# 1 Elliptic Curves

## 1.1 Geometry Intuition

Before we go on with more theoretical topics, let us examine another application of the quite abstract theory of algebra we have studied in the previous chapter, namely elliptic curves cryptography. It should be mentioned that linear algebra is by far not the only topic relevant in this context. In fact, important aspects of the theory of elliptic curves require the understanding of function analysis – where it actually comes from – but here we will focus on algebra and group theory.

Elliptic Curves (EC) provide the mathematical background for variants of public key cryptography. This kind of cryptography is being developed since the eighties, but it took a while until it was accepted by industry. It is the main public key cryptographic scheme today. Its acceptance was accelerated by the smartphone boom. In smartphones and other devices with restricited resources, classic cryptographic schemes are not very practical. Their drawback is the computational overhead resulting from key size. Many cryptoanalytical attacks on classic cryptography are known that force cryptographers to use huge keys. To achieve 128-bit security with RSA, we need keys with at least 3072 bits. The same level of security can be reached with EC cryptography, according to known attacks today, with 256 bits. A huge improvement!

EC cryptography is different from classic cryptography in various respects. First, it includes much more math. That is to say, it does not include theory from only one or two branches of mathematics like number theory in classic cryptography, but from many different branches. This has huge impact on cryptoanalysis. Hidden attacks may linger around in apparently remote fields of mathematics that we did not account for. However, the theory surrounding EC is very well understood today and, as said, it is the mainline cryptography today.

Second, the basic means, especially the group we need for public key cryptography, are much more "engineered" than in classic cryptography. Classic schemes are based mainly on modular arithmetic, which was well known centuries before anyone thought of cryptography. The groups found in modular arithmetic, in particular the multiplicative group, was then used to define cryptographic tools. In elliptic curves, there are no such groups "by nature". They are constructed on the curves with the purpose to use them in cryptography. Therefore, EC may sometimes feel a bit artificial. It is important to understand that the group we define on the curves is defined voluntarily according to our purpose. When we speak of *point addition* in this context, one must not confuse this operation with the arithmetic operation of addition. It is something totally different.

Anyway, what are elliptic curves in the first place? Elliptic curves are polynomials that were intensively studied in the late $19^{th}$ century, especially by German mathematician Karl Weierstrass (1815 – 1897), who was of huge importance in the sound fundamentation of analysis. We will meet him again in the third part. He studied polynomials of the form

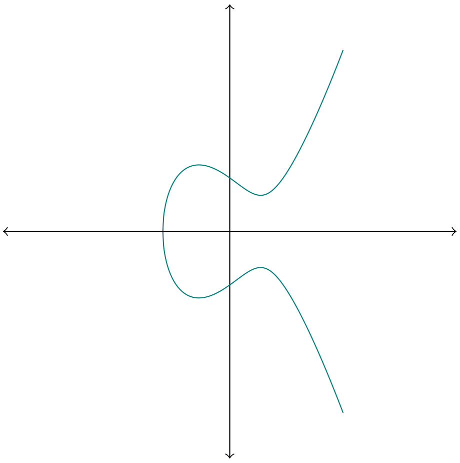
$$y^2 = x^3 + ax + b, \tag{1.1}$$

which is said to be in *Weierstrass form*. Obviously, we can easily transform this equation into a form that looks more like something that can be computed, namely:

$$y = \sqrt{x^3 + ax + b}. \tag{1.2}$$

But be careful! Weierstrass polynomials are not functions, at least not in $\mathbb{R}$, since there is not exactly one $y$ for each $x$. When the expression $x^3 + ax + b$ becomes negative, there is, in the world of real numbers, no solution for the right-hand side of the equation. That is, there is no $y$ for $x$ that cause that expression to be negative.

This is quite obvious, when we look at the geometric interpretation of that polynomial. It looks – more or less – like in the following sketch:

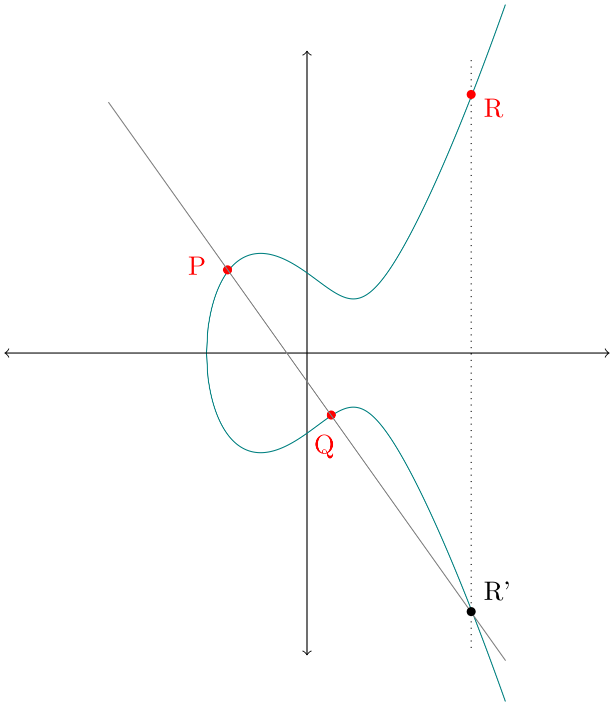


The exact shape depends on the coefficients $a$ and $b$. The bubble on the left may sometimes be a circle or ellipse completely separated from the “tail” on the right; it may, in other cases, be less clearly distinguished from the tail on the right, forming just a tiny bulge in the tail.

Anyhow, the point is that the curve "ends" on the left-hand side for some $x < 0$. More precisely, it ends where $x^3$, for a negative value, becomes greater than $ax + b$, because, then, the whole expression becomes negative and there is no real square root for it anymore.

We will now start to construct a group on this kind of curves. We call it an *additional group*, but be aware that this is not addition in the sense of the arithmetic operation. It has nothing to do with that! Is a way to combine points with each other that can be captured in a – more or less – simple formula. We will start by giving a geometric interpretation of this operation. This will help getting an intuition. But, again, be aware that we are not dealing with geometry. We will soon deviate from geometry and talk about curves in a quite abstract way.

The following sketch shows an elliptic curve with three points $P$, $Q$ and $R$, all coloured in red. These points are in the relation $P + Q = R$. The meaning of this operation is indicated by the lines:



When adding two points $P$ and $Q$ on an elliptic curve, we draw a straight line through them (the grey one). From the nature of the elliptic curve, it is obvious that the straight line will meet the curve once again. At that cross point, we draw a helper point, $R'$. Then we draw another line (the dotted one) that goes straight up crossing $R'$. This line will meet the curve again, namely at a point with the same $x$ coordinate, but with the inverse of the $y$ coordinate $-y$. That point is $R$, the result of $P + Q$.

You see that this operation has in fact nothing to do with arithmetic addition. It is an arbitrary construction to relate three different points. Nevertheless, it is carefully designed to give rise to a group based on this operation, as we will see later.

For the moment, our main question is how can we compute $R$ from $P$ and $Q$. We start by computing the straight line. A straight line is defined by a formula of the form

$$y = mx + c, \tag{1.3}$$

where $m$ is the slope and $c$ the $y$-intercept. What we need to do now is to find the third point, $R'$, which, like $P$ and $Q$, lies on both, the straight line and the elliptic curve. To find such a point, we set the two formulas equal. Since an elliptic curve is defined as

$$y^2 = x^3 + ax + b, \tag{1.4}$$

we can derive

$$(mx + c)^2 = x^3 + ax + b. \tag{1.5}$$

By subtracting $(mx + c)^2$ on both sides, we get

$$x^3 + ax + b - (mx + c)^2 = 0. \tag{1.6}$$

Using the binomial theorem we can expand this to

$$x^3 - m^2x^2 - 2mxc - c^2 + ax + b = 0. \tag{1.7}$$

We know already two points, where this equation is fulfilled, namely $x_P$ and $x_Q$. We also know that these values are roots of the above equation. We can hence use them for factoring that equation into $(x - x_P)(x - x_Q)\Psi$, where $\Psi$ is yet another factor. But we know even more. We just have to look at the sketch above to see that there are three roots and, hence, three factors. The third root is at $x_{R'}$, so we can conclude that

$$x^3 - m^2x^2 - 2mxc - c^2 + ax + b = (x - x_P)(x - x_Q)(x - x_{R'}). \tag{1.8}$$

From here it is quite simple. We just apply the trick of the *opposite sum of the roots* and get

$$m^2 = x_P + x_Q + x_{R'}, \tag{1.9}$$

which we can easily transform to obtain

$$x_{R'} = m^2 - x_P - x_Q. \tag{1.10}$$

Since $R$, the point we are finally looking for, is the reflection of $R'$ across the $x$-axis, we have $x_R = x_{R'}$, *i.e.* the points have the same the $x$-coordinate.

Computing $y_{R'}$ is again quite simple. We use the straight line for this. Since it is a straight line, the $y$-values increase at a constant rate. So, we compute the value it should increase on the segment between $x_P$ and $x_R$, which is $m(x_R - x_P)$ and add this to the already known $y$-value at point $P$:

$$y_{R'} = y_P + m(x_R - x_P). \tag{1.11}$$

But hold on! This is $y_{R'}$. What we are looking for is $y_R$, the reflection of $y_{R'}$ across the $x$-axis. But that is simply $(x, -y)$. Alternatively, we can compute that value directly by rearranging the equation to

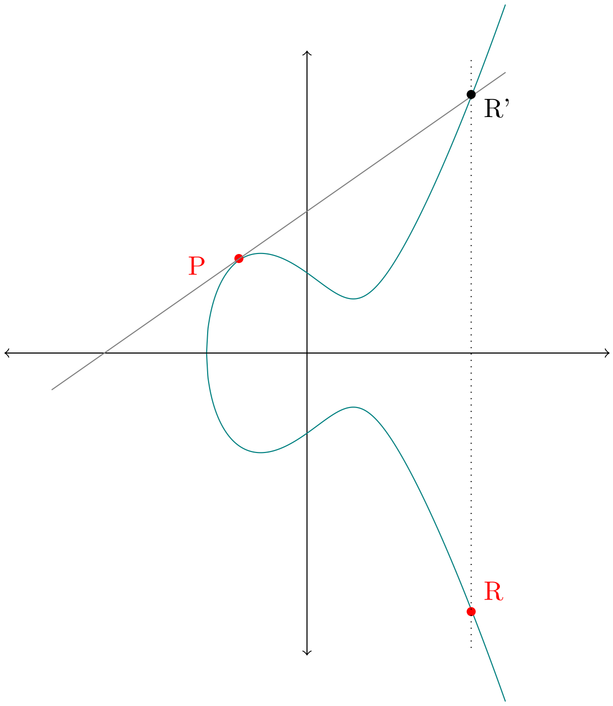
$$y_R = m(x_P - x_R) - y_P. \tag{1.12}$$

The final piece missing now is the slope, which we are using all the time. The slope, as we know, can be expressed as a fraction:

$$m = \frac{y_Q - y_P}{x_Q - x_P}. \tag{1.13}$$

With this equation, however, we get into trouble. Everything is fine, when we assume that we add two distinct points $P$ and $Q$. But if we had $P = Q$, *i.e.* if we want to add a point to itself, then the denominator of the above fraction becomes negative. That, clearly, is to be avoided.

To avoid that, we use, instead of a secant line that intersects the curve, the tangent line at point $P$, which, as we already know, measures the slope of the curve at $P$. Geometrically, this corresponds to the following sketch:



Here, we draw the tangent line in P. Where the tangent line intersects the curve again, we draw the helper point $R'$ and reflect it across the $x$-axis to obtain the point $R = P + P = 2P$.

As you hopefully remember, the slope of a curve at a given point can be calculated with the derivative of that curve. We will apply that derivative trick to get the tangent line at $P$. This task, however, is a bit more difficult than for the trivial cases we have seen so far. Until now, we have seen derivatives of simple functions of the form $f(x) = x^2$, whose derivative is $f'(x) = 2x$ and so on. Now, we have the equation

$$y^2 = x^3 + ax + b. \tag{1.14}$$

We can interpret this equation as an application of two different functions. The first function, say, $g$, is $g(x) = x^3 + ax + b$. The second function, $f$, is $f(x) = \sqrt{x} = x^{\frac{1}{2}}$.

For such cases, we have the *chain rule*, which we will discuss more thoroughly in part 3. The chain rule states that the derivative of the composition of two functions is

$$(f \circ g)' = (f' \circ g) \times g'. \tag{1.15}$$

That is, the derivative of the composition of two functions $f$ and $g$ is the composition of the derivative of $f$ and $g$ times the derivative of $g$. So, let us figure out what the derivatives of our $f$ and $g$ are. The derivative of $g$ is easy:

$$g'(x) = 3x^2 + a$$

A bit more difficult is $f'$. If $f(x) = x^{\frac{1}{2}}$, then $f'$ is

$$\frac{1}{2}x^{\frac{1}{2}-1} = \frac{1}{2}x^{-\frac{1}{2}} = \frac{x^{-\frac{1}{2}}}{2},$$

which is

$$f'(x) = \frac{1}{2x^{\frac{1}{2}}}.$$

Now, we apply this to the result of $g(x)$. The result of $g(x)$ is $y^2$. If we plug $y^2$ into the equation above, we get

$$\frac{1}{2y^{2\times\frac{1}{2}}} = \frac{1}{2y}.$$

This, we now multipy by $g'$ and finally get

$$\frac{3x^2 + a}{2y}.$$

When we use this formula for $x = x_P$, we get the formula to compute $m$:

$$m = \frac{3x_P^2 + a}{2y_P}. \tag{1.16}$$

So, we finally have an addition formula that covers both cases, $P \neq Q$ and $P = Q$:

$$x_R = \begin{cases} m^2 - x_P - x_Q & \text{if } x_P \neq x_Q \\ m^2 - 2x_P & \text{otherwise} \end{cases} \tag{1.17}$$

and

$$y_R = m(x_P - x_R) - y_P, \tag{1.18}$$

where

$$m = \begin{cases} \frac{y_Q - y_P}{x_Q - x_P} & \text{if } x_P \neq x_Q \\ \frac{3x_P^2 + a}{2y_P} & \text{otherwise.} \end{cases} \tag{1.19}$$

## 1.2 Projective Geometry

## 1.3 EC modulo a Prime

## 1.4 EC Crypto Systems

## 1.5 Cryptoanalysis

## 1.6 Mr. Frobenius

## 1.7 Mr. Schoof

## 1.8 Mr. Elkies and Mr. Aktin

## 1.9 EC in Practice