Lecture Notes, Week 5

Math 480A2: Mathematics of Blockchain Protocols, Fall 2022

Lecturer: Bryan Gillespie

Elliptic Curves

We next describe the background and construction of *elliptic curve groups*, a class of abelian groups which is the core object of an important area of modern cryptography.

Definition 1. An **elliptic curve** is a smooth projective plane curve of genus 1 with at least one point over its underlying field.

This definition is not especially comprehensible without some significant background in algebraic geometry. For our purposes, we can adopt the following definition, which can be derived from the previous one.

Definition 2 (Short Weierstrass Form)**.** Let K be a field of characteristic not equal to 2 or 3. An **elliptic curve** E over K is the set of solutions of a polynomial equation of the form

$$(\operatorname{char} K \neq 2, 3) \qquad y^2 = x^3 + ax + b, \quad a, b \in K, \quad 4a^3 + 27b^2 \neq 0$$

The form of the equation in the above definition is a reduction from a general form which is valid over fields in which the numbers 2 and 3 are nonzero. For fields of these small characteristics, more general equations are required.

Definition 3 (Long Weierstrass Forms)**.** Let K be any field, and for values $a_1, a_3, a_2, a_4, a_6 \in K$, define

$$\begin{aligned} b_2 &= a_1^2 + 4a_4, & b_4 &= 2a_4 + a_1a_3, \\ b_6 &= a_3^2 + 4a_6, & b_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 - a_4^2 \end{aligned}$$

An **elliptic curve** E over K is the set of solutions of a polynomial equation of the form

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

where we require the following condition on the coefficients to hold:

$$-b_2^2b_8 - 8b_4^3 - 27b_6^2 + 9b_2b_4b_6 \neq 0$$

If K has characteristic larger than 2, then the following simpler equation is equivalent:

$$(\operatorname{char} K \neq 2) \qquad y^2 = 4x^3 + b_2x^2 + b_4x + b_6$$

with the same restriction on values of b_2, b_4, b_6, and b_8.

Due to the complexity of these more general expressions, we will restrict our attention to the short Weierstrass form, and assume that we are working over fields of characteristic not equal to 2 or 3.

Example 4. The following cubic equations give representative examples of the different "shapes" that the graph of a short Weierstrass elliptic curve equation can have over the rational numbers.

$$\begin{aligned} y^2 &= x^3 - x & \qquad y^2 &= x^3 \\ y^2 &= x^3 - x + 1 & \qquad y^2 &= x^3 - 3x + 2 \end{aligned}$$

The equations in the left column represent the shapes of valid elliptic curves, where the condition $4a^3 + 27b^2 \neq 0$ is satisfied, so the curve is "smooth". The first has two connected components, one of which is bounded and the other of which is unbounded. The second has only a single unbounded component.

The second column demonstrates the two failure modes of the short Weierstrass equation. Both equations satisfy $4a^3 + 27b^2 = 0$. The first represents a "cuspidal" cubic, which has a pointed "cusp" singularity at the origin. The second represents a "nodal" cubic which has a "node" singularity at the point $(1, 0)$, where the curve smoothly passes through the point in two separate directions. Both of these possibilities are avoided by the restrictions placed on the coefficients.

A group structure may be associated with the points of an elliptic curve in the following way.

Definition 5. Let K be a field, let E be an elliptic curve over K, and let $G = E \cup \{O\}$, where O is an additional point called the **point at infinity**. For $P, Q \in G$, let L be the line connecting P and Q, with the following edge cases:

- If $P, Q \in E$ with $P = Q$, then L is the line tangent to E at P
- If $P \in E$ and $Q = O$, then L is a vertical line passing through P
- If $P = Q = O$, then L is a special **line at infinity**

We define $P * Q$ to be the third point of intersection of L with G, with the following edge cases:

- If L is a tangent line to E at a point which is not an inflection point, then this point is counted twice as an intersection point
- If L is a tangent line to E at an inflection point, then this point is counted three times as an intersection point
- If L is a vertical line, then one of the intersection points of L with G is the point at infinity
- If L is the line at infinity, then all three of the intersection points of L with G are the point at infinity

We define the group operation "+" by taking $P + Q$ to be the reflection of $P * Q$ across the x-axis. Here, we let the reflection of O across the x-axis again be O.

Note that in the above, the definition makes use of the notion of a "tangent" line for the sum of a point with itself. While this definition relies on the geometric structure of a curve in the real plane $\mathbb{R}^2$, the algebraic representation of this construction (e.g. using the gradient of the defining equation) makes sense for more general fields.

The steps for computing the sum of points in an elliptic curve group algebraically are given in detail in Algorithm 1 for the case of fields of characteristic not equal to 2 or 3. The algorithm is similar for elliptic curve groups over binary or ternary fields, except that the algebraic expressions must be derived from the more general Weierstrass form of the defining equation.

Algorithm 1 Computing the sum of two points in an elliptic curve group

Precondition: E defined by $y^2 = x^3 + ax + b$ over a field K with $\operatorname{char} K \neq 2, 3$

```
function Sum($P, Q$) → $E$
    if $P = \mathcal{O}$ then                                  ▷ $P = \mathcal{O}$
        return $Q$
    else if $Q = \mathcal{O}$ then                             ▷ $Q = \mathcal{O}$
        return $P$
    $(x_P, y_P) \leftarrow P$
    $(x_Q, y_Q) \leftarrow Q$
    if $x_P = x_Q$ and $y_P = -y_Q$ then                       ▷ $Q = -P$
        return $\mathcal{O}$
    else
        if $x_P = x_Q$ then                                    ▷ $Q = P$
            $m \leftarrow (3x_P^2 + a)/(2y_P)$
        else                                                   ▷ $Q \neq \pm P$
            $m \leftarrow (y_Q - y_P)/(x_Q - x_P)$
        $x_R \leftarrow m^2 - (x_P + x_Q)$
        $y_R \leftarrow y_P + m(x_R - x_P)$
        return $(x_R, -y_R)$
```

Note that the parameter m used in several expressions in Algorithm 1 represents the slope of a line: either the line connecting P and Q if P and Q have different x-coordinates, or the line tangent to E at P if $P = Q$.

The elliptic curve group operation satisfies the following properties.

Proposition 6. *As defined above, $(G, +)$ is an abelian group with identity element O satisfying the relation $P + Q + R = O$ for any three points which are the intersection points (with multiplicity) of a line with G.*

Proof. The abelian property follows because the line L associated with the sum $P+Q$ doesn't depend on the order of the defining points.

To see that O is an identity element, note first that clearly $O + O = O$ by the definitions above. If $P \in E$, then the line connecting P and O is a vertical line through P. Thus if

$P = (x_0, y_0) \in E$ with $y_0 \neq 0$, then the third intersection of L with G is $P' = (x, -y)$, and the reflection of P' across the x-axis is again P. This demonstrates that $P + O = P$.

If $y_0 = 0$, then the vertical line through P is tangent to E. To see this, we compute the gradient of $f(x, y) = y^2 - x^3 - ax - b$ at P:

$$\Delta f = (\partial f/\partial x, \partial f/\partial y) = (-3x^2 - a, 2y)$$

At P, this evaluates to $(-3x_0^2 - a, 0)$, which is the normal vector to a vertical line in the case that $-3x_0^2 - a$ is nonzero.

To see that this is the case, suppose that this value is equal to 0, and apply the relation $x_0^2 = -a/3$ in the formula for E:

$$0 = (x_0^3) + ax_0 + b = (-a/3)x_0 + ax_0 + b = (2a/3)x_0 + b$$

This implies $x_0 = -3b/(2a)$. But substituting this back into the original expression for x_0, we obtain

$$-3x_0^2 - a = -27b/(4a^2) - a = 0$$

This expression can be rearranged to conclude that $4a^3 + 27b^2 = 0$, which contradicts the condition assumed on the coefficients a and b for the Weierstrass form of an elliptic curve.

From the definitions, it is straightforward to see that for $P \in G$, the inverse element $-P$ is given by the reflection of P across the x-axis. This remains true if P is on the x-axis, as we have seen that E has a vertical tangent line at such a point, and at $P = O$ as the reflection of O is defined to be O.

From these two properties, the relation $P + Q + R = O$ follows directly. Associativity is a complicated algebraic or geometric argument which we will leave as an exercise for the (determined) reader. □

It is worth remarking that the definition of the elliptic curve group involves a bunch of cases and a strange point at infinity. The addition rule can be phrased more simply with a few more concepts from algebraic geometry which we won't cover in full detail, but will describe in general terms with a few examples.

First, we explain the mysterious "point at infinity". An elliptic curve can be more completely described using so-called *projective* solutions to *homogeneous* polynomial equations.

Definition 7. Let K be a field, and let $f \in K[x, y]$ be a polynomial with total degree d. The **homogenization** of f is the polynomial in $K[x, y, z]$ obtained by multiplying each monomial $ax^{d_1}y^{d_2}$ by $z^{d-(d_1+d_2)}$. Then the homogenization of f is a homogeneous polynomial of total degree d.

If (x_0, y_0, z_0) is a solution to a homogeneous equation, then any scalar multiple (rx_0, ry_0, rz_0) is as well, and in particular the origin is always a solution for a non-constant homogeneous equation. The solution can thus be described by *projective* coordinates.

Definition 8. Let K be a field. Then the relation $\sim$ on $K^n \setminus 0$ given by

$$(x_1, \ldots, x_n) \sim (y_1, \ldots, y_n) \quad \iff \quad (y_1, \ldots, y_n) = (rx_1, \ldots, rx_n), r \in K \setminus \{0\}$$

is an equivalence relation. The equivalence class of $(x_1, \ldots, x_n)$ in $(K^n \setminus 0)/\sim$ is written as $(x_1 : x_2 : \cdots : x_n)$.

Any solution (x, y) to a polynomial equation $f(x, y) = 0$ can be thought of as a solution $(x : y : 1)$ to the projective (homogeneous) equation $g(x, y, z) = 0$, where g is the homogenization of f. However, there might be additional solutions to the projective equation having $z = 0$ which are not caught by the original non-projective equation. Such points are said to be on the "line at infinity".

Example 9. Let K be a field of characteristic not equal to 2 or 3, and let E be the elliptic curve described by $y^2 = x^3 + ax + b$, $4a^3 + 27b^2 \neq 0$. Then the homogeneous equation of E is given by

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

Aside from the points with nonzero z-coordinate, setting $z = 0$ gives the resulting equation $x^3 = 0$, which has solutions $(0, y, 0)$ for any y. Interpreting as (nonzero) projective coordinates, this is only a single *projective* point, which can be rescaled to make the y-coordinate 1, giving the single point of E on the line at infinity, $O = (0 : 1 : 0)$.

Example 10. Any two distinct lines meet in a single projective point. Let L be the line defined by the equation $ax + by + c = 0$, and let L' be the line defined by the equation $a'x + b'y + c' = 0$. Here we must have $(a, b), (a', b') \neq (0, 0)$ in order for each of the equations to describe a line.

Homogenizing these equations gives projective equations of the form $ax + by + cz = 0$, which describes a plane through the origin with normal vector (a, b, c). Note that in this formulation, it is only necessary to have $(a, b, c) \neq (0, 0, 0)$ in order to give a valid expression for a plane. This gives rise to an additional *projective* line described by $z = 0$. This is the *line at infinity* which was mentioned when defining the elliptic curve group operation.

The solutions of the system of equations $ax + by + cz = 0$ and $a'x + b'y + c'z = 0$ is the set of vectors perpendicular to both the coefficient vectors at once, which is the set of points spanned by the vector $(a, b, c) \times (a', b', c')$. This is a nonzero vector when the input vectors are not colinear, which is the case exactly when L and L' describe distinct lines. In particular, the set of solutions is a line through the origin, which corresponds to a single projective point.

This point can be written explicitly as

$$\det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a & b & c \\ a' & b' & c' \end{pmatrix} = (bc' - b'c, ca' - c'a, ab' - a'b)$$

corresponding to the projective point $(bc' - cb' : ca' - ac' : ab' - ba')$. If $ab' - ba' \neq 0$ (which occurs exactly when the original lines L, L' are not parallel), then this can be translated to a solution in our original coordinates (x, y) by rescaling by the third component:

$$(x : y : z) = \left(\frac{bc' - cb'}{ab' - ba'} : \frac{ca' - ac'}{ab' - ba'} : 1 \right)$$

Otherwise, this solution gives a point with z-coordinate 0, which corresponds with a projective solution on the line at infinity.

Next we justify in general terms the cases which came up above when defining the elliptic curve group operation that required "counting an intersection point of curves multiple times". If $f, g \in K[x, y]$ are polynomials, then it is possible to associate to an intersection point α, β with $f(\alpha, \beta) = g(\alpha, \beta) = 0$ a positive integer called the *multiplicity* of the intersection. A fully general definition of multiplicity requires nontrivial background in algebraic geometry, but in less precise terms this number counts the number of intersection points that the intersection splits into when you slightly change the coefficients of f and g (where solutions are taken in the algebraic closure of K, so that no intersection points are "missing").

For the intersection of a line with a curve represented as $f(x, y) = 0$ for a polynomial f, we can give a simpler definition.

Definition 11. Let C be a plane curve defined by a polynomial equation $f(x, y) = 0$, and let L be a line with parametric equation $x(t) = at + \alpha$, $y(t) = bt + \beta$. Then the **multiplicity** of the intersection of C and L at a point $P = (x(t_0), y(t_0))$ is defined to be the order of t_0 as a root of the polynomial $f(x(t), y(t))$.

Example 12. Let C be the curve given by $f(x, y) = x = 0$, consisting of the y-axis, and let L be the x-axis, parametrized by $x(t) = t, y(t) = 0$. Then $f(x(t), y(t)) = f(t, 0) = t$, so we have that the point $(0, 0) = (x(0), y(0))$ is an intersection of L with C of multiplicity 1.

Example 13. Let C be the curve given by $f(x, y) = x^2 + y^2 - 1 = 0$ consisting of the unit circle. If L is the x-axis, then the intersections of L with C are described by the roots of $f(t, 0) = t^2 - 1 = (t - 1)(t + 1)$. The points on L for $t = \pm 1$ thus both are intersections of multiplicity 1, and correspond to the points $(1, 0)$ and $(-1, 0)$.

On the other hand, if L is the line $y = 1$, then L intersects C as a tangent. In this case, L can be parametrized as $x(t) = t, y(t) = 1$, and we have $f(x(t), y(t)) = t^2 + 1^2 - 1 = t^2$, and so $t = 0$ is a root of order 2 of this polynomial. This corresponds to an intersection of C and L at the point $(0, 1)$ with multiplicity 2.

If L is the line $y = \sqrt{2}$, then the intersection of L with C over the reals is empty. However, we can compute $f(t, \sqrt{2}) = t^2 + 2 - 1 = t^2 + 1 = 0$. We see then that this polynomial has no real roots, but if we allow points in the complex numbers, then $t = \pm i$, and so over $\mathbb{C}$ the line L intersects C at two points with multiplicity 1, given by $(\pm i, \sqrt{2})$.

Example 14. Let C be the curve given by $f(x, y) = y - (x^3 + x) = 0$ describing the graph of the function $y = x^3 + x$, and let L be the line $y = x$ paramatrized as $x(t) = y(t) = t$. Then the intersection of L with C is described by $f(t, t) = t - (t^3 + t) = t^3 = 0$, so the point $(0, 0)$ is an intersection of C and L with multiplicity 3.

Lemma 15. *Let C be a curve described by a polynomial equation $f(x, y) = 0$, $f \in K[x, y]$, and let $P = (x_0, y_0)$ be a point on C such that the gradient $\Delta f = (f_x, f_y)$ at P is nonzero. Then the tangent line of C at P, defined by the parametric equations $x(t) = x_0 + f_y(x_0, y_0)t$, $y(t) = y_0 - f_x(x_0, y_0)t$, intersects C with multiplicity at least 2.*

Proof. Because we assume that P is a point on C, we know already that $t = 0$ is a root of $f(x(t), y(t))$, since this is $f(x_0, y_0) = 0$. To check that $t = 0$ is a multiple root, we need to verify that the formal derivative of this polynomial also has a root at 0. This can be computed by the chain rule:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(f(x(t),y(t)) &= f_x(x(t),y(t))x'(t) + f_y(x(t),y(t))y'(t)\\ &= f_x(x(t),y(t))f_y(x_0,y_0) - f_y(x(t),y(t))f_x(x_0,y_0)\end{aligned}$$

Evaluating at $t = 0$, we see that this is again a root of the formal derivative. As we have seen, this means that t^2 divides $f(x(t), y(t))$, so the multiplicity of the intersection at (x_0, y_0) is at least 2. □

The following important theorem describes the very nice result of passing to projective intersections of polynomial curves, counted with multiplicity, over an algebraically closed field. We state the result without complete definitions or proof.

Theorem 16 (Bézout's Theorem)**.** *Let K be a field with algebraically closed field extension E, and let X, Y be projective plane curves defined by homogeneous polynomials $f, g \in K[x, y, z]$ with no irreducible factor in common. Then X and Y intersect in* $\deg(f)\deg(g)$ *(projective) points with coordinates in E, counted with multiplicities.*

In the context of elliptic curves in Weierstrass form, what this amounts to is saying that: a line which intersects an elliptic curve E in two points must intersect E in a third point, where these points may be considered with multiplicity, and may include the "point at infinity", which is a projective solution with projective coordinates $(0 : 1 : 0)$. The multiple "edge-cases" described in the definition handle all of these different contingencies explicitly.