Chapter 3

Elliptic curves

At the end of the previous chapter we mentioned Sylvester's problem: determine which prime numbers p can be written as the sum of two rational cubes

$$u^3 + v^3 = p, \qquad u, v \in \mathbf{Q}.$$

We can be more ambitious and ask: what are all the solutions (if such exist) $(u, v) \in \mathbf{Q}^2$ of this equation?

Here is another example of such a problem: given $c \in \mathbf{Z}$, can we write c as the difference of a rational square and a rational cube? What are all the solutions (if any) $(x, y) \in \mathbf{Q}^2$ of the equation

$$y^2 - x^3 = c?$$

A couple of surprising things happen here.

The first is that the two problems are actually very closely related.

Proposition 3.1. *The function*

$$f(u, v) = \left(\frac{12p}{u+v}, \frac{36p(u-v)}{u+v} \right)$$

is a bijection from the set of rational solutions $(u, v) \in \mathbf{Q}^2$ *of* $u^3 + v^3 = p$ *to the set of rational solutions* (x, y) *of* $y^2 - x^3 = -432p^2$*. Its inverse is*

$$g(x, y) = \left(\frac{36p + y}{6x}, \frac{36p - y}{6x} \right).$$

Proof. Tutorial question. □

This means that if we know how to find all rational solutions to the equations E_c (at least for c of the form $-432p^2$ for primes p), then we can find all rational cubes that add up to p. (Since I mentioned that the latter problem is still open, this is an indication that the former problem is probably hard.)

The second surprise, which was found in 1621 by Bachet, is that given a solution $(x_0, y_0) \in \mathbf{Q}^2$ of E_c, if we let

$$x_1 = \frac{x_0^4 - 8cx_0}{4y_0^2}, \quad y_1 = \frac{-x_0^6 - 20cx_0^3 + 8c^2}{8y_0^3},$$

then (x_1, y_1) is another solution of E_c. In fact, it turns out that if $c \neq 1, -432$ and $x_0 y_0 \neq 0$, then as we iterate this formula we get infinitely many distinct solutions of E_c.

It also turns out that Bachet's discovery is not an accident, but a feature of equations that give rise to so-called elliptic curves.

Before getting into that, let's note that we have been asking about rational solutions to $y^2 - x^3 = c$. What about integer solutions?

Example 3.2. Consider the equation $y^2 - x^3 = -2$. After some trial and error we may find $(x, y) = (3, 5)$, and of course since y is squared in the equation, we automatically get $(3, -5)$ as another integer solution.

Bachet's trick doesn't help here: taking $(x_0, y_0) = (3, 5)$ we compute

$$(x_1, y_1) = \left(\frac{129}{100}, -\frac{383}{1000}\right)$$

which is not integral, and it only gets worse from here on.

Fermat conjectured that $(3, \pm 5)$ are the only integer solutions of this equation, but it took several hundred years before this was proved.

3.1 Basic properties of elliptic curves

It is time to define elliptic curves. We will do so in a fairly concrete manner, and we will restrict our attention to certain fields of particular interest to us.

Let $\mathbf{F}$ denote one of the following fields[1]: $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$, $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$ for primes $p \geqslant 5$.

A *Weierstrass equation* is a polynomial equation of the form

$$y^2 = x^3 + ax + b, \qquad a, b \in \mathbf{F}.$$

The *discriminant* of the above Weierstrass equation is defined to be $\Delta = -16(4a^3 + 27b^2)$.

We say that the Weierstrass equation defines an *elliptic curve* if $\Delta \neq 0$.

Most of the time we conflate the Weierstrass equation and the elliptic curve it defines.

Nonsingularity condition

What is the meaning of the condition $\Delta \neq 0$?

It has to do with the curve being nonsingular. The intuitive picture is that at each point (x_0, y_0) on the curve, there exists a well-defined (unique) tangent line.

This is a multivariable calculus topic. Generally, if we fix a differentiable function F in two variables and a point (x_0, y_0) such that $F(x_0, y_0) = 0$, we say that the curve defined by $F(x, y) = 0$ is *singular at (x_0, y_0)* if

$$\frac{\partial F}{\partial x}(x_0, y_0) = 0 = \frac{\partial F}{\partial y}(x_0, y_0).$$

If this condition does not hold, the curve is said to be *nonsingular at (x_0, y_0)*. The curve is *nonsingular* if it is nonsingular at every point.

What does this mean for our curves? Write $f(x) = x^3 + ax + b$ and $F(x, y) = y^2 - f(x)$, then $F(x, y) = 0$ is the type of equation we are considering. Suppose the resulting curve is singular at (x_0, y_0), then

$$0 = \frac{\partial F}{\partial x}(x_0, y_0) = -f'(x_0) = -3x_0^2 - a \quad \text{and} \quad 0 = \frac{\partial F}{\partial y}(x_0, y_0) = 2y_0.$$

So $f(x_0) = y_0 = 0$ and $f'(x_0) = 0$. Therefore x_0 is at least a double root of the polynomial $f(x)$.

We can follow this to a criterion for (non)singularity of Weierstrass equations:

[1] Other fields, for instance of characteristic 2, would of course be highly desirable, especially for computational applications. These require more complicated Weierstrass equations, and additional care in many of the proofs, so we are not considering them here.

Proposition 3.3. *Let $y^2 = x^3 + ax + b$ be a Weierstrass equation and let $\Delta = -16(4a^3 + 27b^2)$ be its discriminant. The curve defined by this equation is singular if and only if $\Delta = 0$.*

Proof. Suppose (x_0, y_0) is a singular point, then by the discussion before the proposition we have that $x_0^3 + ax_0 + b = 0$ and $3x_0^2 + a = 0$.

If $a = 0$, then $3x_0^2 = 0$ so $x_0 = 0$ so $b = 0$ so $\Delta = 0$.

If $a \neq 0$, note that

$$0 = 3(x_0^3 + ax_0 + b) - x_0(3x_0^2 + a) = 2ax_0 + 3b,$$

so $x_0 = -\dfrac{3b}{2a}$. Plugging this into $3x_0^2 + a = 0$ we end up with $27b^2 + 4a^3 = 0$ so $\Delta = 0$.

Conversely, suppose $\Delta = 0$. If $a = 0$ then $b = 0$ so the equation is $y^2 = x^3$, which is singular at $x_0 = 0$ (0 is a triple root of x^3).

If $a \neq 0$, consider $x_0 = -\dfrac{3b}{2a}$. We have

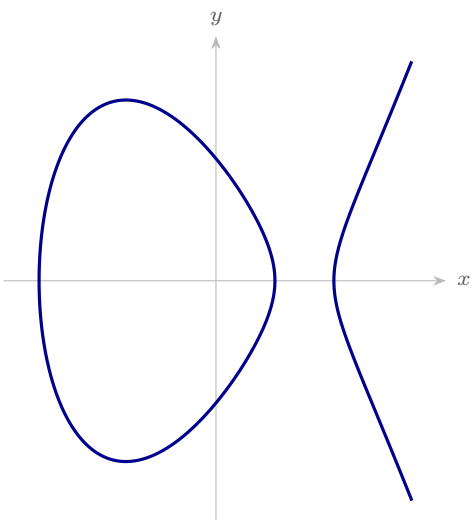
$$f'(x_0) = 3\,\frac{9b^2}{4a^2} + a = \frac{27b^2 + 4a^3}{4a^2} = 0$$

and

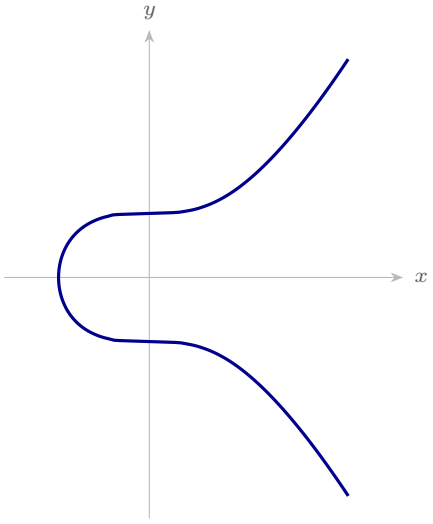
$$f(x_0) = -\frac{27b^3}{8a^3} - a\,\frac{3b}{2a} + b = \frac{-27b^3 - 12a^3b + 8a^3b}{8a^3} = \frac{4a^3b - 4a^3b}{8a^3} = 0. \qquad \square$$

So the nonvanishing of the discriminant Δ is equivalent to the cubic polynomial $x^3 + ax + b$ having simple roots.

Over $\mathbf{R}$, $x^3 + ax + b$ has either one or three real roots. Correspondingly, the picture for an elliptic curve over the real numbers is of one of the following two types:



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

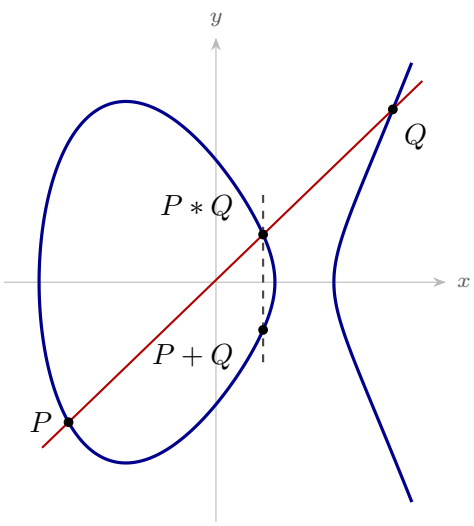


$E:\ y^2 = (x+1)(x^2 - x + 1)$

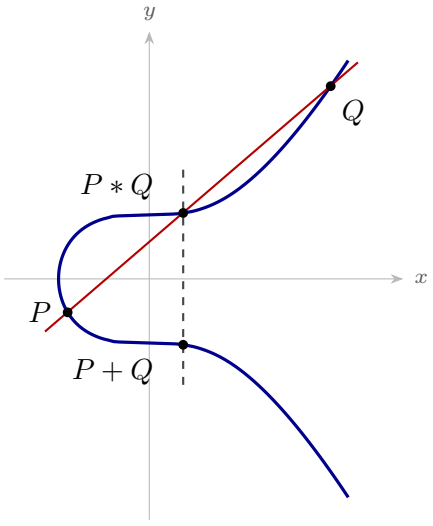
The miracle: elliptic curves have a group law

We now get to the property that really distinguishes elliptic curves from other types of plane curves: the set of points on an elliptic curve has the additional structure of an abelian group.

The group operation can be defined geometrically, via the so-called "chord-and-tangent" rule.



$E: \ y^2 = x^3 - 7x + 6$



$E: \ y^2 = x^3 + 1$

Suppose we have two points $P \neq Q$ on an elliptic curve. Consider the line passing through the points P and Q; since the line has degree 1 and the curve has degree 3 then, generically, there will be a third point of intersection, which we call $P * Q$. We then let $P + Q$ denote the reflection of the point $P * Q$ about the x-axis. (This final point $P + Q$ will be on the curve, since the Weierstrass equation is symmetric about the x-axis.)

I claim that the operation $(P, Q) \mapsto P + Q$ makes the elliptic curve into an abelian group. Making sense of this claim requires answering a number of questions:

- What is the neutral element $\mathcal{O}$? It should have the property that $P + \mathcal{O} = P$ for all points P.

  Thinking through the recipe given above, this forces $P * \mathcal{O}$ to be the reflection of P about the x-axis, hence $\mathcal{O}$ should be the third point of intersection between the curve and the straight line joining P to its reflection. But that straight line is vertical so… how can this be?

  The answer is that we should add a point to the curve. It is a point $\mathcal{O}$ that lies infinitely far along any vertical line in the plane. We refer to it as *the point at infinity*.

  We make the convention that $\mathcal{O} + \mathcal{O} = \mathcal{O}$, and that the reflection of $\mathcal{O}$ about the x-axis is $\mathcal{O}$.

  (There is a much less hand-wavy definition involving the projective plane and the projectivisation of the curve, but for our purposes this intuitive approach suffices.)

- What is the inverse $-P$ of a point P?

  Having established what $\mathcal{O}$ is, the point $-P$ should be the unique point on the curve such that $P + (-P) = \mathcal{O}$. This means that $P * (-P) = \mathcal{O}$; as we argued above, this forces $-P$ to be the reflection of P about the x-axis. In terms of coordinates, the inverse of $P = (x, y)$ is $-P = (x, -y)$.

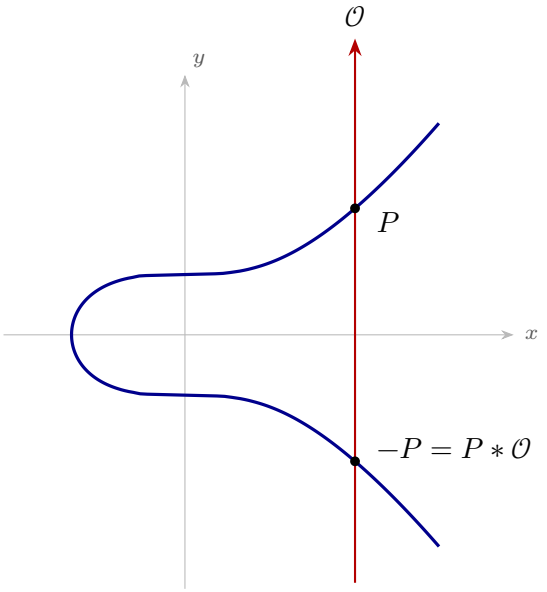
- The group operation $P + Q$ should be defined for all pairs of points (P, Q), but in the above description we asked that $P \neq Q$. So how do we define $P + P$?

  We appeal to calculus ideas here. If $P \neq Q$ then the line PQ is secant to the curve. As the point Q approaches P along the curve, this secant line becomes the tangent line at P.

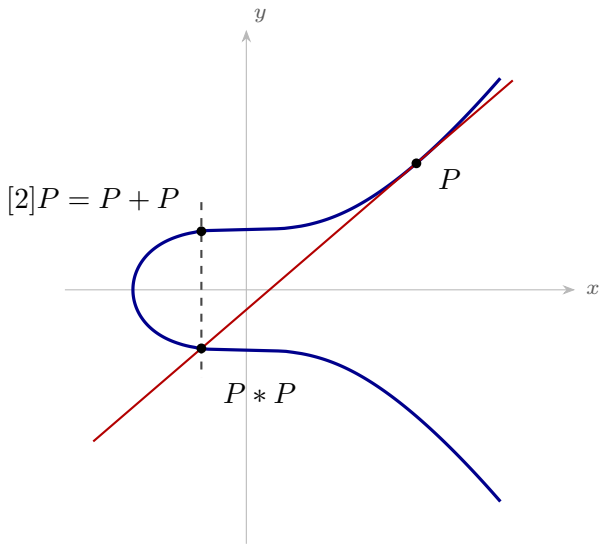
  So we apply the same recipe using the tangent line at P. Since this line intersects the curve with multiplicity 2, there is another point of intersection, which we call $P * P$, then $P + P$ is the reflection of $P * P$ about the x-axis.

We should really have a different notation for $P+P$; we will prefer to write $[2]P$ instead of $2P$ so we do not confuse this with multiplying the coordinates of P by 2.

More generally, we write $[n]P$ for the n-fold sum $P+P+\cdots+P$ for a positive integer P, as well as $[0]P$ for $\mathcal{O}$ and $[-n]P$ for $[n](-P)$.



$E:\ y^2 = x^3 + 1$



$E:\ y^2 = x^3 + 1$

- Why are the (other) axioms of an abelian group satisfied in this setting? This is extremely tedious and we will take it for granted. (There are more conceptual ways to obtain the group law that require a lot more theory and a lot less tedious calculation.)

Computing the group operation

While the elliptic curve pictures are pretty, we need a more practical way of determining $P+Q$ for points P and Q. Expressing this algebraically also allows us to work with elliptic curves over fields like $\mathbf{C}$ or $\mathbf{Z}/p\mathbf{Z}$, where the pictures do not really make sense.

Let E be the elliptic curve defined by the Weierstrass equation

$$E:\ y^2 = x^3 + ax + b, \qquad a, b \in \mathbf{F}.$$

Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be two points on the curve E.

If $x_1 = x_2$, then $y_1^2 = y_2^2$, so $(y_1 - y_2)(y_1 + y_2) = 0$. Since $\mathbf{F}$ is a field, we conclude that $y_1 = y_2$ (in which case $P = Q$) or $y_1 = -y_2$ (in which case $P = -Q$). In the first case we are really asking what $[2]P = P + P$ is, and we treat this separately a little later; in the second case we have $P + (-P) = \mathcal{O}$.

So we assume for now that $x_1 \neq x_2$.

Proposition 3.4 (Addition Formula)**.** *With the above notation, we have $P + Q = (x_3, y_3)$, where*

$$\begin{aligned} x_3 &= m^2 - x_1 - x_2 \\ y_3 &= -(mx_3 + n) \\ m &= \frac{y_2 - y_1}{x_2 - x_1} \\ n &= y_1 - mx_1 = y_2 - mx_2. \end{aligned}$$

Proof. This is a matter of translating the geometric construction into algebra.

It suffices to determine $P * Q = (x_3, Y_3)$, then $y_3 = -Y_3$.

The line passing through P and Q has slope $m = \dfrac{y_2 - y_1}{x_2 - x_1}$ and equation

$$y = mx + n, \qquad n = y_1 - mx_1 = y_2 - mx_2.$$

We plug this into the Weierstrass equation of the curve:

$$(mx + n)^2 = y^2 = x^3 + ax + b,$$

so that after some manipulation

$$x^3 - m^2x^2 + (a - 2mn)x + (b - n^2) = 0.$$

This equation has as roots the x-coordinates of the three points P, Q, $P * Q$ of intersection between the line and the curve. In other words

$$x^3 - m^2x^2 + (a - 2mn)x + (b - n^2) = (x - x_1)(x - x_2)(x - x_3).$$

Now we compare the quadratic terms on both sides and find that $-m^2 = -x_1 - x_2 - x_3$, therefore

$$x_3 = m^2 - x_1 - x_2.$$

The rest follows from $Y_3 = mx_3 + n$ and $y_3 = -Y_3$. □

It is of course possible to express m^2 in terms of x_1, x_2, y_1, y_2 in the formula for x_3 and get a somewhat unwieldy formula expressing x_3 solely in terms of the coordinates of P and Q, but it is a bit easier to do the computations in steps.

Example 3.5. Consider the curve given by the equation $E:\ y^2 = x^3 + 17$. A short search gives points $P = (2, 5)$ and $Q = (-1, 4)$.

Then

$$m = \frac{1}{3}, \qquad n = 5 - \frac{2}{3} = \frac{13}{3},$$

so that

$$x_3 = \frac{1}{9} - 2 + 1 = -\frac{8}{9}$$

and

$$Y_3 = \frac{1}{3} \cdot \left(-\frac{8}{9}\right) + \frac{13}{3} = \frac{109}{27}.$$

Therefore

$$P * Q = \left(-\frac{8}{9}, \frac{109}{27}\right)$$
$$P + Q = \left(-\frac{8}{9}, -\frac{109}{27}\right).$$

The remaining case $P = Q$ is treated similarly but involves tangent lines instead of secant lines.

Proposition 3.6 (Duplication Formula)**.** *Let $P = (x_1, y_1)$ be a point on the elliptic curve given by*

$$E:\ y^2 = x^3 + ax + b.$$

If $[2]P \neq \mathcal{O}$ *then* $[2]P = (x_2, y_2)$*, where*

$$\begin{aligned} x_2 &= m^2 - 2x_1 \\ y_2 &= -(mx_2 + n) \\ m &= \frac{3x_1^2 + a}{2y_1} \\ n &= y_1 - mx_1. \end{aligned}$$

Proof. We find the tangent line at $P = (x_1, y_1)$ to the curve $y^2 = x^3 + ax + b$ by implicit differentiation:

$$2y\,\frac{dy}{dx} = 3x^2 + a,$$

so the slope of the tangent line is

$$m = \frac{dy}{dx} = \frac{3x_1^2 + a}{2y_1}.$$

The equation of the tangent line is then $y = mx + n$ with $n = y_1 - mx_1$.

Plugging this into the Weierstrass equation and proceeding as in the proof of the addition formula gives

$$x^3 - m^2x^2 + (a - 2mn)x + (b - n^2) = (x - x_1)^2(x - x_2).$$

From the quadratic terms we get $-m^2 = -2x_1 - x_2$, so

$$x_2 = m^2 - 2x_1.$$ □

For future reference we record the formula for $[2]P$ solely in terms of P:

Corollary 3.7. *If* $P = (x_1, y_1)$ *satisfies* $[2]P \neq \mathcal{O}$ *then* $[2]P = (x_2, y_2)$*, where*

$$\begin{aligned} x_2 &= \frac{x_1^4 - 2ax_1^2 - 8bx_1 + a^2}{(2y_1)^2}, \\ y_2 &= \frac{x_1^6 + 5ax_1^4 + 20bx_1^3 - 5a^2x_1^2 - 4abx_1 - a^3 - 8b^2}{(2y_1)^3}. \end{aligned}$$

Isomorphisms

Consider a Weierstrass equation over a field $\mathbf{F}$ (feel free to replace $\mathbf{F}$ with your favourite field that is not $\mathbf{Z}/2\mathbf{Z}$ or $\mathbf{Z}/3\mathbf{Z}$):

$$E:\ y^2 = x^3 + ax + b, \qquad a, b \in \mathbf{F}.$$

Pick $u \neq 0$ and define $\hat{x}$ by $x = u^2\hat{x}$ and $\hat{y}$ by $y = u^3\hat{y}$. Then $\hat{x}$ and $\hat{y}$ are related by

$$(u^3\hat{y})^2 = (u^2\hat{x})^3 + a(u^2\hat{x}) + b \quad \Rightarrow \quad u^6\hat{y}^2 = u^6\hat{x}^3 + au^2\hat{x} + b,$$

in other words by the Weierstrass equation

$$\hat{E}:\ \hat{y}^2 = \hat{x}^3 + \hat{a}\hat{x} + \hat{b}, \qquad \text{where } a = u^4\hat{a}, b = u^6\hat{b}.$$

Note that $\Delta = -16(4a^3 + 27b^2) = u^{12}\hat{\Delta}$, so that E defines an elliptic curve if and only if $\hat{E}$ defines an elliptic curve.

It turns out that this kind of change of coordinates is the only algebraic transformation that takes a Weierstrass equation to another Weierstrass equation. We say that two elliptic curves E and $\hat{E}$ are *isomorphic* if there exists $u \neq 0$ such that their Weierstrass equations are related by such a change of coordinates. In this case we write $E \cong \hat{E}$.

Example 3.8. Recall the family of Weierstrass equations considered by Bachet:

$$E_c : \ y^2 = x^3 + c.$$

Fix $n \in \mathbf{Z}_{\geqslant 1}$ and let $u = 1/n$, then proceeding as above we get

$$\hat{E} : \ \hat{y}^2 = \hat{x}^3 + n^6 c,$$

so we conclude that $E_c \cong E_{n^6 c}$ for all $n \in \mathbf{Z}_{\geqslant 1}$.

Example 3.9. Take a Weierstrass equation over $\mathbf{Q}$ (or $\mathbf{R}$)

$$E : \ y^2 = x^3 + ax + b$$

and let $u = i = \sqrt{-1}$. Then we get

$$\hat{E} : \ \hat{y}^2 = \hat{x}^3 + a\hat{x} - b.$$

Depending on what we are trying to do, we may wish to allow such isomorphisms, or to disallow them by imposing that u should be an element of the field of definition.