An introduction to elliptic curve cryptography - Part 1

Sources

- Craig Costello, *Pairings for beginners*
  - https://static1.squarespace.com/static/5fdbb09f31d71c1227082339/t/5ff394720493bd28278889c6/1609798774687/PairingsForBeginners.pdf

Contents

1. Elliptic curve cryptography basics 2
Elliptic curve points 2
Group Law 4
Projective coordinates 6
Order of the group 7
Discrete logarithm, and why cryptopgraphers like elliptic curves 8

1. Elliptic curve cryptography basics

Elliptic curve points

Let $\mathbb{F}$ be a field. An elliptic curve over $\mathbb{F}$ is the set of points (x, y) over $\mathbb{F}^2$ which satisfy the equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

parameterized by $a_1, ..., a_6$. We also consider the point at infinity, which we will call O, to be a solution to the equation and thus a point on the curve. One intuition about the point at infinity can be found if you graph a curve over field $\mathbb{R}$, it will look like it extends infinitely (see the diagram below). We will provide more intuitions later.

You can transform x and y linearly to make some of the a_i become 0. This does not affect any mathematical property of the curve that we care about but in practice makes computation more efficient. We often linearly transform to make $a_1 = a_2 = a_3 = 0$, so we say (without loss of generality) that our curve has equation

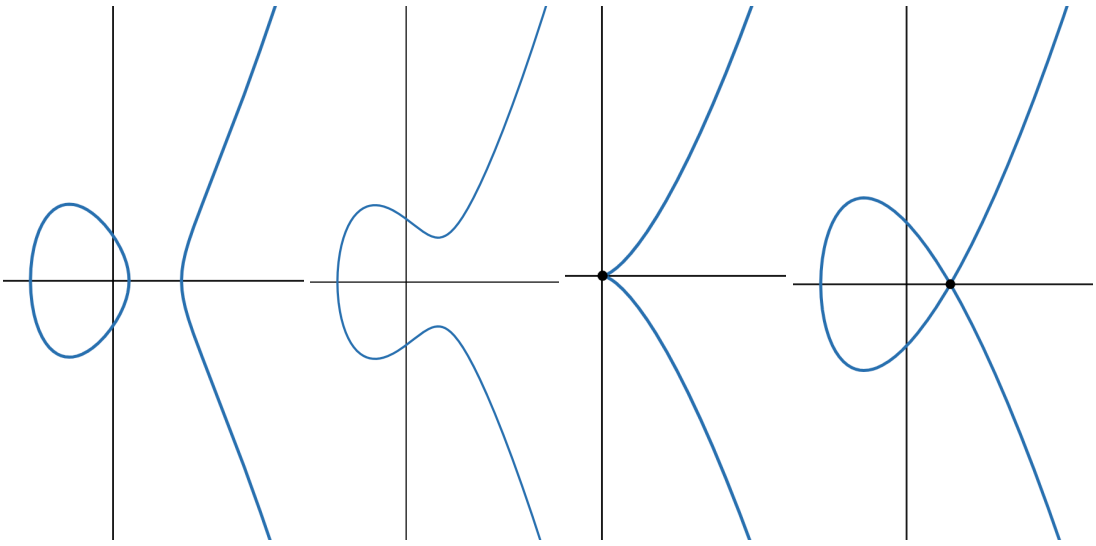
$$y^2 = x^3 + ax + b$$

for some a, b in our field. The details of this transformation are omitted.

Actually, the relevant linear transformations on x and y involve denominators of 2 and 3, so in order to make this simplification, we must be working in a field whose characteristic is netiher 2 nor 3.

For the purposes of cryptography, we only consider finite fields of large characteristic, so we may assume that every curve's equation can be expressed as $y^2 = x^3 + ax + b$.

For visualization purposes, here are four curves over the field $\mathbb{R}$.



$y^2 = x^3 - 3x + 1$ $y^2 = x^3 - 3x + 4$ $y^2 = x^3$ $y^2 = x^3 - 3x + 2$

The last third and fourth curves have "singular points" $(0, 0)$ and $(1, 0)$, respectively, so we say that the curves are singular. The first and second curves don't have singular points, so we call them "smooth". We don't formally define singular points now, but vaguely mention that the local behavior of the curve is qualitatively different at these points. The phenomenon of singular points extends to curves over finite fields as well, even though it cannot be easily visualized.

As an optional note, the curve defined by $y^2 = x^3 + ax + b$ has a singular point if and only if the right side of the equation has a multiple root, or equivalently, its discriminant $-4a^3 - 27b^2$ equals 0.

In a sense, "most" curves over large fields are smooth. We will only work with smooth curves, because they have nice properties that singular curves do not.

Summary.

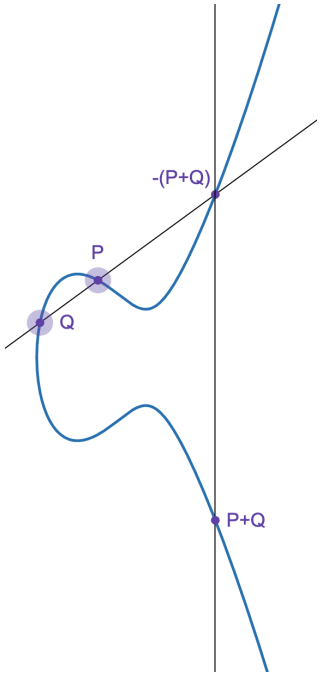
- Given a field $\mathbb{F}$ (not of characteristic 2 or 3) and parameters $a, b \in \mathbb{F}$, we can consider the equation
$$y^2 = x^3 + ax + b.$$
- This equation determines an elliptic curve, which for now we regard as the set of points
$$E = \{(x, y) \in \mathbb{F}^2 \mid y^2 = x^3 + ax + b\} \cup \{O\}$$
where O is called the "point at infinity" (we will get a better grasp on this point later).
- Some curves have features called singular points. Curves with no singular points are called smooth curves.
- We will focus on smooth curves over large finite fields.

Group Law

What makes this seemingly arbitrary cubic curve worth studying? The main reason is that their points can be given a natural addtition operation that turns them into an abelian group.

(We will use $\oplus$ to denote group addition and $\ominus$ to denote the additive inverse.)

A rudimentary description of the addition operation is as follows: Given points P and Q on the curve, we can draw the line through P and Q, which intersects the curve at a third point, R. Then the reflection of R over the x-axis also lies on the curve, which we call $P \oplus Q$. This process is illustrated in the diagram below, where we label $R = \ominus (P \oplus Q)$ for reasons that will make sense soon.



We have used $\mathbb{R}$ instead of a finite field in the illustration to make things easier to visualize.

Exercise: Convince yourself that the constructions in this section, such as "line" and "reflection over the x-axis" have natural analogues in $\mathbb{F}^2$ instead of $\mathbb{R}^2$, where $\mathbb{F}$ is a finite field.

We say that the above definition is rudimentary because it fails to deal with certain edge cases:

1. What if the line PQ is vertical and thus doesn't intersect the curve again?
2. What if $P = O$ or $Q = O$?
3. What if $P = Q$?

Naturally, the first two objections can be dealt with by stipulating that O lies on all vertical lines (and is its own reflection over the x-axis).

To address the third objection, we say that if $P = Q$, the relevant line through P and Q is the tangent to the curve at P. (If additionally $P = O$, we can perhaps say that the line is the vertical line at infinity and actually intersects the curve a third time at O.)

Now, it is true that any line which intersects the curve at ≥ 2 points actually intersects the curve at exactly 3 points (including multiplicity, i.e. a tangency point counts as at least 2 intersections).

Exercise: (Optional) Make the above claim rigorous, and prove it.

This is a lot to take in, but a clean way to remember it is as follows:

- We are giving the set E of points on the curve (including the point O at infinity) an additive structure that turns E into a group.
- O will be the identity element of the group.
- For any line intersecting the curve at 3 points, we stipulate that the sum of these points is the identity.
  - We say that O lies on all vertical lines.

In particular, not only does the "intersect and reflect" method for addition of generic points fall out of the above, edge cases are also covered.

To spell things out:

- The reflection P' of P over the x-axis is the inverse of P because some vertical line passes through O, P, P', meaning $P \oplus P' = O \oplus P \oplus P' = 0$.
- To add points P and Q, you find the third intersection of line PQ with the curve, which should be $\ominus (P \oplus Q)$ because the sum of any three collinear curve points is 0.
- Then its reflection over the x-axis is $P \oplus Q$.

Why does the addition operation we have just defined induce a group structure?

- The axioms of identity and inverse follow immediately from our definition.
- Closure follows from the fact (which we stated without proof) that any line which intersects the curve at ≥ 2 points actually intersects the curve at exactly 3 points (including multiplicity).
- Associativity is the most difficult to show, but it follows from the Cayley–Bacharach theorem.

Exercise: (Optional) Learn the statement of the Cayley-Bacharach theorem and use it to prove that the addition operation defined above is associative.

Thus, E can be turned into a group with the addition operation defined above. We can also see that this group is abelian because its addition operation does not depend on the order of P and Q.

In practice, elliptic curve addition can be computed by applying a certain rational function to the coordinates of the addends, which we supply but will not prove.

Given distinct points $P = (x_P, y_P)$ and $Q = (x_Q, y_Q)$ that we wish to add, first compute the slope $\lambda = \frac{y_Q - y_P}{x_Q - x_P}$ of line PQ. Then

$$(x_P, y_P) \oplus (x_Q, y_Q) = (x_R, y_R) = (\lambda^2 - x_P - x_Q, \lambda x_P - \lambda x_R - y_P).$$

To add a point P to itself, use the above formula with $(x_Q, y_Q) = (x_P, y_P)$, and instead use $\lambda = \frac{3x_P^2 + a}{2y_P}$.

The edge case where $P = O$ or $Q = O$ can be handled separately.

The formulas above are not conceptually important, but their existence means that computers can add elliptic curve points reasonably efficiently.

Summary.

- The set E of points on the curve can be given an additive structure, which turns E into an abelian group.
- Geometrically, we can think of the addition operation as follows:
  - To add points P and Q, find the third intersection of line PQ with the curve E, and reflect this intersection over the x-axis, to get another point on the curve, which we say is $P + Q$.
  - As stated, this operation is undefined over certain edge cases, which we can fix with some care.
- This addition can be computed reasonably efficiently in terms of the coordinates of P and Q.

Projective coordinates

Feel free to skip this section if you have never seen projective coordinates used elsewhere; it's not written well enough to be a first introduction to them.

Recall that E is the set $\{(x, y) \in \mathbb{F}^2 \mid y^2 = x^3 + ax + b\} \cup \{O\}$ (now endowed with group addition).

We are going to put the curve in the projective plane.

- We homogenize the curve equation and write instead

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

  so that the equation is homogenous of degree 3 in x, y, z.
- We define the following equivalence relation on $\mathbb{F}^3 \setminus \{(0,0,0)\}$:
  - $(x_1, y_1, z_1) \sim (x_2, y_2, z_2)$ iff there exists $k \in \mathbb{F}$ such that $(x_2, y_2, z_2) = (kx_1, ky_1, kz_1)$.
  - The collection $(\mathbb{F}^3 \setminus \{(0,0,0)\})/\sim$ of equivalence classes is called the projective plane.
- Note if two points are in the same equivalence class, then one satisfies the homogenous equation if and only if the other does.
- Now we can define $E_{\text{proj}} \subseteq (\mathbb{F}^3 \setminus \{(0,0,0)\})/\sim$ to be the collection of equivalence classes whose points satisfy the homogenous equation.
- The points of E and E_{proj} are in a simple bijection:
  - $(x, y) \in E$ corresponds to the equivalence class $[(x, y, 1)]$ in E_{proj}.
  - $O \in E$ corresponds to the equivalence class $[(0, 1, 0)]$ in E_{proj}.

In particular, by extending the plane $\mathbb{F}^2$ to the projective plane, we gain an explanation for why the point at infinity belongs to our curve.

Another thing to note is that writing points in projective coordinates gives us a more efficient way to compute group addition:

Given points $P = (x_P, y_P, z_P)$ and $Q = (x_P, y_P, z_P)$ on the curve, we can compute their sum $R = (X_R, Y_R, Z_R)$ given by

$$\begin{aligned}
X_R &= (X_P Z_Q - X_Q Z_P)\Big(Z_P Z_Q (Y_P Z_Q - Y_Q Z_P)^2 - (X_P Z_Q - X_Q Z_P)^2 (X_P Z_Q + X_Q Z_P)\Big) \\
Y_R &= Z_P Z_Q (X_Q Y_P - X_P Y_Q)(X_P Z_Q - X_Q Z_P)^2 \\
&\quad - (Y_P Z_Q - Y_Q Z_P)\Big((Y_P Z_Q - Y_Q Z_P)^2 Z_P Z_Q - (X_P Z_Q + X_Q Z_P)(X_P Z_Q - X_Q Z_P)^2\Big) \\
Z_R &= Z_P Z_Q (X_P Z_Q - X_Q Z_P)^3.
\end{aligned}$$

Although these expressions look much scarier, they can be computed without taking a multiplicative inverse in $\mathbb{F}$ (an operation which is usually many times more expensive than multiplication), so this method of addition is somewhat faster than that shown at the end of the previous section .

Summary.

- By putting the curve in the projective plane, we can better explain the point at infinity.
- We can also do arithmetic on curve points faster.

Order of the group

How many points are in the group E? If E is over a finite field $\mathbb{F}_q$ with q elements, we might use the following heuristic as an estimate: There are q^2 pairs (x, y) of points, and each of them has a "$1/q$ probability" of satisfying the equation $y^2 = x^3 + ax + b$ in $\mathbb{F}_q$. (And there's also the point at infinity.) Based on this heuristic, we might expect that a typical curve over $\mathbb{F}_q$ has approximately $q + 1$ points.

Indeed, while the order N of the group depends on not only q but also a and b, the Hasse-Weil bound (which we will not prove yet) guarantees that it satisfies

$$|N - (q + 1)| \leq 2\sqrt{q}$$

regardless of a and b.

As we will see later, cryptographers are mostly interested in curves where the order N of the group is prime, or at least has a very large prime factor. This is because we often want to find a group (or subgroup) which has prime order.

There is no known way to choose parameters a and b that provides fine control over N, so in practice, finding cryptographically desirable elliptic curves often requires a lot of searching. (Even more so due to the fact that cryptographers typically require several more properties of an elliptic curve than just near-primeness of the group order.)

Summary.

- The order of the elliptic curve group given by the solutions in $\mathbb{F}_q^2$ to $y^2 = x^3 + ax + b$ depends on q, a, and b, but is approximately q.
- Selecting appropriate curves for cryptography takes a lot of computation and care, because one often needs the group order to be prime or divisible by a large prime, among other properties.

Discrete logarithm, and why cryptopgraphers like elliptic curves

Almost all cryptographic protocols rely on hardness assumptions. That is, we prove statements of the form, "If it is computationally feasible for an adversary to break the security of our protocol, then it is computationally feasible for that adversary to compute the answer to this other problem." The other problem is usually a self-contained, well-established problem, whose hardness we can be confident in because motivated attackers have been trying and failing to solve it for many years.

One of the longest-standing cryptographic assumptions is that the discrete logarithm is computationally hard. Informally, this assumption states that given a large prime p, a generator g of the multiplicative group $\mathbb{Z}_p^\times$, and given a random element $h \in \mathbb{Z}_p^\times$, it is computationally infeasible to compute an integer x such that $g^x = h$.

For example, for carefully-chosen primes p around the size of 2^{3072}, the discrete logarithm over the cyclic group $\mathbb{F}_p^\times$ offers about 128 bits of security—that is, the best known algorithms take 2^{128} operations to have a reasonable probability of finding the discrete logarithm of a given random h. This level of security is enough for all purposes today. Arithmetic modulo a prime about as large as 2^{3072} is expensive but easily within reason for modern computers.

We can also consider the same problem over elliptic curve groups (or their subgroups) that happen to be cyclic. Given a generator P of the group and a random element Q, the discrete logarithm problem is that of finding an integer x such that

$$xP = \underbrace{P + P + \cdots + P}_{x \text{ times}} = Q.$$

(We still call it the discrete logarithm problem even though the elliptic curve group uses additive notation.)

We can achieve the same 128-bit security level on a carefully-chosen elliptic curve group with prime order around only 2^{256}. The group operation on such a curve can be considerably faster to compute than the group operation of $\mathbb{Z}_p^\times$ (for $p \approx 2^{3072}$), so cryptographers often use elliptic curve groups instead of the multiplicative group of a large prime when they need the discrete logarithm assumption.

Summary.

- Many cryptographic protocols rely on the discrete logarithm assumption, that computing the logarithm of a random element of a large group is computationally infeasible.
- We can use a smaller group to achieve the same level of infeasibility if we use an elliptic curve group instead of a multiplicative group $\mathbb{Z}_p^\times$.
- So, elliptic curves are often the group of choice for protocols using the discrete logarithm assumption.