1 Introduction

Depending how you look at them, elliptic curves can be deceptively simple. Using one of the easier definitions, we are just looking at points (x, y) that satisfy a cubic equation, something like $y^2 = x^3 + 1$. This is barely more complicated than looking at points that satisfy a quadratic equation, like a circle or a parabola. This is something covered in any precalculus course.

However, when we move up from quadratic equations to cubic equations, we open up an incredible wealth of new mathematics. This is based mainly on three main facts: (1) elliptic curves show up in arclength formulas for ellipses, (2) we can put an interesting group structure on an elliptic curve, and (3) we can find an isomorphism to a doubly periodic map on the complex numbers. With these observations, elliptic curves have a prominent role in algebraic geometry, algebraic number theory, arithmetic number theory, cryptography, function theory, physics, and complex analysis.

Elliptic curves show up in the arclength formula of ellipses and the true description of a swinging pendulum. They were used prominently in the proof of Fermat's Last Theorem, once the most famous unsolved problem in mathematics. One of the seven Millennium problems, the Birch/Swinnerton-Dyer Conjecture, is a question about elliptic curves (a question whose solution is worth one million dollars). Elliptic curve cryptography is one of the more popular forms of cryptography used today. Elliptic curve factorization works about as well as any other known method. There are at least four professors at Harvard who specialize in elliptic curves.

Many of the theorems involving elliptic curves are straightforward and relatively easy to understand. The proofs, however, are very deep, and they are some of the most difficult works of mathematics out there. As such, we can get a feel of the important results about these objects, but we will not be able to provide justification for our results.

One of the goals for this topic is to bring together as many different fields of mathematics as possible. This course will include topics from function theory (glorified calculus), linear algebra, foundations, number theory, abstract algebra, complex analysis, cryptography, computational number theory, topology, and geometry.

2 $E(\mathbb{R})$

In this section, we lay down the foundation for the definition of an elliptic curve or a field $\mathbb{K}$, with $\mathbb{R}$ being our main example. For our definition, we need affine coordinates, projective coordinates, homogeneous polynomials, algebraic curves, nondegenerate curves and nonsingular curves. We define projective equivalence, which allows us to simplify the study of our curves. In particular, all interesting quadratic curves are equivalent to the unit circle, and all elliptic curves reduce down to a fairly standard form. Finally, we give the geometric and algebraic definition of the group structure on $E(\mathbb{K})$. While the geometric definition works for $E(\mathbb{R})$, it is clear from the algebraic definition that we can put the same group structure on $E(\mathbb{K})$ for any field $\mathbb{K}$.

2.1 Affine and Projective Coordinates

In the beginning, we will think of elliptic curves and a subtopic of *algebraic curves*. These are simply the solution sets to polynomial equations. For instance, you have looked at the solutions to $x^2 + y^2 = 1$ since algebra. We can formalize these ideas:

Definition 2.1. Let $\mathbb{K}$ be a field, and let $f(x, y)$ be a polynomial function with coefficients in $\mathbb{K}$. We say $C = \{(x, y) \in \mathbb{K}^2 | f(x, y) = 0\}$ is the *affine algebraic curve defined by* $f(x, y)$ *over* $\mathbb{K}$. If we want to emphasize the fact that we are looking at ordered pairs (x, y) over the field $\mathbb{K}$, then we use the notation $C(\mathbb{K})$.

For now, you may assume that $\mathbb{K}$ is just the set of real numbers. Through the course of this material, we will let $\mathbb{K}$ be the set a real numbers, complex numbers, rational numbers, and various finite fields.

Example 2.2. If $f(x, y) = x^2 + y^2 - 1$, then $C(\mathbb{R})$ is just the unit circle.

Example 2.3. If $g(x)$ is a polynomial function in one variable, its graph is an algebraic curve defined by $f(x, y) = y - g(x)$.

If you look at the graph of, say, $y = x^2$, the points on the graph go off to infinity. It is sometimes useful for us to add these points at infinity to the set of points that we are studying. Unfortunately, our standard coordinates in the plane (x, y) is not sufficient to include these points. We need a new set of points, which are called *projective coordinates*.

Definition 2.4. Let $\mathbb{K}$ be a field. The set $\mathbb{K}P^2$, the *projective plane*, is defined to be the set of equivalence classes on $\mathbb{K}^3 - (0, 0, 0)$, where $(x_1, y_1, z_1) \equiv (x_2, y_2, z_2)$ if and only if there is a $\lambda \neq 0$ such that $x_1 = \lambda x_2$, $y_1 = \lambda y_2$, and $z_1 = \lambda z_2$. A representative point of an equivalence class is denoted by $[x, y, z]$.

As an example, in $\mathbb{R}P^2$, the following represent the same point:

$$[1, 2, 3] = [-1, -2, -3] = [2, 4, 6] = [200, 400, 600] = [\sqrt{2}, 2\sqrt{2}, 3\sqrt{2}]$$

Our standard $\mathbb{K}^2$ is called the *affine plane*. We represent the points in $\mathbb{K}^2$ as (x, y), and we associate them with points in $\mathbb{K}^3$ by $(x, y) = [x, y, 1]$. Hence we can think of $\mathbb{K}^2$ as a subset of $\mathbb{K}^3$. This representation gets every point in $\mathbb{K}^3$ *except* for the points with a third coordinate equal to zero, i.e., $[x, y, 0]$. These will be our "points at infinity". A good way to think about them is to say that any line of the form $ax + by = c$ will "end" at the point $[b, -a, 0]$. Hence lines that are typically parallel in $\mathbb{K}^2$ will now actually intersect at a point at infinity. For a general curve, the point at infinity describes the slope of the curve as its coordinates go off to infinity.

How do we talk about curves like $y = x^2$ in projective coordinates? We have two problems: the first is that each projective point is represented by three numbers (hence we need three variables), the second is that we need our curve to be well-defined over the equivalent representations of the same point. We satisfy both requirements by looking at *homogeneous polynomials*.

Definition 2.5. A *homogeneous polynomial* over $\mathbb{K}P^3$ is a polynomial of three variables with coefficients in $\mathbb{K}$ such that each monomial has the same degree.

Example 2.6. The following are homogeneous polynomials over $\mathbb{R}P^3$:

$$x^3 + xyz + 3xz^2 + z^3 \qquad (x+y)(x+z)(y+z) \qquad 23x$$

The following are not homogeneous polynomials over $\mathbb{R}P^3$:

$$x + y^2 \qquad x^3 + xyz + z^4 \qquad x^{45} - y^{54}$$

If you want to find the set of points satisfying $f(x, y, z) = 0$, this set is well-defined on $\mathbb{K}P^3$ because a replacement of $[x, y, z]$ with $[\lambda x, \lambda y, \lambda z]$ will still give a triple that makes the polynomial equal to zero.

Since we can relate affine points to point in the projective plane, we should also relate affine polynomials to homogeneous polynomials. We call this "homogenizing the polynomial".

Definition 2.7. Let $f(x, y)$ be an affine polynomial of degree n. Define $F(x, y, z) = z^n f(x/z, y/z)$. The resultant polynomial in three variables is homogeneous. On the flip side, if we have a homogeneous polynomial $F(x, y, z)$, we can *dehomogenize* it by replacing it with $f(x, y) = F(x, y, 1)$.

Example 2.8. If $f(x, y) = y - x^2$, it's corresponding homogeneous polynomial is $F(x, y, z) = yz - x^2$. Note that all you need to do to homogenize is to add z's to any monomial that does not have the same degree as $f(x, y)$.

In the other direction, if we start out with $F(x, y, z) = y^2z - x^3 - z^3$, its corresponding affine form is $f(x, y) = y^2 - x^3 - 1$.

2.2 Projective Equivalence

We can ask ourselves, "When are two algebraic curves the same?" The answer will depend on what we mean by "same". The most restrictive answer we can give is that two curves are the same if and only if they have the same polynomial. But this does not really work, because $g(x, y)$ and $kg(x, y)$ define the same curve for any nonzero number k. We could use that as the definition: two curves are the same if they have the same point set. But then we could ask if $x^2 + y^2 = 1$ and $(x-1)^2 + (y-1)^2 = 1$ are really different curves. They are both circles of radius one, but they have different centers. If we thought location was not important, then we could treat these two curves the same. And are two circles of different radii really all that different? Again, not really.

We could classify by shape: ellipse, hyperbola, parabola, pair of lines. This is better, and this is really good if we are doing things in affine coordinates. But when we do projective coordinates, the only difference between an ellipse, hyperbola, and parabola is whether the curve goes through the line at infinity zero, two, or one time. In projective coordinates, the points are infinity are not really all that different than other points. So in projective coordinates, we will actually want to think of all of these things as the same.

For us, we want to consider to curves to be the same if they are *projectively equivalent*. This is really just going to be a change of coordinates.

Definition 2.9. Two projective curves $f(x,y,z) = 0$ and $g(x,y,z) = 0$ are *projectively equivalent* if there is a nonsingular 3×3 matrix A such that $f(\overline{x},\overline{y},\overline{z}) = g(x,y,z)$, where

$$\begin{bmatrix} \overline{x} \\ \overline{y} \\ \overline{z} \end{bmatrix} = A \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

Example 2.10. The curves $x^2+y^2-z^2 = 0$ and $(x-z)^2+(y-z)^2-z^2 = 0$ are projectively equivalent using the matrix

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Thus the affine curves $x^2+y^2 = 1$ and $(x-1)^2+(y-1)^2 = 1$ are projectively equivalent.

2.3 $C(\mathbb{R})$

A *real quadratic curve* is the solution set to a quadratic polynomial

$$g(x,y) = ax^2 + 2bxy + 2cx + dy^2 + 2ey + f = 0$$

where all of the coefficients are in $\mathbb{R}$. Let $g(x,y)$ by this polynomial, and $G(x,y,z)$ be the corresponding homogeneous polynomial

$$G(x,y,z) = ax^2 + 2bxy + 2cxz + dy^2 + 2eyz + fz^2$$

It is convenient to think of G as the result of a matrix multiplication. In particular, we could write G as

$$G(x,y,z) = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

In such a notation, it is easier to see what a projective transformation does. In particular, if $G = \mathbf{x}^T Q \mathbf{x}$, then a projective transformation given by the matrix A will replace this with the polynomial $H = \mathbf{x}^T A^T Q A \mathbf{x}$. In linear algebra, you have seen that a symmetric matrix Q can be diagonalized, and it will have real eigenvalues. So, for an appropriate choice of projective transformation, we can get rid of any mixed terms in the quadratic:

Proposition 2.11. *Let $G(x,y,z)$ be a homogeneous real quadratic. Then $G(x,y,z)$ is projectively equivalent to one of the following:*

(i) $x^2+y^2-z^2$

(ii) $x^2+y^2+z^2$

(iii) x^2+y^2

(iv) x^2-y^2

(v) x^2

The associated algebraic curve in each situation is: (i) a circle, (ii) the empty set, (iii) a single point, (iv) two distinct lines, (v) one "double" line.

Proof The matrix Q is diagonalized, and then its nonzero eignvalues are normalized. The cases simply list the possible combinations of the signs of the eigenvalues. □

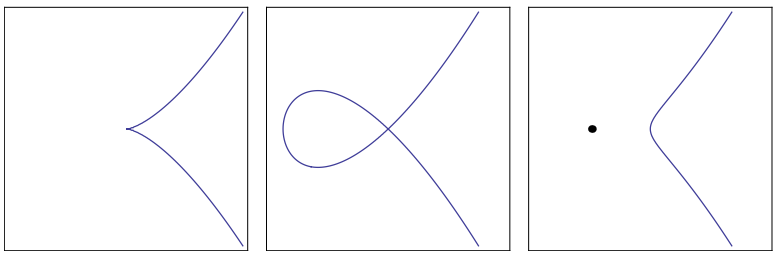
Moral of the story: if we like projective equivalence, the ONLY quadratic curve worth studying is the unit circle. All other curves are either degenerate, or they are equivalent to the circle.

2.4 Nonsingular and Nondegenerate Curves

We need two more definitions before we can define elliptic curves.

Definition 2.12. An algebraic curve $f(x, y) = 0$ is *singular* if there is a point (x_0, y_0) on the curve where there is not a well defined tangent line. An algebraic curve is *nonsingular* if it has no singular points.

Singular points show up as cusps ($y^2 = x^3$), intersections ($y^2 = x^3 + x^2$), or isolated points ($y^2 = x(x + 1)^2$).



Three types of singularities

A point (x_0, y_0) is singular if and only if $\nabla f(x_0, y_0) = \mathbf{0}$, i.e., $f_x = f_y = 0$ at (x_0, y_0). If we use projective coordinates, then $F(x, y, z) = 0$ has a singularity at (x_0, y_0, z_0) if and only if $\nabla F(x_0, y_0, z_0) \times \langle x_0, y_0, z_0 \rangle = \mathbf{0}$.

Definition 2.13. An algebraic curve $f(x, y) = 0$ is *degenerate* if the polynomial $f(x, y)$ factors. An algebraic curve is *nondegenerate* if the polynomial $f(x, y)$ factors.

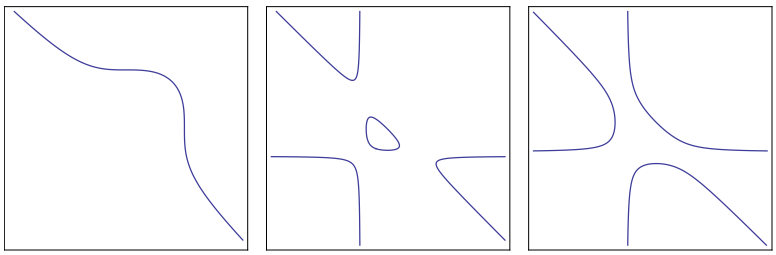
The curve $x^2 - y^2 = 0$ is degenerate. It really should be thought of as two linear equations as opposed to one quadratic equation. With these definitions, we can now say, up to projective equivalence, the only nondegenerate, nonsingular quadratic curve is the unit circle.

2.5 Elliptic Curves

Finally, we can define an elliptic curve. In one sense, this is going to be one step up from a quadratic curve, in that we are going to use cubic polynomials instead of quadratic. And like the quadratic case, we really do not care about cubic curves that are singular or degenerate.

Definition 2.14. An *elliptic curve* is an algebraic curve defined by a nondegenerate, nonsingular cubic polynomial (either affine or projective). We write $E(\mathbb{K})$ to describe the point set of our elliptic curve, where $\mathbb{K}$ is our field of interest.

Example 2.15. Note that $E(\mathbb{R})$ is nothing more than the graph of the cubic equation. Here are some pictures of $E(\mathbb{R})$ for various cubic polynomials:



Various elliptic curves

Singular cubics will have some type of cusp, intersection, or isolated point. All three curves below Definition 2.12 are examples of singular cubic curves (and thus not quite elliptic curves).

Cubic equations have ten coefficients, and they are somewhat unwieldy. Like quadratic equations, we would like to do a projective transformation that would perhaps reduce the cubic into something more understandable. There are several *normal forms* for elliptic curves that people use, but this is one of the more popular:

Theorem 2.16. *Let $f(x, y)$ be a nondegenerate, nonsingular cubic polynomial. Then by a projective transformation, the equation $f(x, y) = 0$ can be rewritten in the normal form*

$$y^2 = x^3 + ax + b$$

where a and b are numbers in our chosen field and $4a^3 + 27b^2 \neq 0$.

Proof Because it is easier to see the projective transformations while working in projective coordinates, we begin with a cubic homogeneous polynomial in three variables:

$$ax^3 + bx^2y + cxy^2 + dy^3 + ex^2z + fxz^2 + gxyz + hy^2z + jyz^2 + kz^3 = 0$$

It is a fact that if the coefficients are real or complex, then this cubic must have an inflection point on it (this follows from a cool little theorem called Bezout's Theorem, which is not too difficult, but it is more of a tangent than we want to do right now). By a projective transformation, we may assume that (1) the point $[0, 1, 0]$ is on the curve, (2) the tangent

line to the curve at $[0,1,0]$ is given by the equation $z=0$, and (3) the point $[0,1,0]$ is an inflection point. These three conditions force the coefficients of y^3, x^2y, and xy^2 to be zero. Thus we have:

$$ay^2z + bxyz + cyz^2 = dx^3 + ex^2z + fxz^2 + gz^3$$

where the coefficients in this expression are not the same as the coefficients in the original expression (if we kept renaming coefficients, we would run out of letters, or we would need to subscript our coefficients, which would be annoying). Now if a was zero, then our curve would be singular at $[0,1,0]$, so we may assume that $a \neq 0$. Hence we can replace z with z/a to get (again, renaming the coefficients)

$$y^2z + bxyz + cyz^2 = dx^3 + ex^2z + fxz^2 + gz^3$$

Then we can replace y with $y - bx/2 - cz/2$ to get (again, renaming the coefficients)

$$y^2z = ax^3 + bx^2z + cxz^2 + dz^3$$

Here we may assume (again) that $a \neq 0$, else our polynomial is degenerate. Thus we can set the coefficient of x^3 equal to 1 by replacing z with az and then dividing everything by a (which does not change the point set of the polynomial).

$$y^2z = x^3 + bx^2z + cxz^2 + dz^3$$

Then we can get rid of the x^2z term by setting x equal to $x - b/3z$:

$$y^2z = x^3 + axz^2 + bz^3$$

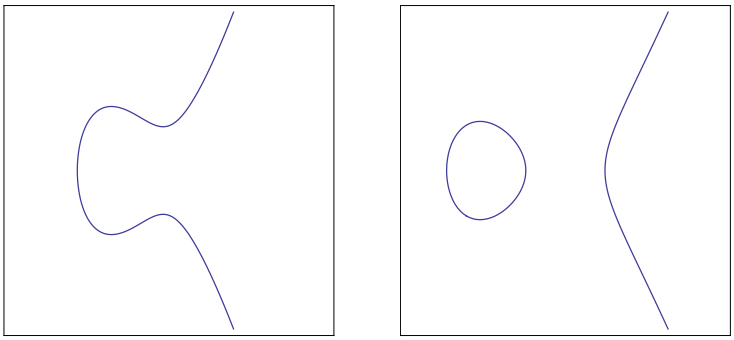
Finally, set $z=1$ to turn the homogeneous polynomial into an affine polynomial:

$$y^2 = x^3 + ax + b$$

The curve is obviously nondegenerate, no matter what a and b are. But we still need to require that this curve is nonsingular. If $f(x,y) = x^3 + ax + b - y^2$ is singular, then $\nabla f(x,y) = \langle 3x^2 + a, -2y \rangle$ must be zero. This requires $y=0$ and $x^2 = -a/3$. Substituting these into $f(x,y)$, we find that we also require $-ax/3 + ax + b = 0$, so $x = -3b/(2a)$. Substituting this back into $x^2 = -a/3$, we get $9b^2/(4a^2) = -a/3$, which can be rewritten as $4a^3 + 27b^2 = 0$. Thus if this term is nonzero, then our curve is nonsingular. □

The expression $\Delta = 4a^3 + 27b^2$ is known as the *discriminant*. It tells how many real solutions the equation $x^3 + ax + b = 0$ has (just like $b^2 - 4ac$ tells how many real solutions the quadratic equation has). In particular, if $\Delta > 0$, we have one real solution and two complex solutions. If $\Delta = 0$, we have a double or triple root. If $\Delta < 0$, we have three real solutions.

In our normalized form, our elliptic curve can take one of two basic shapes, depending on whether Δ is positive or negative:



Two normalized forms of an elliptic curve

There are other normalized forms that people use. For instance, you can find sources that reduce their elliptic curves to $y^2 = 4x^3 + ax + b$ or $y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ or $y^2 = x(x-1)(x-\lambda)$. We will tend to stick to $y^2 = x^3 + ax + b$, but if a particular curve is easier to describe in another form, we will do it.

Finally, while we will use the affine form of the elliptic curve almost exclusively now, we will always keep in mind that the point $[0, 1, 0]$ is on this curve. We will be using this point extensively, and we will frequently call it the "point at infinity".

2.6 The Group Structure of $E(\mathbb{K})$

Much of the interest of elliptic curves boils down to an interesting group structure that one can put on it's point set. We want to define this group now. First, let us recall the definition of a group:

Definition 2.17. Let G be a set. We say that G is an *abelian group* if there exists a binary map (written as "+") on G satisfying the following properties:

- For all $P, Q \in G$, $P + Q = Q + P$.
- For all $P, Q, R \in G$, $(P + Q) + R = P + (Q + R)$.
- There exists an element of G, called the identity, notated by O, such that $P + O = O + P = P$ for all $P \in G$.
- For all $P \in G$, there exists an element $-P$ such that $P + (-P) = (-P) + P = O$. This element is called the inverse of P.

If we left out the first bullet, then we would be dealing with just a group, as opposed to an abelian group. However, the group structure on elliptic curves happens to be abelian.

To describe the group structure, we need to create a different binary function, called the pound operator. It can be defined geometrically, and it is based on the following theorem:

Theorem 2.18. *Let $E(\mathbb{R})$ be an elliptic curve. Let ℓ be a line that intersects the curve in at least two points. Then ℓ intersects the curve at exactly three points.*

Proof Let $y^2 = x^3 + ax + b$ be the polynomial for our elliptic curve. If ℓ is not vertical, then we can express ℓ by the equation $y = mx + c$. Assume that ℓ and $E(\mathbb{R})$ intersect at the points (x_1, y_1) and (x_2, y_2). Combining the two equations, we know that the x coordinates of the points of intersection satisfy:

$$(mx + c)^2 = x^3 + ax + b$$

This is a cubic equation, and hence has exactly three solutions. We know that two of the solutions are x_1 and x_2, both of which are real. Since complex solutions of a real polynomial equation must come in pairs, the third solution, x_3, must be real also. The corresponding y value is $mx_3 + c$, which is also real. There can be no more intersections, because there are no more solutions to the cubic.

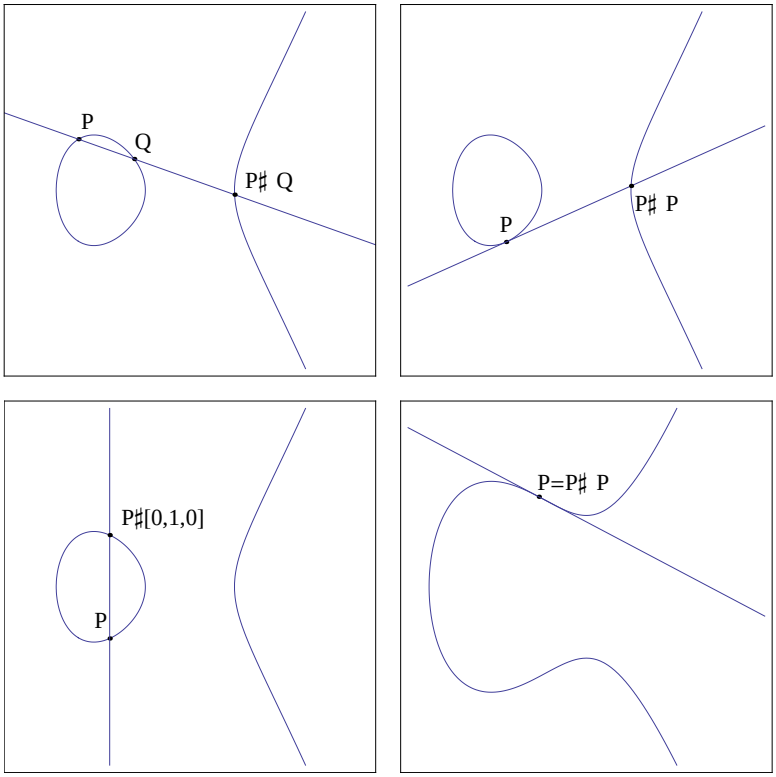
If ℓ is vertical, then ℓ intersects $E(\mathbb{R})$ at $[0, 1, 0]$. Because the curve is symmetric with respect to the x-axis, a vertical line will intersect the curve in zero or two non-infinite places. If it does not intersect the affine portion, then the theorem does not apply. If it does intersect the affine portion, then the theorem is satisfied.

Finally, we need to consider the case when ℓ is the line at infinity. In this case, ℓ intersects $E(\mathbb{R})$ only at $[0, 1, 0]$. But since $[0, 1, 0]$ is an inflection point with tangent line equal to the line at infinity, the intersection counts with a multiplicity of three. □

Note that we need to be specific when we talk about intersecting. For our purposes, a tangent line will intersect the curve at the point of tangency with a multiplicity of two, and hence will count as two points of intersection (so for the three solutions x_1, x_2, and x_3 of the cubic in the proof, two of the solutions are equal). For a tangent line at an inflection point, we want to consider this as an intersection with a multiplicity of three, and hence will count as three points of intersection.

With this theorem in mind, given two points P and Q on an elliptic curve (possibly the same point), there is a unique line that connects P to Q (or if $P = Q$, then we think of the tangent line at P). This unique line intersects the elliptic curve at least twice, and so by the theorem, it must intersect the curve exactly three times. The third point of intersection is uniquely determined by P and Q, and thus we have a binary operator on our elliptic curve.

Definition 2.19. Let $E(\mathbb{R})$ be an elliptic curve. Define the *pound operator* as a binary map, represented by the symbol #, such that if P and Q are two points on $E(\mathbb{R})$, then $P\#Q$ is defined as the third point of intersection between $E(\mathbb{R})$ and the line through P and Q.



Various configurations of the pound operator

While it is a binary operator, the pound operator does NOT turn our elliptic curve into a group. The operation is commutative, but it is not associative, and it does not have an identity element (and hence no inverses either). However, we can use the pound operator to create our addition operator, and hence our group structure.

Theorem 2.20. *Let $E(\mathbb{R})$ be an elliptic curve. Pick any point on the curve, and call it O. Define an addition operator on $E(\mathbb{R})$ by*

$$P + Q = (P\#Q)\#O$$

for all $P, Q \in E(\mathbb{R})$. Then with this operator, $E(\mathbb{R})$ is a group with identity element O. Inverses are defined by

$$-P = P\#(O\#O)$$

Proof As always, we have to show that $E(\mathbb{R})$ with this definition of addition satisfies all of the properties of being a group:

- Commutativity: Trivial, because the pound operator is commutative.
- Associativity: Annoying. We need to show:

  $$(P + Q) + R = P + (Q + R)$$

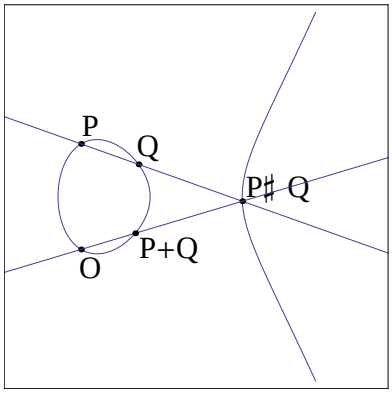
  which, in terms of the pound operator, turns into

  $$(((P\#Q)\#O)\#R)\#O = (P\#((Q\#R)\#O))\#O$$

  We can show this geometrically, but it is not pretty.

- Identity: This isn't too bad. It's an exercise.
- Inverse: Again, not too bad. It's an exercise.

□



Geometric definition of addition

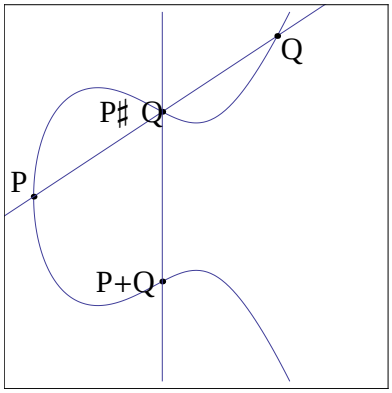
Example 2.21. Let our elliptic curve be $y^2 = x^3 + 1$. There are six points on this curve that have integer coordinates: $(-1, 0)$, $(0, 1)$, $(0, -1)$, $(2, 3)$, $(2, -3)$, and $[0, 1, 0]$ (see handout). We can choose any of these points to be our identity: we will pick $(-1, 0)$. With this point as our identity, it turns out that these six points form a subgroup of $E(\mathbb{R})$, meaning that they are closed under addition. We can work out the addition table:

$+$	$(-1,0)$	$(0,1)$	$(0,-1)$	$(2,3)$	$(2,-3)$	$[0,1,0]$
$(-1,0)$	$(-1,0)$	$(0,1)$	$(0,-1)$	$(2,3)$	$(2,-3)$	$[0,1,0]$
$(0,1)$	$(0,1)$	$(2,3)$	$(-1,0)$	$[0,1,0]$	$(0,-1)$	$(2,-3)$
$(0,-1)$	$(0,-1)$	$(-1,0)$	$(2,-3)$	$(0,1)$	$[0,1,0]$	$(2,3)$
$(2,3)$	$(2,3)$	$[0,1,0]$	$(0,1)$	$(2,-3)$	$(-1,0)$	$(0,-1)$
$(2,-3)$	$(2,-3)$	$(0,-1)$	$[0,1,0]$	$(-1,0)$	$(2,3)$	$(0,-1)$
$[0,1,0]$	$[0,1,0]$	$(2,-3)$	$(2,3)$	$(0,1)$	$(0,-1)$	$(-1,0)$

It might be a bit easier to visualize if we name the points. So, let $(-1, 0)$ be O, $(0, 1)$ be P, $(2, 3)$ be Q, and $[0, 1, 0]$ be ∞. Then $(0, -1)$ is $-P$ and $(2, -3)$ is $-Q$. Our addition table now looks like:

$+$	O	P	$-P$	Q	$-Q$	∞
O	O	P	$-P$	Q	$-Q$	∞
P	P	Q	O	∞	$-P$	$-Q$
$-P$	$-P$	O	$-Q$	P	∞	Q
Q	Q	∞	P	$-Q$	O	$-P$
$-Q$	$-Q$	$-P$	∞	O	Q	$-P$
∞	∞	$-Q$	Q	P	$-P$	O

It turns out that we get the same group structure no matter which point we pick for our identity. Therefore, it makes sense to pick one that will make our work easier. Since every normalized elliptic curve has a point at $[0, 1, 0]$, we will pick this as our identity. With this, the second step in the calculation of $P + Q$ is easy: if $P\#Q$ is the point (x, y), then $P + Q$ is the point $(x, -y)$.



Addition when $O = [0, 1, 0]$

Our next question: given two points $P = (x_1, y_1)$ and $Q = (x_2, y_2)$, how do we write $P + Q$ in terms of x_1, x_2, y_1, and y_2? The full statement is as follows:

Theorem 2.22. *Let $y^2 = ax^3 + bx^2 + cx + d$ be an elliptic curve with points (x_1, y_1) and (x_2, y_2) on it, and assume that the two points are not reflections of each other about the x-axis. Let our identity by $[0, 1, 0]$. Define the value m as*

$$m = \begin{cases} \frac{y_1 - y_2}{x_1 - x_2} & \text{if } x_1 \neq x_2 \\ \frac{3ax_1^2 + 2bx_1 + c}{2y_1} & \text{if } x_1 = x_2 \end{cases}$$

Then the formula for $(x_3, y_3) = (x_1, y_1) + (x_2, y_2)$ is:

$$\begin{aligned} x_3 &= \frac{1}{a}(m^2 - b) - x_1 - x_2 \\ y_3 &= -m(x_3 - x_1) - y_1 \end{aligned}$$

The three cases not covered by the formula are:

$$\begin{aligned} (x_1, y_1) + (x_1, -y_1) &= [0, 1, 0] \\ (x_1, y_1) + [0, 1, 0] &= (x_1, y_1) \\ [0, 1, 0] + [0, 1, 0] &= [0, 1, 0] \end{aligned}$$

Proof The three special cases can be verified directly using the pound definition. For the formula itself, note that m is the slope of the line connecting the two points (for the second formula, the two points are the same, so the connecting line is the tangent line at the point (x_1, y_1)). Thus the line is $y = m(x - x_1) + y_1$. We substitute this into the equation $ax^3 + bx^2 + cx + d - y^2 = 0$, and we see that the x values for the points of intersection must satisfy:

$$ax^3 + bx^2 + cx + d - (m(x - x_1) + y_1)^2 = 0$$

Compare this polynomial to $a(x - x_1)(x - x_2)(x - x_3) = 0$, and the x^2 coefficients tell us that $ax_1 + ax_2 + ax_3 = m^2 - b$. Our formula for x_3 follows. We get our formula for y_3 by substituting x_3 into the line equation, then negating the result. □

Example 2.23. For our elliptic curve $y^2 = x^3 + 1$, we can work out $(0, 1) + (2, 3)$ using the formula: $a = 1$, $b = 0$, $c = 0$, $d = 1$, $m = 1$, $x_3 = -1$, $y_3 = 0$, and so $(0, 1) + (2, 3) = (-1, 0)$.

Note that our formula does not assume that the cubic has been reduced all the way down to $x^3 + ax + b$. If we wanted to, we could make the formula even more general, but this will work for our purposes.