Last time we stated *Fermat's Last Theorem*: for any integer $n \geq 3$, there are no nontrivial integer solutions to the equation $x^n + y^n = z^n$. (Recall that (x, y, z) is *nontrivial* iff $xyz \neq 0$.) Although first conjectured by Fermat in 1637, this resisted all efforts at proof until the 1990s, when Andrew Wiles (then at Princeton, now at Oxford) succeeded in proving this assertion. Central to Wiles' approach are *elliptic curves*, which are curves of the form $y^2 = x^3 + ax + b$. What's the connection between such curves and the Fermat equation? Let's consider an example.

Suppose we have a nontrivial integer triple (a, b, c) such that $a^3 + b^3 = c^3$. Our goal is to get something in the form of an elliptic curve, i.e. a square of something equalling a cube of something plus some linear correction factors. A bit of playing around might lead you to consider the quantity $(a^3 + 2b^3)^2$:

$$(a^3 + 2b^3)^2 = a^6 + 4(ab)^3 + 4b^6 = a^6 + 4b^3(a^3 + b^3) = a^6 + 4b^3c^3.$$

Dividing by a^6,

$$\frac{(a^3 + 2b^3)^2}{a^6} = 1 + \frac{4b^3c^3}{a^6}$$

Notice the left hand side is a square, and the right hand side is 1 plus something that's almost a cube; the only issue is the 4. To fix this we multiply by 16:

$$\underbrace{\frac{16(a^3 + 2b^3)^2}{a^6}}_{y^2} = 16 + \underbrace{\frac{64b^3c^3}{a^6}}_{x^3}.$$

In other words, starting with a nontrivial solution to FLT with $n = 3$ produces a nontrivial rational point (x, y) on the elliptic curve $y^2 = x^3 + 16$. So to prove FLT for $n = 3$ all we have to do is to show that this elliptic curve doesn't have any nontrivial rational points. This isn't so easy to do, but it turns out there are exactly two rational points on $y^2 = x^3 + 16$: $(0, \pm 4)$.

What about rational points on other elliptic curves? To get a feel for this, here's a sampler of elliptic curves and their rational points (the values in the middle column aren't obvious!):

$\mathcal{E}$	Total # of rational points on $\mathcal{E}$	rational points
$y^2 = x^3 + 16$	2	$(0, \pm 4)$
$y^2 = x^3 + 1$	5	$(-1, 0), (0, \pm 1), (2, \pm 3)$
$y^2 = x^3 - 1$	1	$(1, 0)$
$y^2 = x^3 - 2$	infinitely many	$(3, \pm 5), \ldots$
$y^2 = x^3 - 5$	none	none

This table shows that even for the most bare-bones elliptic curve of the form $y^2 = x^3 + k$ (sometimes called *Bachet's equation* or *Mordell's equation* in the literature), it can be hard to predict how many rational points might be on it.

Earlier we defined an elliptic curve to be an equation of the type $y^2 = x^3 + ax + b$. It will be useful to shift our perspective and re-define an elliptic curve to be the set of points on this curve. Formally:

Definition: An *elliptic curve* is any set of the form

$$\mathcal{E} := \{(x,y) : y^2 = x^3 + ax + b\}$$

where a, b are fixed.

At first glance, this seems a bit pedantic. Who cares about whether we're discussing the equation or the points satisfying that equation?! But there's something vital missing from the above definition: where the points (x,y) live. Are they real numbers? Rational numbers? Complex numbers? The corresponding shape of the set drastically changes depending on the types of inputs we allow. For example, the set of solutions to

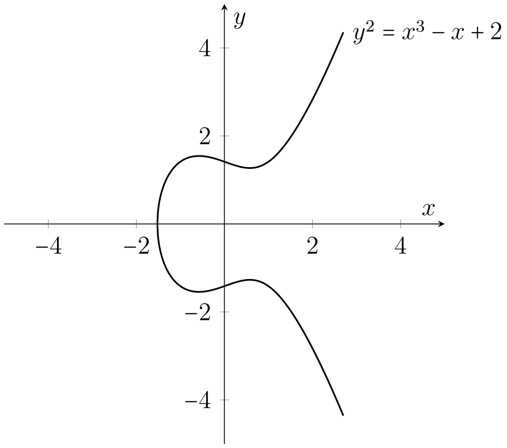
$$x^3 = 2$$

is empty if we require x to be rational, consists of precisely one point if we allow x to be real, and consists of three points if we allow x to be complex.

To specify where our elliptic curve lives, we write

$$\mathcal{E}(S) := \{(x,y) \in S^2 : y^2 = x^3 + ax + b\}.$$

Thus, for example, if our elliptic curve $\mathcal{E}$ is defined by the equation $y^2 = x^3 - x + 2$, then $\mathcal{E}(\mathbb{R})$ is the set of all points in the plane $\mathbb{R}^2$ satisfying the equation $y^2 = x^3 - x + 2$. In other words, $\mathcal{E}(\mathbb{R})$ is the graph of $y^2 = x^3 - x + 2$:

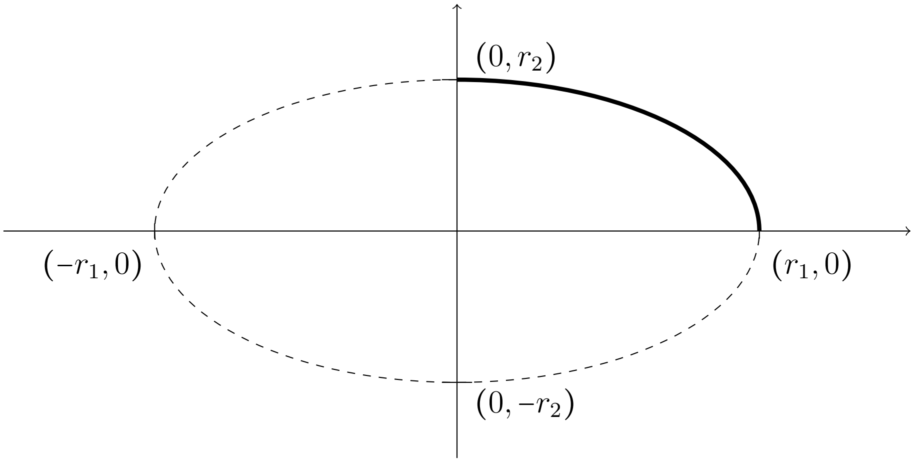


For other elliptic curves, $\mathcal{E}(\mathbb{R})$ might look different; you should try visualizing $\mathcal{E}(\mathbb{R})$ for the curve $y^2 = x^3 - x$.

As we saw above, shifting to $\mathcal{E}(\mathbb{Q})$ may yield no points, or infinitely many points, or just a few; we'll return to this shortly. If we look at $\mathcal{E}(\mathbb{C})$, we have to look at 4 dimensions, since we have two numbers with 2 dimensions each, but it turns out to have the structure of a torus (a fancy word for donut). For this reason, elliptic curves are usually abstractly defined as "curves of genus 1"; here the word *genus* means the number of holes (so a donut is genus 1 because it has one hole).

Digression on nomenclature

One natural question is: what the heck does any of this have to do with ellipses? Why are they called *elliptic curves*? The story is a bit convoluted, so bear with me. Recall that we have nice formulas for the area and circumference of a circle (πr^2 and $2\pi r$, respectively), as well as for the area of an ellipse: $\pi r_1 r_2$, where r_1 and r_2 are the major and minor radii of the ellipse. (This formula is intuitive: take a unit circle (area = π), stretch it in the x-direction by a factor of r_1, and in the y-direction by a factor of r_2.) What you might not have thought about is: what is the circumference of an ellipse? By using techniques from single-variable calculus, it's not too hard to write down a formula for the arc length of the part of the ellipse in the first quadrant:



$$\frac{1}{4} \times \text{circumference} = \int_0^{\pi/2} \sqrt{r_1^2 \cos^2\theta + r_2^2 \sin^2\theta}\, d\theta$$

Remarkably, no simpler formula for the circumference is known. In particular, we don't know how to express the circumference without using an integral sign!

The above represents the arclength of the part of the ellipse with subtended angle $\pi/2$. There are more complicated formulas for the arclength of the part of the ellipse with subtended angle α. While investigating this, Abel was led to ask the opposite question: is there a formula to express the subtended angle α in terms of the arclength, say, $\alpha = f(\ell)$? It turns out that there is, and that this function has an unexpectedly nice property: when viewed as a function over the complex numbers f is doubly periodic, i.e. there exist $\omega_1, \omega_2 \in \mathbb{C}$ such that $f(z) = f(z+\omega_1) = f(z+\omega_2)$ for all $z \in \mathbb{C}$. This led Weierstrass to formulate the notion of an *elliptic function*: any function $f : \mathbb{C} \to \mathbb{C}$ that is non-constant and doubly periodic (with the two periods being linearly independent, i.e. genuinely different). One way to think about this is as the natural generalization of the notion of a trigonometric function (a periodic function on $\mathbb{R}$) to $\mathbb{C}$.

Elliptic functions turn out to be useful in multiple areas, particularly in differential equations and number theory. One of the early examples of an elliptic function is the *Weierstrass* $\wp$ function. I won't state its definition here, but it turns out this function satisfies the differential equation

$$\left(\frac{1}{2}\wp'(z)\right)^2 = \wp(z)^3 + a\wp(z) + b$$

for certain constants a and b. This led Jacobi to study equations of the form $y^2 = x^3 + ax + b$, which became known as *elliptic curves*. In summary: elliptic curves are a family of equations generalizing a differential equation satisfied by a specific elliptic function, and elliptic functions are a generalization of a function coming from the study of the circumference of an ellipse.

As you can see, the appearance of ellipses in the term *elliptic curve* is more of a historical accident rather than descriptive nomenclature. However, a remarkable recent discovery by Agarwal and Natarajan gives a direct connection between ellipses and elliptic curves. Given a triangle, we say an ellipse is *inscribed* in the triangle if it's tangent to all three sides. For a given triangle T, consider the set of all foci of all ellipses that can be inscribed in T; it turns out this forms an elliptic curve. Even more amazingly, every elliptic curve can be realized this way!

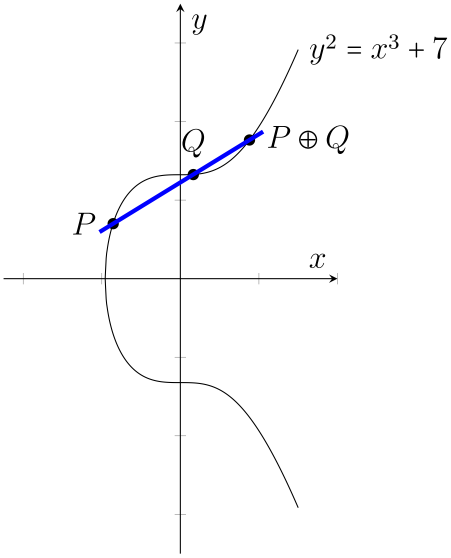
Back to rational points

Recall that (inspired by Fermat's Last Theorem) we were looking at $\mathcal{E}(\mathbb{Q})$, the set of all rational points on a given elliptic curve $\mathcal{E}$.

Question: How do we find rational points on a given elliptic curve?

Rather than answering this question, we pivot and show how, *given* some rational solutions, to generate others.

Idea (Bachet + Fermat, 1600s): Suppose you have two rational points, P, Q on $\mathcal{E}$ (i.e. in $\mathcal{E}(\mathbb{Q})$). Can you imagine a way to use these two points on $\mathcal{E}$ to generate a third? There's a natural geometric way to do so: draw the line that passes through P and Q, and find the third point where it hits $\mathcal{E}$.



Label this third point $P \oplus Q$. Amazingly, it turns out this operation preserves rationality:

Theorem 1: If P, Q are rational points, then $P \oplus Q$ is rational.

Miranda noticed this does not always work. Given a point P on the elliptic curve, consider the point $\overline{P}$ that's a reflection of P across the x-axis.[1] If P is rational, then so is $\overline{P}$. But the line connecting P and $\overline{P}$ doesn't intersect the elliptic curve anywhere else! To formally satisfy Theorem 1, we can write $P \oplus \overline{P} = \infty$. This inspires a definition:

Definition: Given any elliptic curve $\mathcal{E}$, we create a point ∞ and stipulate that $\infty \in \mathcal{E}(\mathbb{Q})$. Moreover, by convention, for any $P \in \mathcal{E}$ we have $P \oplus \overline{P} := \infty$.

This formally resolves Miranda's question, in that Theorem 1 now holds even when the line connecting P and $\overline{P}$ is vertical. Well, not quite: what if $\overline{P} = P$, i.e. if P is on the x-axis? By our definition, we expect $P \oplus \overline{P} = \infty$. Max observed that this makes sense from a geometric perspective as well: we can view the line connecting P to $\overline{P}$ as the vertical line tangent to

[1] The notation is inspired by complex conjugation: in the complex plane, the complex conjugate $\overline{z}$ is the reflection of z across the real axis.

the elliptic curve at P. In fact, this inspires a more general idea: for *any* rational point P, we defined $P \oplus P$ to be the intersection of the line tangent to the elliptic curve at P with the elliptic curve. It turns out Theorem 1 holds in this case as well. Cool!

Ben pointed out that this seems to contradict our table of examples (from page 1): we asserted there that the curve $y^2 = x^3 - 1$ only has the single rational point $(1, 0)$, but Theorem 1 implies that we can add this point to itself to get a new rational point! In fact, $(1, 0) \oplus (1, 0) = \infty$, which we also consider rational. This means that, according to our new convention with ∞, our table is a bit wrong: the number of rational points on $y^2 = x^3 - 1$ is 2, and more generally, all the numbers in the middle column should increase by one.

Mia observed that implicit in our discussion of Theorem 1 is the assumption that a line cannot intersect an elliptic curve more than 3 times. (If this could happen, the $\oplus$ operation wouldn't be well defined!) To justify our assumption, write the equation of a line and of the elliptic curve; these are two equations in two variables, so we can solve them simultaneously to find the intersection points. When you go through this process, you'll see that this boils down to a cubic equation in x. Thus, if we're given two points with distinct x-coordinates, there's exactly one remaining solution to the cubic; this produces two points on the elliptic curve, but only one of these can be on the line through the two given points. (You should carry out this exercise with a specific elliptic curve and a specific line to get a better sense of the argument.)

Ben asked about how useful the operation $\oplus$ is to finding rational points on an elliptic curve. Starting with two given rational points P and Q, we can use the operation to generate a new rational point $P \oplus Q$, but these three points are collinear and we can't generate anything other new points; for example, $P \oplus (P \oplus Q) = Q$, a point we already knew about. Is there way a to generate more points starting with two points P and Q?

The answer is a resounding yes! First observe that we can look at $P \oplus P$ and $Q \oplus Q$; these potentially give us new points to play with, and then we can look at other combinations (e.g. $P \oplus (P \oplus P)$, etc). The other observation is that if P is a rational point, then so is $\overline{P}$, so we can look at expressions of the form $P \oplus \overline{Q}$. Thus, starting with just two rational points (or even just one!) often allows us to generate many more rational points.

On our table (adjusted to include the point at infinity among the rational points), we saw examples of elliptic curves that had $1, 2, 3, 6$, and infinitely many rational points. Mia asked whether there's any pattern or restrictions on the number of rational solutions. A remarkable theorem discovered by Barry Mazur in 1977 asserts that the number of rational points on an elliptic curve is either one of the numbers $\{1, 2, 3, ..., 16\} \smallsetminus \{11\}$ or is infinite. This inspired Jacob to deduce that once you find 17 rational points on an elliptic curve, the curve must have infinitely many rational points.

It turns out that even in the case that there are infinitely many rational points on an elliptic curve, you can generate them all by starting with some finite set and applying Theorem 1:

Theorem (Mordell, ~1920's): $\mathcal{E}(\mathbb{Q})$ is *finitely generated*, i.e. there exists a finite set of rational points such that all rational points can be obtained using only our $P \oplus Q$ method for constructing new points.

Unfortunately, even though Mordell's theorem implies the existence of a finite generating set, it's an open problem to actually determine this set for a general elliptic curve.

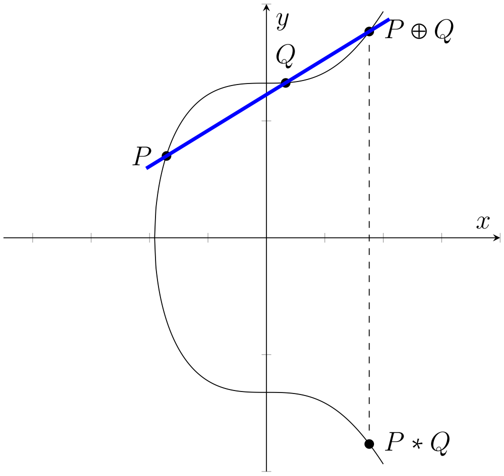
Algebraic properties of $\oplus$

Theorem 1 asserts that $\mathcal{E}(\mathbb{Q})$ is closed under $\oplus$. This is reminiscent of $\mathbb{Z}$ under $+$, and $\mathbb{Z}_n$ under $+ \pmod{n}$, and $\mathbb{Z}_n^\times$ under $\times \pmod{n}$. In these other cases, however, we have further algebraic structure, for example an identity element (0, 0, and 1, respectively).

Question: Does there exist an identity in $\mathcal{E}(\mathbb{Q})$, i.e. some $e \in \mathcal{E}(\mathbb{Q})$ such that $e \oplus P = P$ for any $P \in \mathcal{E}(\mathbb{Q})$?

Playing around a bit convinced us that while for any given P we can find an e such that $e \oplus P = P$, there's no universal identity that satisfies this property simultaneously for *all* points P. However, Alex observed that ∞ behaves *almost* like an identity: $\infty \oplus P = \overline{P}$ for all $P \in \mathcal{E}(\mathbb{Q})$. Inspired by this idea, we were led to ask whether we could tweak the definition of $\oplus$ so that ∞ is literally an identity? And we can! Consider the binary operation

$$P * Q := \overline{P \oplus Q}.$$



It's easy to check that $\infty * P = P$ for all points P, so ∞ is an identity with respect to $*$. This new binary operation enjoys some other nice algebraic properties:

Theorem: Given an elliptic curve $\mathcal{E}/\mathbb{Q}$ (i.e. $\mathcal{E}$ is of the form $y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Q}$). Then

(1) $\mathcal{E}(\mathbb{Q})$ (including ∞) is closed under $*$.

(2) $P * Q = Q * P$ ("$*$ is commutative")

(3) $P * Q * R$ is unambiguous, i.e. $(P * Q) * R = P * (Q * R)$ ("$*$ is associative")

(4) $\infty * P = P$ ("∞ is an identity with respect to $*$")

(5) $P * \overline{P} = \infty$ ("$\overline{P}$ is the inverse of P")

In the language of abstract algebra, properties (1)+(3)-(5) assert that the operation $*$ makes the set $\mathcal{E}(\mathbb{Q})$ into a group; property (2) asserts that $\mathcal{E}(\mathbb{Q})$ is an abelian group.