7 Elliptic Curves

One of the deepest, coolest, and most interesting fields of mathematics is the subject of elliptic curves. Fortunately for us, they're also widely used in cryptography today.

7.1 Groups and Fields

We need to introduce two fundamental ideas from algebra.

7.1.1 Groups

Definition 7.1. A *group* is a set G and a binary operation $\star : G \times G \to G$ with the properties that:

1. There is an *identity element* $e \in G$ such that $e \star g = g \star e = g$ for all $g \in G$;
2. For every $g \in G$, there is an *inverse element* g^{-1} such that $g \star g^{-1} = g^{-1} \star g = e$;
3. And the operation is *associative*, i.e. for every $f, g, h \in G$ we have $(f \star g) \star h = f \star (g \star h)$.

By convention we treat this operation as multiplication. A group always reflect some sort of symmetries; any time you have a collection of symmetries or a repeating pattern you can describe it as a group.

Example 7.2.
1. The integers with the operation of addition form a group $\mathbb{Z}$.
2. The non-zero rational numbers with the operation of multiplication form a group $\mathbb{Q}$.
3. The set of integers mod n, with the operation of addition mod n, forms a group $\mathbb{Z}/n\mathbb{Z}$.
4. The set of invertible $n \times n$ matrices with the operation of matrix multiplication form a group, called the *general linear group of degree n* or $GL(n)$.
5. The set of $n \times n$ matrices with determinant 1 with the operation of matrix multiplication form a group called the *special linear group of degree n* or $SL(n)$.
6. The set of rotations of a circle, with the operation of composition, form a group.
7. The set of permutations of a n-element set, with the operation of composition, forms a group called the *symmetric group on n letters* or S_n.

You'll note that in some of these groups, the operation is commutative: $a+b=b+a$. In others, like matrix multiplication or the permutations, the operation is not commutative. In general we don't assume groups are commutative; but many of the groups we're interested in in this class are.

Definition 7.3. Let $(G,\star)$ be a group. If the operation $\star$ is commutative, i.e. $g\star h = h\star g$ for all $g,h\in G$, then we say G is an *abelian* group.

If G is an abelian group, we often write $+$ for the operation and write $-g$ for the inverse of g and 0 for the identity. We *never* use addition notation if our group is not abelian.

We often want to fix an element g and consider the set of all the group elements we can get just from g.

Definition 7.4. Let $g\in G$. Then the set $\{g^n : n\in\mathbb{Z}\}$ is called the subgroup of G *generated by* g, often notated $\langle g\rangle$.

The size of this group is the *order of* g, and is the least positive integer $m=\text{ord}_G(g)$ such that $g^m=e$.

If there is some $g\in G$ such that $G=\langle g\rangle$, we say that G is *cyclic* and that g is a *generator* for G.

This should sound very familiar from section 6.1. This is in fact a generalization; the idea of a generator here corresponds to the idea of a primitive root modulo m. We can also generalize one further fact:

Fact 7.5. *If G is a group of order n and $g\in G$ then* $\text{ord}_G(g)|n$.

7.1.2 Fields

A field is in essence a set in which we can do addition, multiplication, and division. Formally:

Definition 7.6. A *field* is a set K together with two operations $+$ and $\cdot$, such that

1. K is an abelian group under the operation $+$;
2. The set $K\setminus\{0\}$ of non-zero elements of K is an abelian group under $\cdot$;
3. and we have the distributive law $k(x+y)=kx+ky$.

Example 7.7. 1. The real numbers, the rational numbers, and the complex numbers are all fields.

2. The integers are *not* a field, since they don't form a group under multiplication: there are no inverses. The requirement that we have a *group* under multiplication is what allows division.

The most important example for us is $\mathbb{Z}/p\mathbb{Z}$. This clearly forms a group under addition. Further, we know that every non-zero element is invertible mod p, so the non-zero elements form a group under multiplication. Thus $\mathbb{Z}/p\mathbb{Z}$ is a field. It is in fact the only field of order p, and when we're thinking of it as a field we often denote it $\mathbb{F}_p$.

In contrast, if m is composite, then $\mathbb{Z}/m\mathbb{Z}$ is not a field, since any factor of m will not be invertible mod m.

Typically when we're looking for solutions to some equation, we need to specify the field we're working in. We know this already: The equation $x^2 - 2 = 0$ has solutions in the reals but not in the rationals, and $x^2 + 1 = 0$ has solutions in the complex numbers but not in the reals.

For our cryptographic applications, we typically want to be working in the field $\mathbb{F}_p$ for some large prime p. But the important idea here is that most of our algebra works the same in any field. We can add, subtract, multiply, and divide the same way regardless.

7.2 Elliptic Curves

Definition 7.8. An elliptic curve over a field K is a smooth projective curve over K of genus 1, together with a point defined over K.

That...doesn't mean very much. Let's try again.

Definition 7.9. An *elliptic curve* over a field K is given by an equation of the form

$$y^2 = x^3 + Ax + B \tag{7.1}$$

where $A, B \in K$, provided that the discriminant $\Delta = 4A^3 + 27B^2 \neq 0$. The set of solutions to this equation with coordinates in K is the set of points of the elliptic curve, and is denoted $E(K)$.

We say a curve written this way is in *Weierstrass Form*.

The condition on the discriminant guarantees that the cubic does not have any repeated roots; if we factor $x^3 + Ax + B = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3)$, then $\Delta = 4A^3 + 27B^2 \neq 0$ if and only if $\alpha_1, \alpha_2, \alpha_3$ are distinct. This turns out to be important for the group law we wish to build.

Remark 7.10. An elliptic curve is *not* an ellipse. The name comes from the relationship to elliptic integrals, which are the integrals you need to compute the circumference of an ellipse.

Over the complex numbers, an elliptic curve forms a torus. Over the reals, we get a characteristic pinched curve or circle-and-wiggle curve.

Elliptic curves start off as an example of a basic number theory question: when does a given equation have solutions in the rational numbers? But elliptic curves have a very nice property that makes this question in some ways much easier: the group law.

7.2.1 Bezout's Theorem

The group law depends on a result from algebraic geometry known as Bezout's Theorem. We say that a planar curve is of *degree d* if it is defined by a polynomial whose highest term is degree d.

Theorem 7.11 (Bezout's Theorem)**.** *Suppose C_1 is a curve of degree d and C_2 is a curve of degree e. Then there are exactly $e \cdot d$ points in the intersection $C_1 \cap C_2$, up to some technical conditions.*

It's actually really easy to come up with counterexamples to Bezout's Theorem without the technical conditions. One is that Bezout's Theorem holds over the complex numbers but not over the reals (since, say, the intersection of $x = -1$ and $x = y^2$ has no real solutions but two complex solutions). This condition won't actually be important for the application we care about.

There are two other technicalities that are more important. The first is points "at infinity". The lines $y = 1$ and $y = 2$ are each degree 1, so should intersect at exactly one point. Obviously they have no points in common, though. We introduce "points at infinity" so that any pair of parallel lines has exactly one intersection. (The formalization of this idea is called "projective geometry").

The other is that when we're counting the number of intersections, some of them count "more than once". This corresponds to the idea that $(x-1)^5$ has one root "five times", and thus we can say that any degree-d polynomial has exactly d roots over the complex numbers.

We can visualize some of this by thinking of the intersection of a circle and a line. If the line doens't hit the circle, then they have no real intersections (but two complex intersections, since the line has degree 1 and the circle has degree 2). If the line goes through the circle, then there are two actual real intersections we can see. And if the line is *tangent* to the circle, there is one intersection but it counts twice.

7.2.2 Geometry and the Group Law

How does this let us find points on an elliptic curve? The elliptic curve has degree 3, so any line should intersect it in exactly three points. Thus if we have any two points on an elliptic curve, we can draw a line through the two points and find a third point of intersection.

This won't get us terribly far on its own, though—we just get one new point. In order to get us out of the rut we could get stuck in, we actually reflect our new point across the line $y = 0$—because every y in our defining equation is squared, then whenever (x, y) is a point on the curve, so is the point $(x, -y)$.

Definition 7.12. Let P, Q be two points on an elliptic curve. We can draw a line through them and find a third point on the elliptic curve, which we'll call R. Let R' be the reflection of R across the x-axis. Then we define $P \oplus Q = R'$.

Remark 7.13. This is very specifically *not* the result we'd get by adding the coordinates of P and Q; doing that will probably not give another point on the elliptic curve.

There are a couple special cases we need to deal with. The first is when our line has intersection with multiplicity. In particular, the only way this can happen is if the line is tangent to the elliptic curve, in which case it intersects it with multiplicity two at the point of tangency. So if the line through P and Q is tangent to the curve at P, then our point R from this algorithm is just P. And if we want to compute $P \oplus P$ then we draw the line tangent to the curve at P and find the third point of intersection (and then reflect across the x-axis).

The other special case is to reintroduce the projectivity. Recall that Bezout's theorem only holds if you allow points at infinity. In particular, we say that every elliptic curve contains a point at infinity which is intersected by every vertical line; we call this point $\mathcal{O}$.

Remark 7.14. We can actually recast our "addition" law as drawing a line through P and Q and finding the third point of intersection R; and then drawing a line through R and $\mathcal{O}$ and finding the third point of intersection R', and then defining $P \oplus Q = R'$. This is the same definition as earlier, slightly generalized.

How does the $\oplus$ operation work with $\mathcal{O}$? Every line through $\mathcal{O}$ and another point is vertical. So we can see that if P is any point, then the line through $\mathcal{O}$ and P also intersects the reflection of P across the x-axis; reflecting this back across the x-axis gives P, so $\mathcal{O} \oplus P = P$. Thus $\mathcal{O}$ is an identity for the $\oplus$ operation.

(You might want to check here that $\mathcal{O} \oplus \mathcal{O} = \mathcal{O}$. This is in fact true, because a line can intersect $\mathcal{O}$ with multiplicity 3. But proving that is far beyond the scope of this discussion,

so we'll just take it as a definition for now).

We'd like to show that this operation gives us an abelian group. The operation is clearly commutative, since the line through P and Q is the same as the line through Q and P. Associativity is straightforward, but incredibly tedious. And we already have an identity. So we just need to show that we have inverses.

Let P be a point, and let Q be the reflection of P across the x-axis. Then the line through P and Q is vertical, and the third point of intersection is $\mathcal{O}$. The line through $\mathcal{O}$ and $\mathcal{O}$ has its third point of intersection in $\mathcal{O}$, so we see that $P \oplus Q = \mathcal{O}$, and by definition Q is an inverse of P. Thus in general we notate the reflection of P across the x-axis by $-P$.

Proposition 7.15. *Let E be an elliptic curve defined over a field K. Then the set of points $E(K)$ forms a group under the operation $\oplus$ defined above.*

Definition 7.16. We write $P - Q$ for $P \oplus (-Q)$ and we write $nP = P \oplus \cdots \oplus P$.

This operation can be characterized in another, possibly more intuitive way.

Fact 7.17. *If P, Q, R are points of an elliptic curve all on the same line, then $P \oplus Q \oplus R = \mathcal{O}$.*

Thus if P, Q, R are all on the same line, then $P \oplus Q = -R$.

7.3 Elliptic Curves over $\mathbb{Q}$

Let's first discuss elliptic curves over the rational numbers. These are not per se useful for cryptography, but they have two nice properties. One is that they're in many ways more theoretically rich and interesting; the other is that they are far easier to visualize.

It's a very difficult question in general to find all the points on an elliptic curve—or, indeed, any of them. It's not trivial to figure out whether an elliptic curve *has* any points over $\mathbb{Q}$ other than $\mathcal{O}$. The primary use of the group law is to constrain how many points an elliptic curve can have, and then to help us find them. If we start with one point, we can use the group law to repeatedly add it to itself to generate more.

Fact 7.18. *If $E(\mathbb{Q})$ is finite, then either $E(\mathbb{Q}) = \mathbb{Z}/n\mathbb{Z}$ for $n \in \{1, 2, \ldots, 9, 10, 12\}$, or $E(\mathbb{Q}) = \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ for $n \in \{2, 4, 6, 8\}$.*

Conjecture 7.19. *Exactly 50% of elliptic curves have an infinite number of points.*

However, if we do have a point, we can do arithmetic with it straightforwardly. Suppose we want to add the points (x_1, y_1) and (x_2, y_2). Then the line including both of these points will have equation

$$y = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) + y_1. \tag{7.2}$$

If we substitute this into our elliptic curve equation, we should get three solutions for x; two of them will be x_1 and x_2, and the third will be x_3, the x-coordinate we're looking for. We plug x_3 back into the equation for the line to get $-y_3$, and multiply the y-coordinate by -1 to get our final point $(x_1, y_1) \oplus (x_2, y_2)$.

Example 7.20. Let's consider the elliptic curve $E : y^2 = x^3 - 15x + 18$. We can calculate the discriminant $\Delta = 4(-15)^3 + 27(18)^2 = -4752 \neq 0$, so this is in fact an elliptic curve.

We can check that $P = (7, 16)$ and $Q = (1, 2)$ are both in $E(\mathbb{Q})$. Let's compute $P \oplus Q$. The line through these points is given by

$$y = \frac{16-2}{7-1}(x-1) + 2 = \frac{7}{3}(x-1) + 2 = \frac{7}{3}x - \frac{1}{3}.$$

Substituting this into the equation for the elliptic curve gives

$$\begin{aligned} (7x/3 - 1/3)^2 &= x^3 - 15x + 18 \\ 49x^2/9 - 14x/9 + 1/9 &= x^3 - 15x + 18 \\ 0 &= x^3 - \frac{49}{9}x^2 - \frac{121}{9}x + \frac{161}{9}. \end{aligned}$$

This is a cubic equation, which is kind of annoying to solve. (There is a cubic formula analogue to the quadratic formula, but it's considerably more complex). However, we have an advantage since we already know two of the roots. This gives us two easy ways out.

The less easy way is to do polynomial long division. We can divide this polynomial by $(x-1)$ and then by $(x-7)$ to find the third term; we get

$$x^3 - \frac{49}{9}x^2 - \frac{121}{9}x + \frac{161}{9} = (x-7)(x-1)(x+23/9)$$

so the third root is $-23/9$.

But there's an even easier way. We know that our polynomial is equal to $(x-7)(x-1)(x-x_3)$; we can check then that the coefficient of x^2 must be equal to $-7-1-x_3$. Thus we set up

$$\begin{aligned} -49/9 &= -7 - 1 - x_3 \\ -49/9 + 8 &= -x_3 \\ -23/9 &= x_3. \end{aligned}$$

We know the x-coordinate of the third point on the line, so now we need the y-coordinate. We can get this easily by substituting back into the equation for our line, and we get

$$y = \frac{7}{3}(-23/9) - \frac{1}{3} = \frac{-161}{27} - \frac{1}{3} = \frac{-170}{27}.$$

Finally we reflect across the x-axis, to get $P \oplus Q = (-23/9, 170/27)$.

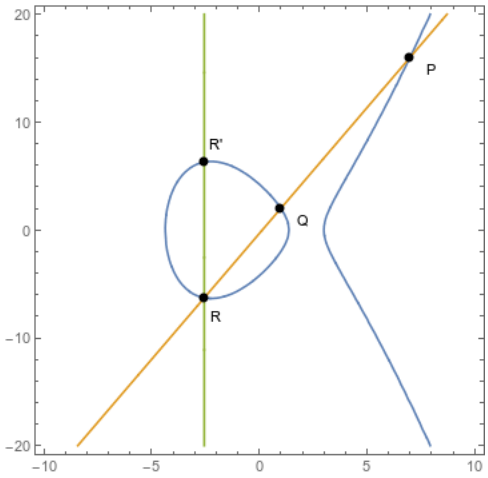

Figure 7.1: Calculating $P \oplus Q$

Example 7.21. Now let's compute $Q \oplus Q$. Recall that when we have a point occur twice, that means we want the tangent line at that point. To find the slope we use calculus. We know that

$$\begin{aligned} y^2 &= x^3 - 15x + 18 \\ 2yy' &= 3x^2 - 15 \\ 2 \cdot 2 \cdot y' &= 3 \cdot (1)^2 - 15 \\ y' &= \frac{-12}{4} = -3. \end{aligned}$$

Thus the tangent line at Q is given by

$$y = -3(x-1) + 2 = -3x + 5.$$

Substituting this into our curve gives

$$\begin{aligned} (5-3x)^2 &= x^3 - 15x + 18 \\ 25 - 30x + 9x^2 &= x^3 - 15x + 18 \\ 0 &= x^3 - 9x^2 + 15x - 7. \end{aligned}$$

Using the same trick as before, we know that $x^3 - 9x^2 + 15x - 7 = (x-1)(x-1)(x-x_3)$ so we have $-1 - 1 - x_3 = -9x^2$, giving us $x_3 = 7$.

To find the y-coordinate of the intersection, we substitute this into our linear equation, and get $y = -21 + 5 = -16$. Reflecting this across the x-axis, we get $Q \oplus Q = (7, 16) = P$.

Thus we've also seen that $3Q = (-23/9, 170/27)$.

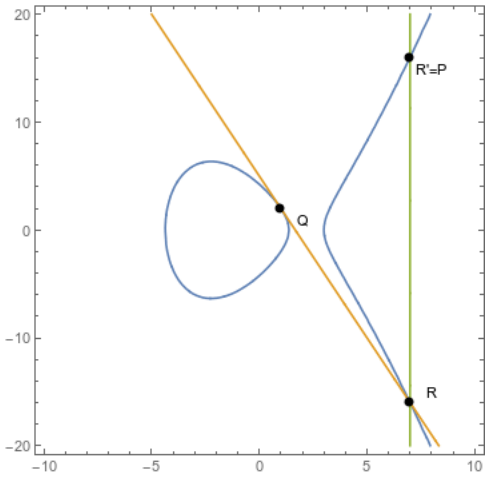

Figure 7.2: Calculating $2Q$

References

[Singh(2015)] Vikram Singh. A practical key exchange for the internet using lattice cryptography. Cryptology ePrint Archive, Report 2015/138, 2015. http://eprint.iacr.org/2015/138.