Lecture 25

Exploration day: Elliptic curves

25.1 Goals

1. Understand that every line intersects a cubic in "three" points, where tangencies count as two.
2. Perform some computations to get used to the group operation on an elliptic curve

25.2 Getting used to a cubic

For no reason whatsoever, I want to study the set of points (x, y) in the plane satisfying the equation

$$y^2 = x^3 - 4x + 1. \tag{25.2.0.1}$$

See Figure 25.1. I will call this set C, so

$$C := \{(x, y) \in \mathbb{R}^2 \,|\, y^2 = x^3 - 4x + 1\}.$$

C stands for "curve."

Exercise 25.2.1. Verify that the following values of (x, y) satisfy (25.2.0.1):

$$(-2, 1), \qquad (-1, 2), \qquad (0, 1), \qquad (2, 1), \qquad (3, 4).$$

Exercise 25.2.2. (a) If a pair (x, y) satisfies (25.2.0.1), show that the pair $(x, -y)$ does, too.

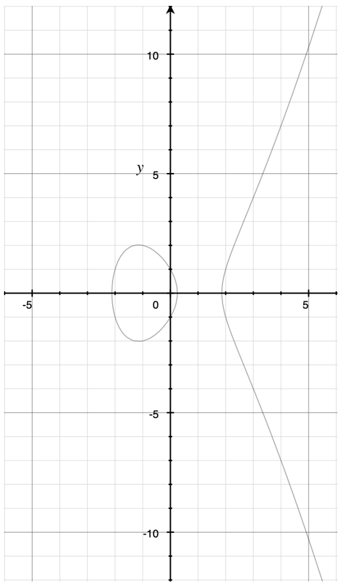


Figure 25.1: $y^2 = x^3 - 4x + 1$

(b) Aside from the pairs in Exercise 25.2.1 and their "negative y" versions, can you find more pairs (x, y) satisfying (25.2.0.1) where both x and y are integers? (This is hard, try only for a couple minutes.)

(c) How about where both x and y are rational numbers? (This is also hard.)

Exercise 25.2.3. Consider the two points

$$P = (-2, 1) \qquad \text{and} \qquad Q = (-1, 2).$$

(a) Find the equation of the line L passing through P and Q.

(b) It turns out that L intersects the curve C at a third point. (See Figure 25.2.) Find the coordinates of this third point. We will call this third point X.

Hint: Once you find your equation for L, and can in particular write y as a function of x, the x-coordinates of the intersection $L \cap C$ can be found by substituting the expression for y into (25.2.0.1)*. This becomes a cubic polynomial in x, but you already know two of the roots, so you can factor to find the third root.*

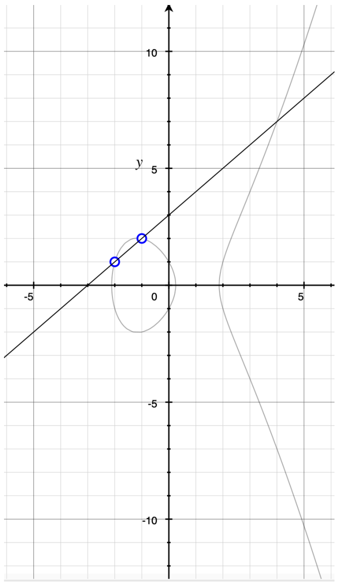


Figure 25.2: Points P and Q on $y^2 = x^3 - 4x + 1$, and the line through them (Exercise 25.2.3).

Then, you can compute the y-coordinate using the equation for L.

(c) Are the coordinates of X rational?

25.3 “Adding” distinct points on C

Given Exercise 25.2.3, we seem to have a way to take two elements of C, and output a third. Is this starting to feel like a binary operation? Good. But there’s a twist.

Notation 25.3.1 (Addition on C)**.** For no reason whatsoever, flip the sign of the y-coordinate of X. We will call this point $-X$. We define

$$P + Q := -X.$$

In words: $P + Q$ is a point on C obtained by (i) drawing the line L through P and Q, (ii) finding the third point X in the intersection of L with C, and (iii) declaring $P + Q$ to be the point across the x-axis from X.

This $P+Q$ makes sense for any two distinct points on C. I am going to eventually claim that this + operation on points of C turns out to define a *group* operation!

Exercise 25.3.2. When $P \neq Q$, explain why $P+Q=Q+P$.
(Hint: You should not have to perform any computations.)

Exercise 25.3.3 (Visualization and playing)**.** Here is a desmos website where you can play with adding different P and Q on C:

`https://www.desmos.com/calculator/kanuniawpd`

Because of my technological deficiencies, you can only choose P and Q with positive y-coordinate.

(a) Verify that, for almost all choices of P and Q, the line L through P and Q passes through a third point.

(b) Though the Desmos demo doesn't allow for it, what if P and Q lie on a vertical line?

(c) Can you find an example of a line that passes through only two points and is *not* vertical? (Make sure that your example is a valid one—don't forget to Zoom out!) What geometric property does such a line have? (After a few minutes, for the sake of time, you can give up on this if you can't find such a line.)

25.3.1 "Adding" a point to itself

To really have a binary operation on C, we have to fix a few things. First, what if $P=Q$–what do we mean by the "line through P and Q?" Second, what if a line doesn't pass through three points?

Exercise 25.3.4. Let $P=(-2,1)$ be the point from before.

(a) Find the line L that is tangent to C at P. (Hint: Implicit differentiation. See Figure 25.3)

(b) It turns out L intersects C at exactly one more point. Find the coordinates of this point.

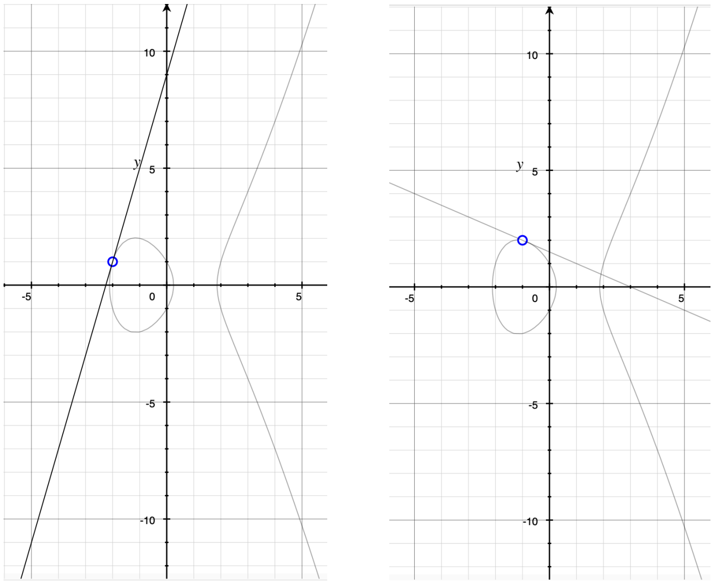


Figure 25.3: The tangent line at the point P to $y^2 = x^3 - 4x + 1$ (Exercise 25.3.4) Though the intersection point does not fit on this window, the tangent line does intersect C at point with large y coordinate. Drawn also is the tangent to Q.

I had you do the above exercise, because it tells you how to define $P + P$ in general:

Definition 25.3.5. Fix a point $P \in C$. Let L be the tangent line to C at P, and let X be the other point[1] in $L \cap C$. We define

$$P + P = -X.$$

(That is, we again flip the y-coordinate of X to obtain $P + P$.)

Remark 25.3.6. Now, it turns out C does not have an identity element; we'll fix this later, but I don't have a fun exercise for you illustrating this, so you can read about the identity in Section 25.5.

After fixing the situation to obtain an identity element, it turns out that the inverse of any point $P = (x, y)$ is the point $-P = (x, -y)$.

For now, you can just take for granted that there's some way we can make $+$ and C into a group (after adding an identity element).

25.4 The subgroup of rational points

Exercise 25.4.1. Suppose P and Q are arbitrary points on C for which the coordinates of P and the coordinates of Q are all rational numbers.

Show that $P + Q$ is another point whose coordinates are all rational numbers.

While we omit the details, the exercise above is the main step toward proving that the subset of C consisting of only rational points is a subgroup (after adjoining the same additive identity as C).

For more on this, see Section 25.8.

25.5 The actual definition of the elliptic curve and the group operation

That's the end of the exploratory exercises for today. The rest are just notes for your perusal.

[1] When the tangent line is vertical, there is no such other point! This issue will be resolved when we introduce the identity element.

I mentioned earlier that C doesn't have an identity element. We're going to do something funny to fix this:

Definition 25.5.1. We let $\{*\}$ be a one-element set, and we will informally call its element $*$ "the point at infinity."

We define the *elliptic curve* associated to the equation (25.2.0.1) as the set

$$E = \{*\} \bigcup C$$

given by adjoining $*$ to C. (So the only difference between E and C is the one element $*$.)

Now we are ready to define addition on all of E:

Definition 25.5.2. We define addition on E as follows:

(1) If $P, Q \in C$ and $P \neq Q$, then $P + Q$ is as before (Notation 25.3.1).

(2) If $P \in C$ and the tangent line to C at P is not vertical, we define $P + P$ as before (Definition 25.3.5).

(3) If $P \in C$ and the tangent line to C at P is vertical, we define $P + P$ to be $*$.

(4) For any $P \in C$, we declare

$$P + * = P \qquad \text{and} \qquad * + P = P,$$

and

(5) $* + * = *$.

We will not prove the following. The only hairy part is the verification that $+$ is associative:

Theorem 25.5.3. *E, with addition defined as above, is an abelian group.*

Remark 25.5.4. This is of course highly non-trivial—not every binary operation has an identity, nor admits inverses, nor is associative. (Consider the example of cross product in $\mathbb{R}^3$.)

Moreover, how crazy is it that a wonky cubic curve like C should have an addition that has inverses and is associative?

Remark 25.5.5. If you want addition to feel "continuous" then it is inevitable that $P + P$ depends on the tangent line at P. After all, $P + Q$ depends on the secant line through P and Q and (as we know from calculus) as Q approaches P, this secant line approaches the tangent line at P.

25.6 Modern-day applications

E is an example of an abelian group called an *elliptic curve*. I hope you'll find elliptic curves as mysterious as I did when I first learned about them—I had no idea *why* an elliptic curve would be a group, and I could only understand that it simply *was* a group. Very dissatisfying, but very enticing.[2]

Aside from their mystery, elliptic curves have a utility. It turns out that given an element P of an elliptic curve, it is *easy* to compute $P + \ldots + P$. But it is *hard*, given only the data of P and the end result $Q = P + \ldots + P$, how *many* times P was added to itself to attain Q. In other words, it is easy to take n, P and compute $nP = Q$. It is hard to take P and Q and determine what n is.

This fact is at the heart of a lot of modern cryptography. Now-a-days, the security of certain verification systems relies on the difficulty of finding n when someone gives you P and Q.

25.7 Elliptic curves generally

In high school, you studied quadratic polynomials in x and y, and studied their zero locus. For example, you know that

$$x^2 - y^2 - 1 = 0$$

[2] That they become a group is less mysterious once you learn that elliptic curves are also Jacobian varieties, but one needs a great deal of education to understand what this means.

is a hyperbola. And more generally, if you have some reasonable equation consisting of quadratics:

$$x^2 + 2xy - y^2 = 9 - y - x$$

you may know how to apply a coordinate transformation (rotating by an angle determined by the coefficients of a quadratic polynomial), and potentially complete the square, to observe whether the curve is a hyperbola, parabola, or an ellipse.

Let's start today by moving to *cubics*, but a very simply kind of family. First, we'll demand that the only term involving y will always be y^2. Next, we'll demand that x will have no quadratic term, and that x^3 and y^2 have the same coefficient (which may be assumed to be 1 after division). More succinctly, we're only going to study cubic equations that have the following form:

$$y^2 = x^3 + ax + b \tag{25.7.0.1}$$

where a and b are real numbers that we are free to choose. Fixing a and b, the set of all points (x, y) satisfying the above equation forms a curve in the plane. Each choice of $a, b \in \mathbb{R}$ gives rise to a new cubic, hence a new curve in the plane.

Remark 25.7.1. Clearly (25.7.0.1) is only a very particular kind of cubic equation, but it turns out that almost all cubics can be made to take on this form after some clever coordinate transformations. If you're interested in further reading, you can Google "Weierstrass form."

Remark 25.7.2 (More visualization)**.** Here is a desmos website where you can play with changing a and b:

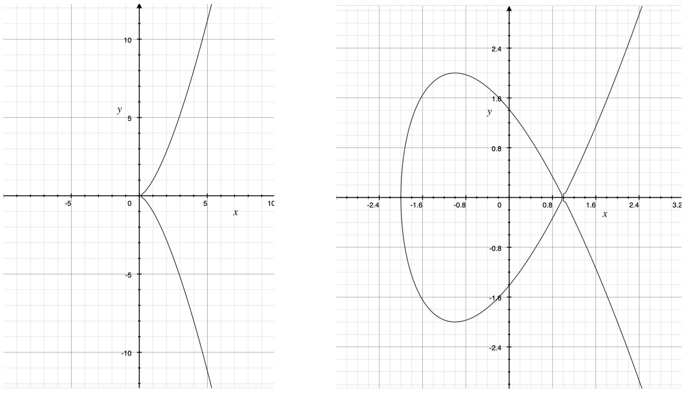
`https://www.desmos.com/calculator/kanuniawpd`

When $a = b = 0$, the curve develops a cusp at the origin (the curve looks point there) so is not "smooth."

More generally, when $4a^3 - 27b^2 = 0$, it turns out that $E_{a,b}$ will have a singular shape (for example, crossing over itself at a node, or having a cusp), and is again not smooth. Try the example $a = -3$ and $b = 2$.

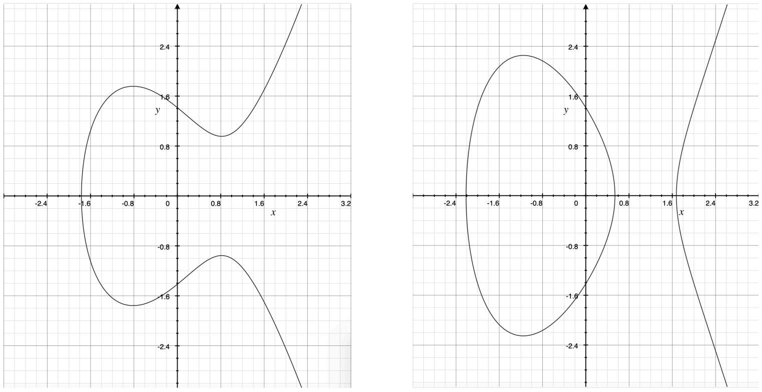
For these singular examples, see Figure 25.7.

For non-singular examples, see Figure 25.7.



$y^2 = x^3$ $\qquad$ $y^2 = x^3 - 3x + 2$

Figure 25.4: Examples of two singular cubics.



$y^2 = x^3 - 2x + 2$ $\qquad$ $y^2 = x^3 - 4x + 2$

Figure 25.5: Examples of two non-singular cubics in Weierstrass form.

Definition 25.7.3. Fix two real numbers a, b for which the graph of (25.7.0.1) is non-singular[3]. Then the associated *elliptic curve* is the set

$$E = \{\infty\} \bigcup \{(x, y) \,|\, y^2 = x^3 + ax + b\}.$$

In words, an elliptic curve has two kinds of points:

(i) Points of the form (x, y) satisfying $y^2 = x^3 + ax + b$, and

(ii) A "point at infinity" that we just write as "∞." To remove the mystery, let me just tell you write now that the points (x, y) come very close to forming a group, but will not have an identity element. If you like, you can imagine that we articificially add on an extra element called ∞ to introduce an identity element to the group.

When we want to make the dependence on a, b explicit, we will write $E_{a,b}$ instead of E.

Remark 25.7.4. You will not lose too much intuition by thinking of E just as the curve in $\mathbb{R}^2$ determined by $y^2 = x^3 + ax + b$. Indeed, by removing just one point from E (the point a infinity) we are left precisely with this curve.

We can define addition on E exactly as in Definition 25.5.2. It then turns out that any elliptic curve is a group.

25.8 The subgroup $E_{\mathbb{Q}}$ and the Mordell-Weil theorem

Proposition 25.8.1. Let E be an elliptic curve defined by a cubic with rational coefficients. Let $E_{\mathbb{Q}} \subset E$ be the set of all points that (i) have only rational coordinates, or (ii) are the identity (the point at infinity). Then $E_{\mathbb{Q}}$ is a subgroup of E.

Remark 25.8.2. In general, E is some crazy, uncountable group. But it turns out that $E_{\mathbb{Q}}$ is a subgroup that can be generated by just finitely many elements. (This is a major theorem in number theory called the Mordell-Weil theorem.)

[3] Let me put in this footnote so you don't miss it. We are excluding choices of a and b that result in cusps, or in self-crossing nodes. By definition, an elliptic curve has no cusps and has no nodes.