factors of n. Let $a, b \in \mathbb{Z}$ and $(a', b') := (-2a, a^2 - 4b)$ as usual. Suppose that $bb' \neq 0$, so that $E : y^2 = x^3 + ax^2 + bx$ defines an elliptic curve over $\mathbb{Q}$. In this case,

$$\operatorname{rank} E \leq \omega(b) + \omega(b') - 1.$$

This inequality is sharp. How might this generalize to number fields? (Hint: 1.21.2 and obstructions at infinity.)

Exercise 2.6.16. A positive rational number $D \in \mathbb{Q}_{>0}$ is said to be *congruent* iff there is a right-angled triangle with rational sides and area equal to D. Note that this property is visibly invariant under multiplication by rational squares, so we might as well focus on the case of squarefree $D \in \mathbb{Z}_{\geq 1}$. We showed in class (1.22.7) that 1 is not a congruent number. The problem of determining which squarefree positive integers are congruent–also known as the *congruent number problem*–is very old and goes back at least to the 10^{th} century (see the discussion of the history in [13, §1.1]). Henceforth, by the term *congruent number*, we mean a *congruent squarefree positive integer*.

(a) List at least 10 distinct congruent numbers. Which of the integers $D \leq 10$ are congruent? Check that the first few numbers listed in the OEIS sequence A003273 agree with your answer to the previous question.
(b) There are infinitely many congruent numbers.
(c) There are infinitely many right-angled triangles with rational sides and area 5.
(d) $(**)$ The number D is congruent iff the elliptic curve $E_D : y^2 = x^3 - D^2x$ over $\mathbb{Q}$ has positive rank, in which case there are infinitely many right-angled triangles with rational sides and area D. (Hint: See 2.3.10, which cites [1, §12].)

For more on the congruent number problem, see [13, §1.1], [32, Chapter 1], and [33].

Exercise 2.6.17. (Adapted from [1, §18, Lemma 1].) Suppose $a, b, c \in \mathbb{Z}_{\geq 2}$ are distinct integers such that $d := abc$ is cubefree. Consider the cubic curves $C_{a,b,c}, D_d, E_d \subset \mathbb{P}^2_{\mathbb{Q}}$ given by

$$C_{a,b,c} = \mathbb{V}(aX^3 + bY^3 + cZ^3), \quad D_d = \mathbb{V}(X^3 + Y^3 - dZ^3), \quad E_d := \mathbb{V}(Y^2Z - X^3 + 432d^2Z^3).$$

If $C_{a,b,c}(\mathbb{Q}) \neq \emptyset$, then $E_d(\mathbb{Q}) \neq \{\mathscr{O}\}$, where $\mathscr{O} = [0 : 1 : 0]$ as usual. (Hint: We met D_d in 2.2.4. We saw in 1.2.B that D_d is isomorphic over $\mathbb{Q}$ to E_d. Now suppose $[X : Y : Z] \in C_{a,b,c}(\mathbb{Q})$. Let ω be a primitive cube root of unity. If $\xi := aX^3 + \omega bY^3 + \omega^2 cZ^3$, then $[\xi : \overline{\xi} : 3XYZ] \in D_d(\mathbb{Q}[\omega])$. Finally, use 2.1.14.)

Exercise 2.6.18. (For those who know about Hensel's Lemma.)

(a) The equation $w^2 = 2u^4 - 34v^4$ of 1.22.15 is everywhere locally unobstructed, i.e., has a nontrivial solution (u, v, w) over $\mathbb{Q}_p$ for each prime $p \leq \infty$. In particular, 1.22.15 gives a genuine counterexample to the local-to-global principle.[5]
(b) In the notation of 2.6.17, let $C = C_{3,4,5} \subset \mathbb{P}^2_{\mathbb{Q}}$; this is the curve of 1.2.7. The curve C is everywhere locally unobstructed, i.e., for each prime $p \leq \infty$, we have $C(\mathbb{Q}_p) \neq \emptyset$.[6]
(c) $(**)$ (This needs algebraic number theory.) The curve C of (b) has no rational points, i.e., $C(\mathbb{Q}) = \emptyset$. In particular, C is another counterexample to the local-to-global principle. (Hint: Use two-descent as in 2.5.17 to compute the rank of E_{60} and hence find all rational points on it.)

2.6.C Challenge Problems

Exercise 2.6.19. (Adapted from [8, Exercises 6.12-13]) For $k \in [0, 1)$, define the *complete elliptic integrals to the modulus* k to be

$$K(k) := \int_0^1 \frac{\mathrm{d}u}{\sqrt{(1-u^2)(1-k^2u^2)}} \text{ and } T(k) := \int_0^1 \sqrt{\frac{1-k^2u^2}{1-u^2}}\mathrm{d}u.$$

(a) We showed in 1.12.A, that the ellipse $C_{a,b} : (x/a)^2 + (y/b)^2 = 1$ with $a \geq b > 0$ has perimeter $P = 4aT(e)$, where $e := \sqrt{1 - (b/a)^2}$ denotes the eccentricity of $C_{a,b}$. The integral $K(k)$ can similarly be expressed in terms of the ellipse $C_{a,b}$; namely, if we let

$$I(a, b) := \int_0^{\pi/2} \frac{\mathrm{d}t}{\sqrt{a^2\cos^2 t + b^2 \sin^2 t}},$$

then $I(a, b) = a^{-1} \cdot K(e)$.
(b) $(*)$ For all $a, b \in \mathbb{R}$ with $a \geq b > 0$,

$$I(a, b) = I\left(\frac{a+b}{2}, \sqrt{ab}\right).$$

[5] If you get stuck, a solution can be found in [18]. Please attempt the problem yourself first!
[6] If you get stuck, a solution can be found in [34]. Please attempt the problem yourself first!

In particular, K satisfies the functional equation

$$K\left(\frac{2\sqrt{k}}{1+k}\right) = (1+k)K(k).$$

(Hint: Use the substitution $x = b\tan t$ to express $I(a,b)$ as the integral of a function from 0 to ∞. Then use the substitution $x = y + \sqrt{y^2+ab}$. A different substitution achieving this–due to Gauss–can be found in [35].)

(c) Let S be the arclength of the lemniscate $r^2 = \cos(2\theta)$. Then

$$S = 2\sqrt{2}\cdot K\left(\frac{1}{\sqrt{2}}\right) = 4\int_0^1 \frac{\mathrm{d}x}{\sqrt{1-x^4}}.$$

(d) For initial values $a, b \in \mathbb{R}$ with $a \geq b > 0$, we recursively define the sequences $\{a_n\}$ and $\{b_n\}$ by $(a_0, b_0) := (a, b)$ and

$$(a_{n+1}, b_{n+1}) := \left(\frac{a_n + b_n}{2}, \sqrt{a_n b_n}\right)$$

for $n \in \mathbb{Z}_{\geq 0}$. Then that

$$a \geq a_1 \geq a_2 \geq \cdots \geq b_2 \geq b_1 \geq b,$$

and $\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n$. This limit is called the *arithmetic-geometric mean* (AGM) of a and b, and is denoted $\mathrm{agm}(a,b)$. Write a simple program in the language of your choice which takes as input two real numbers a, b as above and returns $\mathrm{agm}(a,b)$, and use this to compute $\mathrm{agm}(a,b)$ for various values of a and b. How rapid do you observe the convergence $a_n, b_n \to \mathrm{agm}(a,b)$ to be? Make and prove conjectures.

(e) For $a, b \in \mathbb{R}$ with $a \geq b > 0$, we have that

$$\mathrm{agm}(a,b)\cdot I(a,b) = \frac{\pi}{2}.$$

In particular, for any $k \in [0,1)$ we have that

$$K(k) = \frac{\pi}{2}\cdot\frac{1}{\mathrm{agm}(1,\sqrt{1-k^2})}.$$

This can be used to compute $K(k)$ very efficiently. Finally, we have

$$\int_0^1 \frac{\mathrm{d}x}{\sqrt{1-x^4}} = \frac{\pi}{2}\cdot\frac{1}{\mathrm{agm}(\sqrt{2},1)}.$$

According to [8], "the observation that these two numbers, calculated independely, agreed to eleven decimal places led Gauss to initiate an extensive study of the arithmetic-geometric mean". For more details, see [8, Exercises 6.12-13] and [35].

Exercise 2.6.20. (Adapted from [5, Exercise 2.6.11] and [8, Exercises 3.34-3.36].) Let K be a field. A *nondegenerate elliptic divisibility sequence* (EDS) over K is a sequence (b_k) of elements of K determined by four initial parameters b_1, b_2, b_3, b_4 with $b_1 b_2 b_3 \neq 0$ subject to the recursive relations

$$b_{2k+1} = \frac{1}{b_1^3}\left(b_{k+2}b_k^3 - b_{k-1}b_{k+1}^3\right), \text{ and}$$
$$b_{2k} = \frac{1}{b_1^2 b_2} b_k (b_{k+2}b_{k-1}^2 - b_{k-2}b_{k+1}^2)$$

for all $k \geq 2$.

(a) The identity sequence $b_k = k$ is an EDS. The sequence $b_k = F_{2k-1}$ of the odd-indexed Fibonacci numbers is an EDS. More generally, given $b_1, b_2, u, v \in K$, the sequence (b_k) defined by the linear recursive relation

$$b_k = ub_{k-1} + vb_{k-2}$$

for $k \geq 2$ is an EDS. (Hint: Find a subsequence as for the Fibonacci numbers that *is* an EDS.)

(b) If (b_k) is an EDS, then for each $m \geq 1$ such that $b_m \neq 0$, so is the sequence $(b_{mk}/b_m)_k$. An EDS such that $b_1 = 1$ is said to be *normalized*; given any sequence b we define its *normalization* $\tilde{b}$ to be given by $\tilde{b}_k = b_k/b_1$ for $k \geq 1$. Given a normalized EDS (b_k), we define its *discriminant* to be

$$\Delta := b_4 b_2^{15} - b_3^3 b_2^{12} + 3b_4^2 b_2^{10} - 20 b_4 b_3^3 b_2^7 + 3b_4^3 b_2^5 + 16 b_3^6 b_2^4 + 8 b_4^2 b_3^2 b_2^2 + b_4^4.$$

We say that a EDS is *singular* if the discriminant of its normalization is zero; else it is said to be *nonsingular*. Which of the sequences from (a) are nonsingular?

(c) Assume for simplicity that $\operatorname{ch} K \neq 2, 3$. Let $E : y^2 = x^3 + bx + c$ be an elliptic curve over K, and let $P = (x_0, y_0) \in E(K)$. The sequence (b_k) defined by

$$b_k = \begin{cases} f_k(x_0) & k \text{ odd, and} \\ 2y_0 \cdot f_k(x_0), & k \text{ even,} \end{cases}$$

is an EDS, where the polynomials f_k are as in Exercise 2.2.10. What is the discriminant of (the normalization of) this sequence b_k? Is this sequence singular?

(d) The sequence (b_k) is an EDS iff for each $m > n > r > 0$, we have

$$b_{m+n} b_{m-n} b_r^2 = b_{m+r} b_{m-r} b_n^2 - b_{n+r} b_{n-r} b_m^2.$$

(e) Now suppose that $K = \operatorname{Frac} R$ for some integral domain R, and let (b_k) be an EDS over K such that $b_1, b_2, b_3, b_4 \in R$ and such that $b_1 \mid b_k$ for $k = 2, 3, 4$ and $b_2 \mid b_4$. Then (b_k) is a divisibility sequence in the sense of 2.6.12. If, further, R is a PID and $\gcd(b_3, b_4) = 1$, then (b_k) is a GCD sequence. In particular, these properties hold for the Fibonacci sequence F_k.

(f) Finally suppose that $K = \mathbb{R}$. Suppose that (b_k) is a nonsingular non-periodic EDS. Then there is a real number $h > 0$ such that

$$\lim_{n \to \infty} \frac{\log |b_n|}{n^2} = h.$$

Bibliography

[1] J. W. S. Cassels, *Lectures on Elliptic Curves*. No. 24 in London Mathematical Society Student Texts, Cambridge University Press, 1991.

[2] J. H. Silverman and J. T. Tate, *Rational Points on Elliptic Curves*. Undergraduate Texts in Mathematics, Springer, 2nd ed., 2015.

[3] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, vol. 84 of *Graduate Texts in Mathematics*. Springer, second ed., 1990.

[4] J.-P. Serre, *A Course in Arithmetic*. Graduate Texts in Mathematics, Springer-Verlag New York, 1973.

[5] D. Goel, "Plane Algebraic Curves." `https://gdmgoel.github.io/notes/Curves_Full.pdf`.

[6] W. Fulton, *Algebraic Curves: An Introduction to Algebraic Geometry*. third ed., 2008. `https://users.math.msu.edu/users/magyarp/Math419H/Fulton-Alg-Curves.pdf`.

[7] R. Hartshorne, *Algebraic Geometry*. No. 52 in Graduate Texts in Mathematics, Springer, 1977.

[8] J. H. Silverman, *The Arithmetic of Elliptic Curves*, vol. 106 of *Graduate Texts in Mathematics*. Springer, 2nd ed., 2016.

[9] R. Vakil, *The Rising Sea: Foundations of Algebraic Geometry*. Princeton University Press, 2025.

[10] D. A. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Undergraduate Texts in Mathematics, Springer, fifth ed., 2025.

[11] W. W. Adams and P. Loustaunau, *An Introduction to Gröbner Bases*, vol. 3 of *Graduate Studies in Mathematics*. American Mathematical Society, 1994.

[12] D. Husemöller, *Elliptic Curves*, vol. 111 of *Graduate Texts in Mathematics*. Springer, 2nd ed., 2010.

[13] Á. Lozano-Robledo, *Elliptic Curves, Modular Forms, and Their L-functions*, vol. 58 of *Student Mathematical Library IAS/Park City Mathematical Subseries*. American Mathematical Society, Institute for Advanced Study, 2009.

[14] M. Kronberg, M. A. Soomro, and J. Top, "Twists of Elliptic Curves," *Symmetry, Integrability and Geometry: Methods and Applications*, Oct. 2017. `https://arxiv.org/abs/1707.02893`.

[15] L. V. Ahlfors, *Complex Analysis: An Introduction to the Theory of Analytic Functions of One Complex Variable*. International Series in Pure and Applied Mathematics, McGraw-Hill, Inc., third ed., 1979.

[16] N. D. Elkies and G. Goel, "On powerful integers expressible as sums of two coprime fourth powers," *Research in Number Theory*, vol. 9, no. 4, p. 78, 2023. `https://doi.org/10.1007/s40993-022-00415-9`.

[17] J. E. Cremona, *Algorithms for Modular Elliptic Curves*. online ed. `https://johncremona.github.io/book/fulltext/index.html`.

[18] D. B. Leep, P. Pollack, and D. B. Shapiro, "Revisiting the Lind-Reichardt counterexample to Hasse's Local-Global Principle," *The American Mathematical Monthly*, vol. 132, no. 9, pp. 839–847, 2025. https://doi.org/10.1080/00029890.2025.2540254.

[19] A. Amit, "Answer to *How do you find positive integer solutions to* $\frac{x}{y+z} + \frac{y}{z+x} + \frac{z}{x+y} = 4$*?*." Quora. Accessed at https://www.quora.com/How-do-you-find-the-positive-integer-solutions-to-frac-x-y+z-+-frac-y-z+x-+-frac-z-x+y-4/answer/Alon-Amit on 2026-07-29.

[20] A. Bremner and A. Macleod, "An unusual cubic representation problem," *Annales Mathematicae et Informaticae*, vol. 43, pp. 29–41, 2014. https://dornsife.usc.edu/sergey-lototsky/wp-content/uploads/sites/211/2026/03/ASimpleEquation.pdf.

[21] A. Macleod, "Elliptic Curves in Recreational Number Theory." https://arxiv.org/pdf/1610.03430, 2016.

[22] D. Marcus, *Number Fields*. Universitext, Springer, second ed., 2018.

[23] A. Sutherland, "Constructing elliptic curves over finite fields with prescribed torsion," *Mathematics of Computation*, vol. 81, p. 1131–1147, Aug. 2011. https://arxiv.org/abs/0811.0296.

[24] M. Case, "A Beginner's Guide To The General Number Field Sieve." https://www.cs.umd.edu/~gasarch/TOPICS/factoring/NFSmadeeasy.pdf.

[25] N. M. Katz and B. Mazur, *Arithmetic Moduli of Elliptic Curves. (AM-108)*. Princeton University Press, 1985. http://www.jstor.org/stable/j.ctt1b9s05p.

[26] R. Miranda, *Algebraic Curves and Riemann Surfaces*, vol. 5 of *Graduate Studies in Mathematics*. American Mathematical Society, 1995.

[27] F. Diamond and J. Shurman, *A First Course in Modular Forms*. 228, Graduate Texts in Mathematics, Springer, 2005.

[28] J. Harris, *Algebraic Geometry: A First Course*, vol. 133 of *Graduate Texts in Mathematics*. Springer.

[29] K. Conrad, "Arithmetic Progressions of Four Squares." https://kconrad.math.uconn.edu/blurbs/ugradnumthy/4squarearithprog.pdf.

[30] "Squares in arithmetic progression over number fields," *Journal of Number Theory*, vol. 132, no. 3, pp. 379–389, 2012. https://www.sciencedirect.com/science/article/pii/S0022314X11002101.

[31] E. González-Jiménez, "Squares in arithmetic progression over quadratic extensions of number fields," *International Journal of Number Theory*, vol. 22, p. 1141–1158, Feb. 2026. https://arxiv.org/abs/2602.03251.

[32] N. Koblitz, *Introduction to Elliptic Curves and Modular Forms*, vol. 97 of *Graduate Texts in Mathematics*. Springer, 2nd ed., 1993.

[33] K. Conrad, "The Congruent Number Problem." https://kconrad.math.uconn.edu/blurbs/ugradnumthy/congnumber.pdf.

[34] K. Conrad, "Selmer's Example." https://kconrad.math.uconn.edu/blurbs/gradnumthy/selmerexample.pdf.

[35] David, "The Arithmetic-Geometric Mean of Gauss," *L'Enseignement Mathématique*, vol. 30, pp. 275–330, 1984. https://webspace.science.uu.nl/~wepst101/elliptic/cox_agm.pdf.