2 Elliptic Curves

2.1 Cubic Curves

Recall from our first class that finding solutions to the Diophantine equation

$$X^2 + Y^2 = Z^2, \quad X, Y, Z \in \mathbb{Z}$$

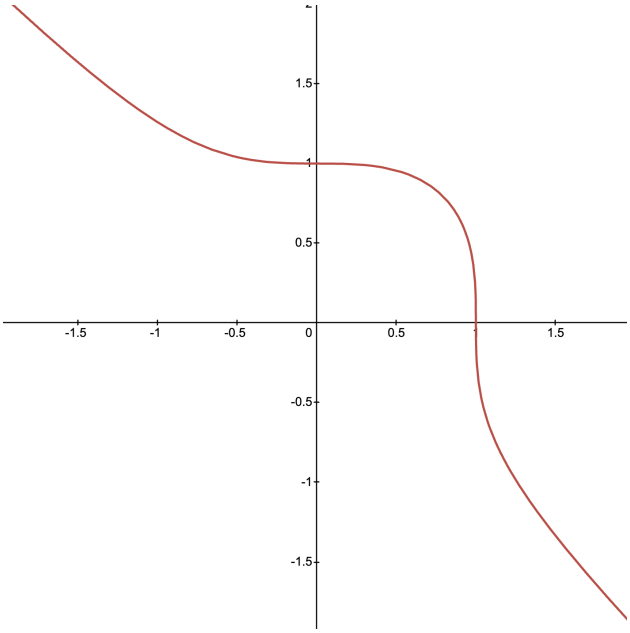
was equivalent to finding rational points $(x, y) \in \mathbb{Q}^2$ that lie on the unit circle $x^2 + y^2 = 1$. Now say instead we want to find solutions to the following Diophantine equation:

$$X^3 + Y^3 = Z^3, \quad X, Y, Z \in \mathbb{Z}.$$

Just as we did in the case of pythagorean triples, we can divide by Z (assuming $Z \neq 0$) and get a curve

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

where $x = X/Z$ and $y = Y/Z$. We'll call this curve the *Fermat cubic*. Now we're interested in finding the rational points on this cubic curve, who's graph looks like the one below.



What are the rational points on the Fermat cubic? By quick inspection, we see that $(1, 0)$ and $(0, 1)$ are both on the curve. It turns out that, unlike the case with the circle, these are the *only* rational points on this cubic curve! This means that the equation

$$X^3 + Y^3 = Z^3, \quad X, Y, Z \in \mathbb{Z}$$

has no solutions at all, unless X, Y, or Z is zero. This is a corollary of the famous *Fermat's last theorem*, which says that there are no positive integer solutions to the equation

$$X^n + Y^n = Z^n, \quad n \geq 3.$$

In one of the most famous (and frustrating) stories of mathematics, Pierre de Fermat in 1637 claimed that he "discovered a truly marvelous proof of this [theorem], which however the margin is not large enough to contain." I wish I could write that in my math exams! However, this theorem proved to be one of the largest challenges of modern mathematics, and it was only proven 350 years later by Andrew Wiles in 1995. Wiles's proof relied on very advanced techniques from the theory of *elliptic curves*.

We won't be able to talk about Wiles's proof (I don't understand it), but let's go back to answering why the Fermat cubic doesn't have any rational points outside of $(1, 0)$ and $(0, 1)$. We'll first carry out a series of transformations that makes the equation of the Fermat cubic slightly simpler.

Let's rewrite our Fermat cubic in coordinates u, v as

$$u^3 + v^3 = 1.$$

Now we introduce new coordinates x, y, related to the old coordinates u, v via the rational functions

$$u = \frac{36 + y}{6x}, \quad v = \frac{36 - y}{6x}.$$

Substituting to our equation for the Fermat cubic in (u, v), we find

$$\begin{aligned} \left(\frac{36 + y}{6x}\right)^3 + \left(\frac{36 - y}{6x}\right)^3 &= 1 \\ \frac{216y^2 + 93312}{216x^3} &= 1 \\ y^2 &= x^3 - 432. \end{aligned}$$

This curve $y^2 = x^3 - 432$ is called a *Weierstrass cubic*. Now since u and v can be expressed as rational functions of x and y, any rational point $(u, v) \in \mathbb{Q}^2$ is mapped to a rational point $(x, y) \in \mathbb{Q}^2$ under the transformation

$$u^3 + v^3 = 1 \quad \rightsquigarrow \quad y^2 = x^3 - 432.$$

We also find that the points $(1, 0)$ and $(0, 1)$ on the Fermat cubic correspond to the following points on the Weierstrass cubic

$$(1, 0) \rightsquigarrow (12, 36), \quad (0, 1) \rightsquigarrow (12, -36).$$

Therefore, to show that there are no rational points $(u, v) \in \mathbb{Q}^2$ on the curve $u^3 + v^3 = 1$ aside from $(1, 0)$ and $(0, 1)$, we can equivalently show that there are no rational points $(x, y) \in \mathbb{Q}^2$ on the curve $y^2 = x^3 - 432$ aside from $(12, \pm 36)$.

Now we're ready for the definition of an elliptic curve.

Definition 1 (Elliptic Curve)**.** An *elliptic curve* is a curve of the form

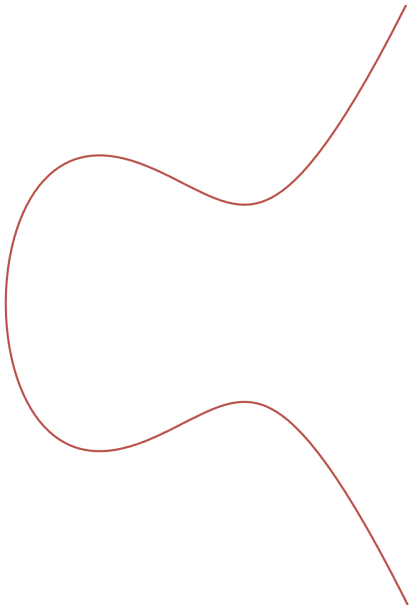
$$y^2 = x^3 + ax^2 + bx + c,$$

where a, b, c are rational numbers.

2.2 The Group Structure of an Elliptic Curve

The most interesting and useful property of elliptic curves is that they are a *group*. Let's explore this claim a little further.

A typical elliptic curve E looks like the following.



How can we define a group, whose elements are the points of this curve? That is, starting with two points $P, Q \in E$, can we find some group operation $*$ that lets us "add" $P * Q$?

We can make the following geometric observation.

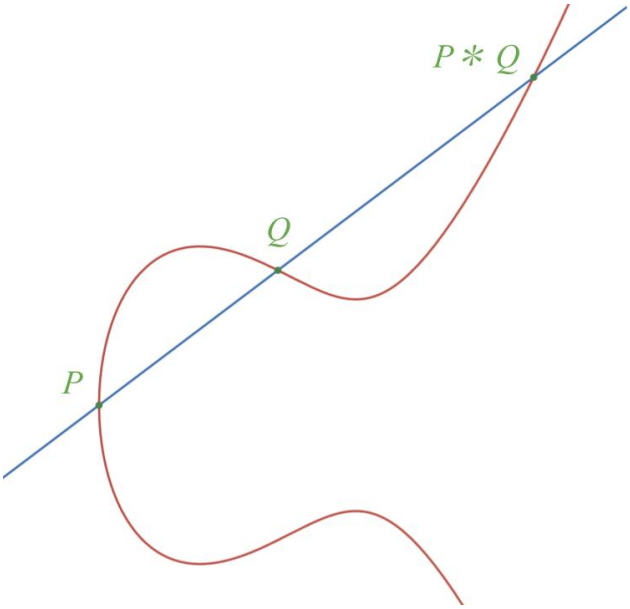
Proposition. *Let P, Q be two distinct points on E with distinct x-coordinates, and let L be the line $\overline{PQ}$. Then the line L always intersects E at three points: P, Q, and a third point $P * Q$.*

Solution. The x-coordinates of P and Q are distinct, so the equation of line L is given by $y = \lambda x + \mu$ with real coefficients. The intersection of E and L has x-coordinates satisfying

$$(\lambda x + \mu)^2 = x^3 + ax^2 + bx + c.$$

This equation looks complicated, but it's just a cubic equation; and we know that if all coefficients of a cubic equation are real, then it has either 1 or 3 solutions. Since L already intersects E at the points P, Q, it follows that L must intersect E at a third point, which we call $P * Q$. □

We now have a binary operation $*$ that takes two points P, Q and produces a third point $P * Q$.



Does $*$ define a group on E? After some inspection, we find that unfortunately $(E, *)$ is not a group: for one thing, there is no point on E that can serve as the identity.

Furthermore, the operation $*$ is not defined for two points P and Q with the same x-value. If this were the case, then the line L is given by the equation $x = \mu$ for some constant $\mu \in \mathbb{R}$. Thus, the line L only intersects E at two points with y-values equal to

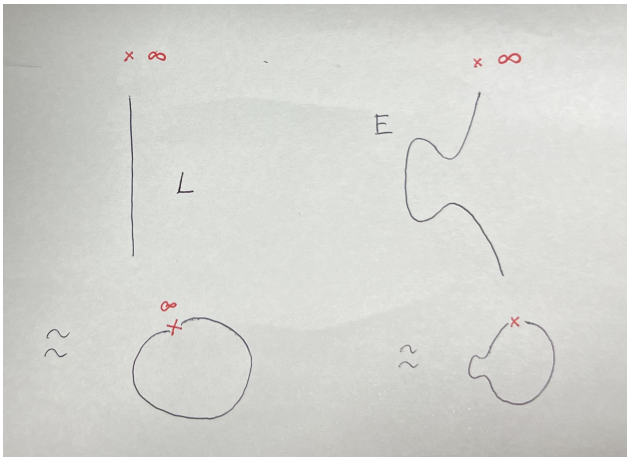
$$y^2 = \mu^3 + a\mu^2 + b\mu + c \quad \rightsquigarrow \quad y = \pm\sqrt{\mu^3 + a\mu^2 + b\mu + c}.$$

We will patch this "incompleteness" of our operation very soon.

Instead, let's take a step back and consider a different question. What point on E could be the identity of our group? This is a bit of a trick question, because we are actually going to introduce a brand new point to E which will serve as our identity!

Currently, the curve E extends infinitely towards the positive and negative y directions. This is unlike the geometry of a circle. Mathematicians much prefer to work with shapes like circles, which are called *compact*. The upshot is that we can make the E compact by adding a point called the "point at infinity."

Given a line L, we imagine the *point at infinity* as a point that exists at "both ends" of the line L. Importantly, we consider both ends of L to be the same point, so that L and the point at infinity create a loop. For the elliptic curve E, we add a point infinitely away towards the positive (or equivalently, negative) y direction. Let's call this point $\mathcal{O}$.



Now we're ready to introduce the group law. It turns out that our guess with $*$ was close—we just need to add one more step. Since an elliptic curve has equation

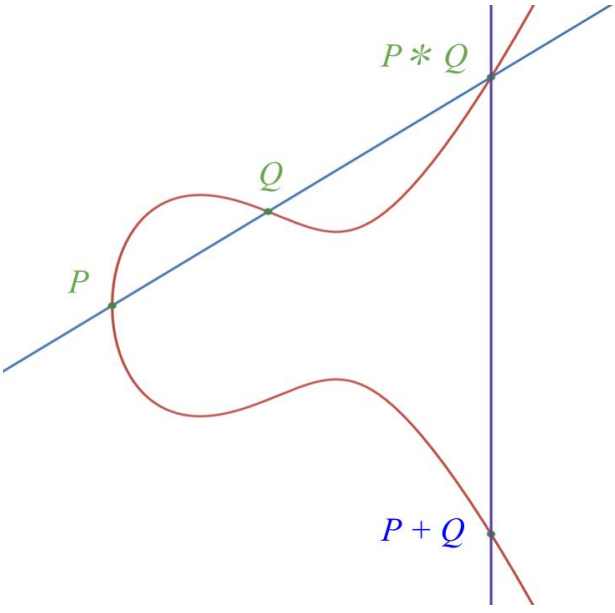
$$E : y^2 = x^3 + ax^2 + bx + c,$$

we notice that if (x, y) is in E, then $(x, -y)$ is also in E. So if $P = (x, y) \in E$, let $-P = (x, -y)$. We define a group operation $+$ on E as

$$P + Q := -(P * Q).$$

That is, $P + Q$ is the point obtained from $P, Q \in E$ by

1. drawing a line through P and Q,
2. taking the third point of intersection between the line and E,
3. and reflecting that point across the x-axis.



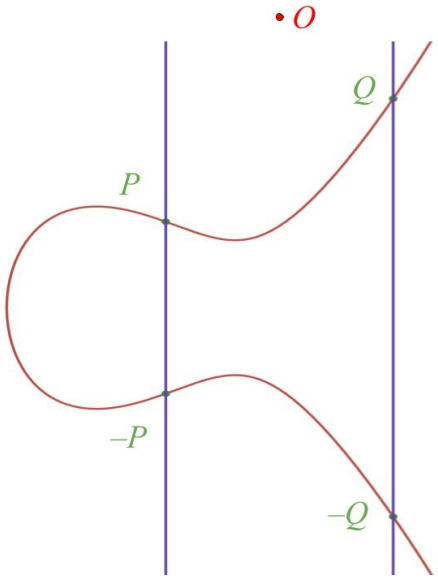
Now let's verify that $(E, +)$ indeed satisfies the group axioms.

First, let's try to understand what it means for the point at infinity $\mathcal{O}$ to be the identity under $+$. We said that the point $\mathcal{O}$ exists infinitely away in the y-direction. If we take any point $P \in E$ and draw a line "through" P and $\mathcal{O}$, we thus get a vertical line. So the line connecting P to $\mathcal{O}$ intersects E at the third point $-P$.

In other words, we've shown that

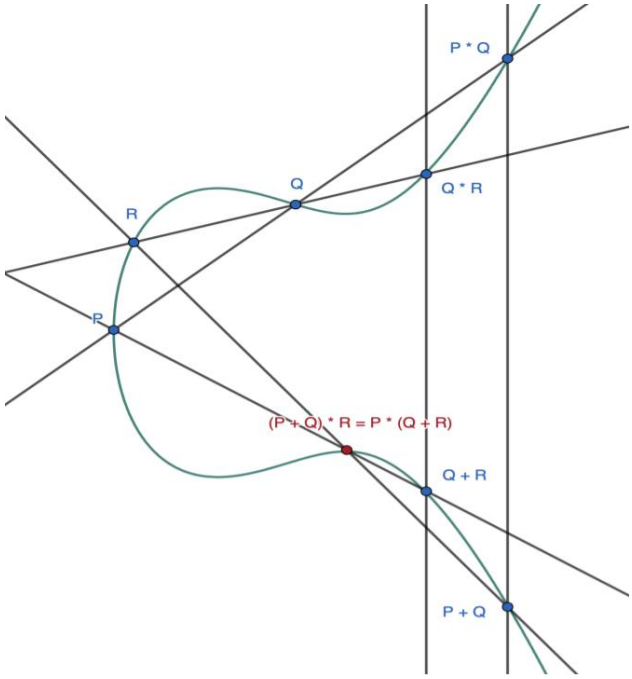
$$P + \mathcal{O} = -(P * \mathcal{O}) = -(-P) = P,$$

and so $\mathcal{O}$ is the identity on $(E, +)$.



We've also basically proved that every element in E has an inverse. If P is a point on E, then its inverse is just its reflection $-P$ across the x-axis.

Finally, we need to show that $+$ is associative. This is a little annoying, so we don't go into details, but you can convince yourself by considering the following diagram.



We'll devote the remainder of this seminar to understanding this (abstract) group. One property we immediately notice is that $(E, +)$ is an *abelian* group: that is, $P + Q = Q + P$. This is because $P * Q$ and $Q * P$ are the same point.

2.3 Explicit formulas for the Group Law

In the previous section, we described the group law $(E, +)$ purely in terms of geometry. Say P and Q are points on E, and let

$$P = (x_1, y_1), \quad Q = (x_2, y_2), \quad P * Q = (x_3, y_3), \quad P + Q = (x_3, -y_3).$$

Let's try to find a formula for x_3 and y_3 in terms of x_1, y_1, x_2, y_2.

The line joining P and Q is given by the equation

$$y = \lambda x + \nu, \qquad \left(\lambda = \frac{y_2 - y_1}{x_2 - x_1}, \quad \nu = y_1 - \lambda x_1\right).$$

Since P, Q and $P * Q$ all (i) lie on this line and (ii) lie on E, the x-coordinates x_1, x_2, and x_3 are solutions to the cubic equation

$$\begin{aligned}(\lambda x + \nu)^2 &= x^3 + ax^2 + bx + c \\ 0 &= x^3 + (a - \lambda^2)x^2 + (b - 2\lambda\nu)x + c - \nu^2.\end{aligned}$$

Since the coefficient on the x^2 term is $a - \lambda^2$, we know

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

Therefore, our formula for x_3 is

$$x_3 = \lambda^2 - a - x_1 - x_2,$$

and it follows that our formula for y_3 is

$$\begin{aligned} y_3 &= \lambda x_3 + \nu \\ &= \lambda(\lambda^2 - a - x_1 - x_2) + \nu. \end{aligned}$$

In other words, we can describe our group law on E as

$$(x_1, y_1) + (x_2, y_2) = (\lambda^2 - a - x_1 - x_2, -\lambda(\lambda^2 - a - x_1 - x_2) - \nu). \qquad (\star)$$

An important observation we make from $(\star)$ is that the formulas for x_3 and y_3 are both rational functions of x_1, y_1, x_2, y_2. In particular, the formula shows that if (x_1, y_1) and (x_2, y_2) are both rational points on E, then λ, ν, and a are all rational, and so the point $(x_3, -y_3)$ is another rational point of E. We've proven that, starting with two rational points $P, Q \in E$, our group law produces a third rational point $P + Q$. Using terminology from the previous unit, this says that the rational points on our elliptic curve, defined

$$E(\mathbb{Q}) = \{(x, y) \in \mathbb{Q}^2 : y^2 = x^3 + ax^2 + bx + c\}$$

forms a *subgroup* of the real points on our elliptic curve $E(\mathbb{R})$, defined

$$E(\mathbb{R}) = \{(x, y) \in \mathbb{R}^2 : y^2 = x^3 + ax^2 + bx + c\}.$$

Or succinctly,

$$E(\mathbb{Q}) \leq E(\mathbb{R}).$$

To understand $E(\mathbb{Q})$, it will be very helpful for us to consider not just the properties of $E(\mathbb{R})$, but also the properties of $E(\mathbb{C})$, the group of *complex points* on an elliptic curve. But what does this mean? As a set, we can define $E(\mathbb{C})$ analogously as

$$E(\mathbb{C}) = \{(x, y) \in \mathbb{C}^2 : y^2 = x^3 + ax^2 + bx + c\}.$$

For example, if E is defined by the curve $y^2 = x^3 + 1$, then the point $(-\sqrt[3]{2}, i)$ would be on $E(\mathbb{C})$. We can't draw the graph of $E(\mathbb{C})$ like we did for $E(\mathbb{R})$, since x and y are now both complex numbers, and thus the "graph" of $E(\mathbb{C})$ inhabits four dimensional space.

However, the group law $(\star)$ works perfectly fine when (x_1, y_1) and (x_2, y_2) are complex coordinates. Therefore, we can make $E(\mathbb{C})$ into a group who's addition operation is defined by formula $(\star)$. We immediately observe that $E(\mathbb{C})$ is a larger group than $E(\mathbb{R})$. Succinctly,

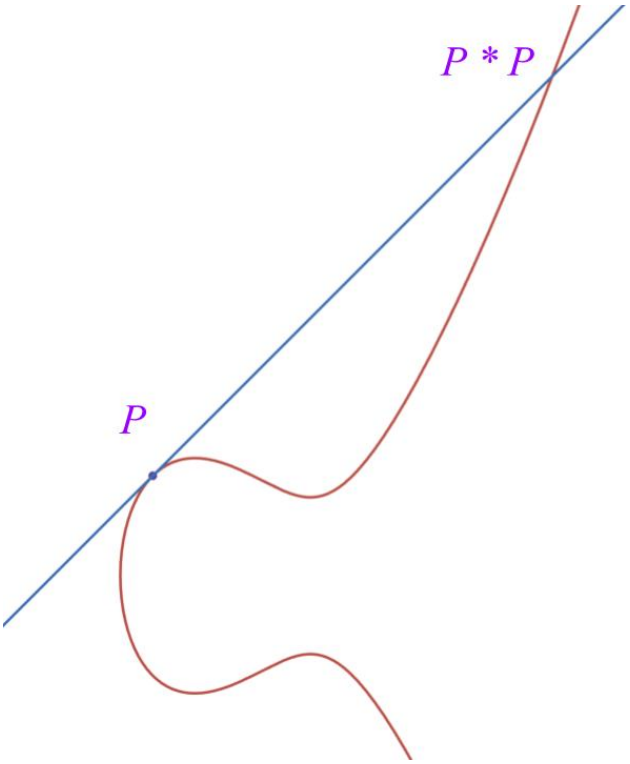
$$E(\mathbb{Q}) \leq E(\mathbb{R}) \leq E(\mathbb{C}).$$

Although we can't draw the "graph" of $E(\mathbb{C})$, mathematicians have devised many clever ways to visualize these high-dimensional objects. In the next lecture, we'll get a glimpse of these techniques, and study the remarkable geometric properties of elliptic curves.

2.4 Multiplying Points of E

Given two distinct point $P, Q \in E$, we've described how to find their sum $P + Q$ on E. How can we double a point on E, i.e. find $2P$ when given P?

As a point Q comes close to P, the line through P and Q approaches the tangent line to E at P. Thus, to multiply a point P by two we can take the *tangent line* to E at P.



Now we know how to triple a point P in E, and in general how to multiply P by any positive integer n, since

$$3P = P + 2P, \quad nP = P + (n-1)P.$$

3 Topology

3.1 What is topology?

Last time, we ended with the group of complex points of an elliptic curve $E(\mathbb{C})$, defined

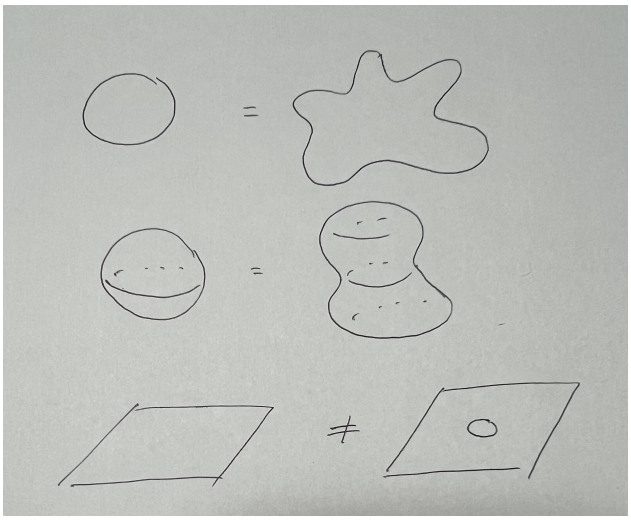
$$E(\mathbb{C}) = \{(x, y) \in \mathbb{C}^2 : y^2 = x^3 + ax^2 + bx + c\}.$$

Unfortunately, we aren't able to "graph" this function, since x and y are both complex numbers: the space $\mathbb{C}^2$ is four-dimensional.

However, it will still be very useful to somehow intuitively understand the geometry of these complex elliptic curves. The tool that will allow us to do this is called topology.

Very roughly speaking, *topology* studies qualitative aspects of spaces, whereas *geometry* studies its quantitative aspects. Questions like "what is the size of this angle?" or "what is the area of this surface?" belong to the realm of geometry, while questions like "how many holes does this shape have?" or "is this shape connected?" belong to the realm of topology.

In particular, topology doesn't care about the precise shape of an object, and two deformations of the same objects are treated as topologically equivalent (called *homeomorphic*). By deformation, we allow bending and stretching, but do not allow puncturing holes or ripping shapes apart.



Let's investigate an example. Consider the x-axis in the xy-plane $\mathbb{R}^2$, defined

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

When we add a point at infinity to this line (at a place an infinite distance away in the x-direction), we said last time that the x-axis behaves like a circle. Indeed, we say that the x-axis is *topologically the same* as a circle. We've actually already seen this homeomorphism in lecture 1, when we projected the points of a circle onto a line!

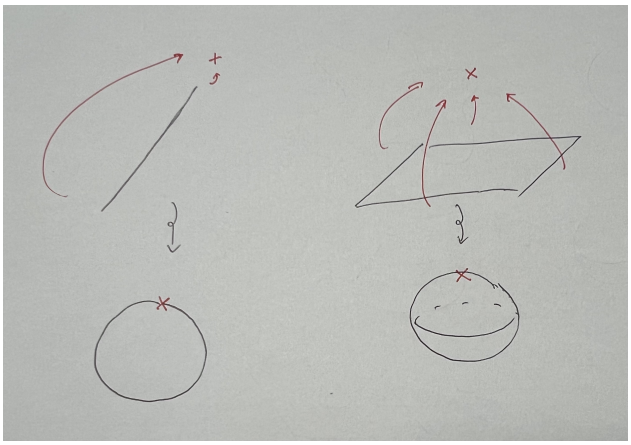
In fact, this shows that any line together with its corresponding point at infinity is topologically the same as the circle.

Now let's consider the analogous space L in $\mathbb{C}^2$, defined

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

Just like the x-axis is a copy of $\mathbb{R}$ inside $\mathbb{R}^2$, the space L is a copy of $\mathbb{C}$ inside $\mathbb{C}^2$. In other words, since we can treat the set of complex numbers as a plane, the space L looks like a plane embedded inside the 4-dimensional space $\mathbb{C}^2$.

Now let's consider what L looks like when we add a point of infinity. Recall that, in the case of a line in $\mathbb{R}^2$, we imagined the point at infinity to exist at the "ends" of the line. Similarly, we'll imagine the point of infinity of the plane L to exist at the "ends," or the "edges," of the plane. Once we identify the "edges" of L as one point, the outcome is topologically equivalent to a sphere—you can think of it like "unpeeling an orange."



3.2 Topology of Curves

Now that we understand that we can visualize lines in $\mathbb{C}^2$ as spheres, let's proceed to investigate the topology of *curves*.

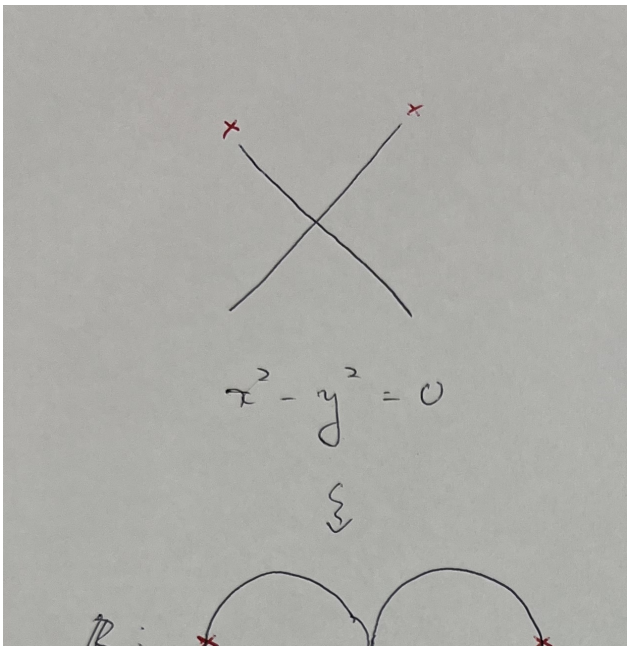
For example, let's take the quadratic curve C_1 defined by the equation $x^2 - y^2 = 1$, and its close neighbor C_0 defined $x^2 - y^2 = 0$. The former is a hyperbola in $\mathbb{R}^2$, whereas the latter is a pair of two lines, since we have

$$x^2 - y^2 = (x + y)(x - y) = 0.$$

Note that we can think of C_1 as a slight deformation of C_0.

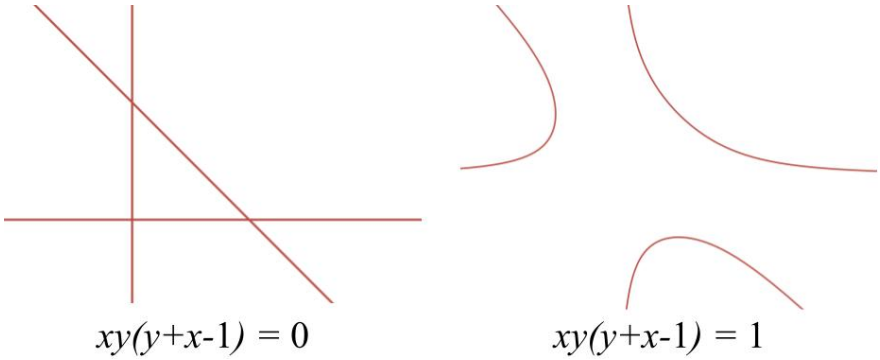
When we add "points at infinity" to $\mathbb{R}^2$, the pair of lines C_0 becomes a pair of circle intersecting at one point. On the other hand, the hyperbola C_1 becomes connected at the points at infinity, and the resulting topology is that of a deformed circle.

We also see this deformation when we try to visualize these two curves inside $\mathbb{C}^2$. The curve C_0 is now a pair of spheres, which intersect at one point. The curve C_1 is also a pair of spheres, but their intersection is deformed, which we can visualize as the two spheres being glued together. Topologically, C_1 is equivalent to a single sphere.



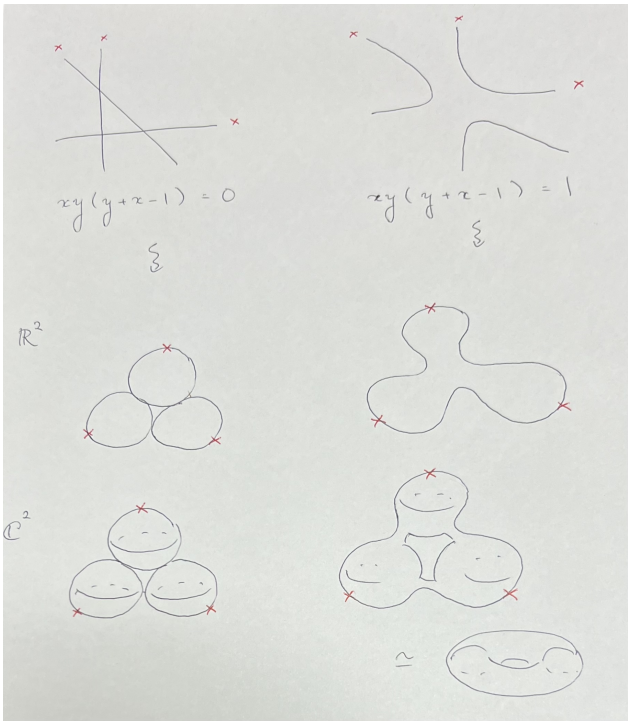
How about for curves of degree 3? Let's consider the curves C_0 and C_1 defined

$$C_0 : xy(y + x - 1) = 0$$
$$C_1 : xy(y + x - 1) = 1.$$



When we add three "points at infinity" to $\mathbb{R}^2$, the curve C_0 becomes three circles, each intersecting one another at a point. The curve C_1 is a deformation of these three circles, forming one "big" circle.

However, the picture is quite different when we look at $\mathbb{C}^2$. Again, the three lines in C_0 are represented by three spheres, each pair of sphere touching at one point. The curve C_1 is a deformation of these three spheres: the deformation is topologically equivalent to a shape called a *torus*, which looks like a (hollow) donut.



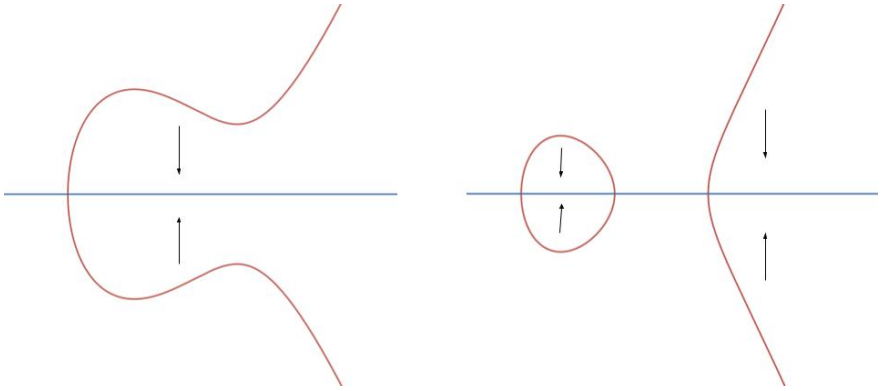
Hopefully this convinces you that cubic curves in $\mathbb{C}^2$ are topologically a torus.

In particular, recall the set of complex points on an elliptic curve $E(\mathbb{C})$ defined

$$E(\mathbb{C}) = \{(x, y) \in \mathbb{C}^2 : y^2 = x^3 + ax^2 + bx + c\}.$$

Since the curve $y^2 = x^3 + ax^2 + bx + c$ is a cubic curve, it follows that the set $E(\mathbb{C})$ can be topologically visualized as a torus!

Let's examine another way to show why the space $E(\mathbb{C})$ looks like a torus. Consider the function f that projects the elliptic curve E down onto the x-axis. Over the real numbers, we see that f takes two points of $E(\mathbb{R})$ to one point of the x-axis, except for the point where E meets the axis. This is similar to how the function $g(x) = x^2$ maps both $\pm a$ to a^2, except for at $a = 0$.



What does this function look like in $\mathbb{C}^2$? Even in the complex numbers, the map f sends

$$(x, \sqrt{x^3 + ax^2 + bx + c}) \mapsto x \quad \text{and} \quad (x, -\sqrt{x^3 + ax^2 + bx + c}) \mapsto x,$$

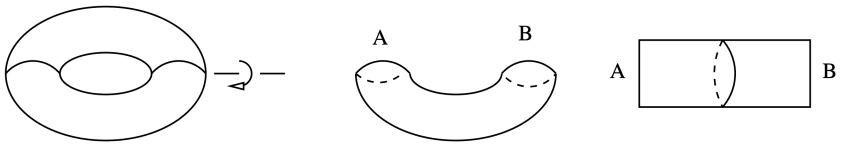
so in most cases takes two points of $E(\mathbb{C})$ and maps it to one point of the plane $x = 0$. The only exceptions are

- where $y = 0$, which corresponds to the three roots of

$$x^3 + ax^2 + bx + c = 0.$$

- the point at infinity $\mathcal{O}$.

We thus call f a *double cover* of the x-plane, *branched* at four points (the three roots of $x^3 + ax^2 + bx + c$ and at the point at infinity). It turns out that, since the x-plane is topologically a sphere, the only way for this to work is for $E(\mathbb{C})$ to be a torus! Here's a visualization.



Imagine skewering the torus with a long stick (maybe you're at a festival... though I don't think I've ever seen doughnuts on skewers before). The map f can be visualized as "folding over" a torus by rotating half of the torus by 180 degrees along this stick. For most points, the map f thus takes two opposite points on the torus and puts them together. The exception is at the four points where the stick meets the torus, which become the four "corners" of the sphere pictured above. The picture looks like a pillowcase, but a pillowcase is topologically equivalent to a sphere.