Elliptic Curves

Reinhard Lüftenegger

Mathematical Foundations of Cryptography – WT 2020/21 – IAIK, TU Graz

Overview

The Very Concrete Introduction to Elliptic Curves

What's Ahead

- **How** and **why** we can calculate with points on cubic curves.
- A hands-on approach to **elliptic curves**.

Nota Bene: For the sake of vividness, we often deal with algebraic curves over the **reals** $\mathbb{R}$. But the discussed concepts are valid in **arbitrary fields** (and thus in finite fields), if not stated otherwise.

Exposition: Cubic Plane Algebraic Curves

Definition

A plane cubic algebraic curve $\mathcal{C}$ over a field $\mathbb{F}$ is the set of points $(a, b) \in \mathbb{F}^2$ which satisfy a polynomial equation

$$f(a, b) = 0,$$

where $f(X, Y) \in \mathbb{F}[X, Y]$ is a polynomial of degree three in two unknowns.

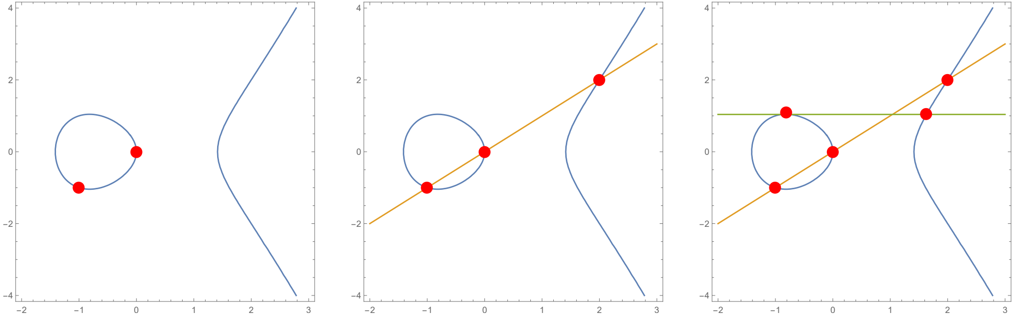
Example: Does the real polynomial $f(X, Y) = X^3 + Y^2X + X + 1$ define a curve in the above sense? What about $g(X, Y) = X^3 + X^2Y^2 + X + 1$?

From now on

- The expression "curve" always denotes a cubic plane algebraic curve.
- We assume that there is at least one point $(a, b) \in \mathbb{F}^2$ on the curve.

To put the cart before the horse...

There is a way to do arithmetic with points on **suitable** cubic curves.



Geometric Intuition: "Chord-and-Tangent-Method"

Steps Towards the Group Structure

"Doing arithmetic" means: endowing algebraic curves with a (additive) group structure.

Requirements from geometric intuition

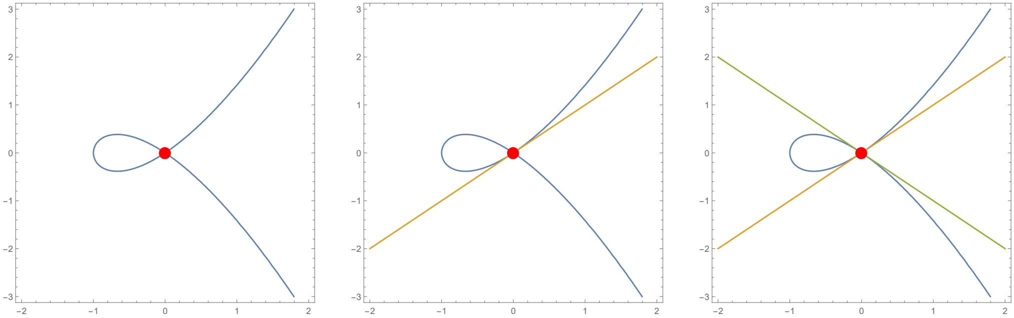
- The line through two points on the curve needs to intersect the curve in a third point, and **nowhere else**.
- Every point on the curve needs to have a **unique tangent**.

Resolutions

- Consider curves in projective space
- Non-singular curves

Example of a Non-Suitable Curve

Consider the real curve defined by $f(X, Y) = Y^2 - X^3 - X^2$:



Problem: With which tangent should we operate?

Tangents we need!

Taylor Series Expansion

Remember: A polynomial function $f : \mathbb{R}^2 \to \mathbb{R}$ in two variables has a Taylor series expansion around every point $(a, b) \in \mathbb{R}^2$.

Example: Expansion of f around (a, b) until first order terms yields

$$f_1(X, Y) = f(a, b) + f_X(a, b) \cdot (X - a) + f_Y(a, b) \cdot (Y - b).$$

Interpretations

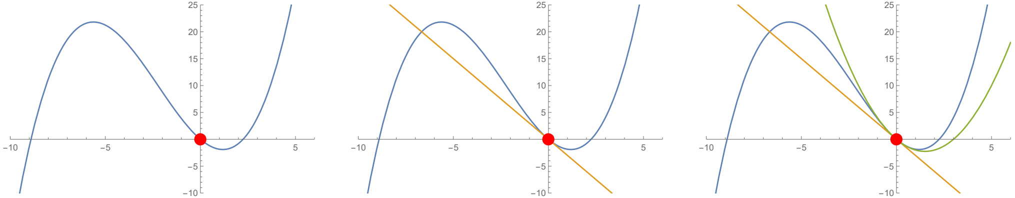
- The function f_1 can be regarded as (first-order) **approximation** of f around (a, b).
- The equation $f_1(x, y) = 0$ describes a line in $\mathbb{R}^2$, which can also be regarded as the **tangent line** at (a, b) to the curve defined by f. If it exists, it is unique.

Example: Taylor Approximation

Below figure demonstrates the first-order and second-order taylor approximation of the univariate polynomial function $f : \mathbb{R} \to \mathbb{R}$ with

$$f(x) := 0.15x^3 + x^2 - 3x$$

around the point $(0, 0)$.



(Formal) Partial Derivatives

Remember: The (first-order) partial derivative with respect to X of a real bivariate monomial $f(X, Y) = aX^nY^m$ is given by

$$f_X(X, Y) := \frac{\partial}{\partial X} aX^nY^m := \begin{cases} 0 & n = 0. \\ n \cdot aX^{n-1}Y^m & n \neq 0. \end{cases}$$

The (first-order) partial derivate of a polynomial is just the sum of the partial derivatives of its monomials.

Question: Can we "imitate" this formalism to introduce a notion of formal (first-order) partial derivatives in arbitrary fields?

Answer: Absolutely!

Example: What is the partial derivate of $f(X, Y) = Y^2 - 3XY^2 - X^3$ over $\mathbb{R}$ and $\mathbb{F}_4$ with respect to X and Y?

Non-Singular Curves and Tangent Lines

Definition

Let $\mathbb{F}$ be a field and $\mathcal{C}$ be a plane curve over $\mathbb{F}$ with defining polynomial $f \in \mathbb{F}[X, Y]$. A point $P = (a, b) \in \mathcal{C}$ is said to be singular, if

$$f_X(a, b) = f_Y(a, b) = 0,$$

otherwise it is called non-singular (or regular or smooth). The curve $\mathcal{C}$ is called non-singular if all points on the curve are non-singular. The set of points $(x, y) \in \mathbb{F}^2$ satisfying the equation

$$f_X(a, b) \cdot (x - a) + f_Y(a, b) \cdot (y - b) = 0$$

is called the tangent line to $\mathcal{C}$ at a non-singular point $P = (a, b)$.

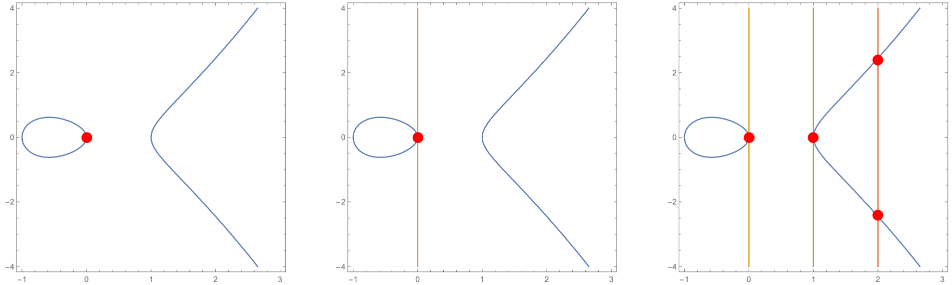
Roundup I

What we have achieved so far

- ☐ The line through two points on the curve needs to intersect the curve in a third point, and **nowhere else**.
- ✓ Every point on the curve needs to have a **unique tangent**.

Complication: Vertical Chord/Tangent Lines I

Example: Consider again the real curve defined by the polynomial $f(X, Y) = Y^2 - X^3 + X \in \mathbb{R}[X, Y]$.



Question: Do above chord/tangent lines intersect the curve in further points?

Answer: No, not in the real plane $\mathbb{R}^2$.

Complication: Vertical Chord/Tangent Lines II

What is the problem here?
For a moment, let's regard the upper part (with non-negative y-coordinate) of the real curve $y^2 - x^3 + x = 0$ as the graph of the function

$$f : \mathbb{R} \to \mathbb{R}, \quad f(x) = \sqrt{x^3 - x},$$

with derivative

$$f_X(x) = \frac{3x^2 - 1}{2\sqrt{x^3 - x}} = \frac{3 - \frac{1}{x^2}}{2\sqrt{\frac{1}{x} - \frac{1}{x^3}}}, \quad x \notin \{0, 1, -1\}.$$

Observation: As $x \to \infty$, $f_x(x) \to \infty$ as well.

In other words: In the limiting case, the curve behaves like a vertical line and is therefore **parallel** to every other vertical line.

Affine Space vs. Projective Space

Idea: Take a space, where parallel lines meet in exactly one point.

Resolution: This idea leads us to **Projective Spaces**. Roughly speaking, they extend ordinary euclidean (or affine) space with intersection points of parallel lines.

Definition

Let $\mathbb{F}$ be a field. The affine n-space over $\mathbb{F}$ is the set of all n-tuples with coordinates in $\mathbb{F}$, i.e. the set

$$\mathbb{A}^n := \mathbb{A}^n(\mathbb{F}) := \{(a_1, \dots, a_n) : a_i \in \mathbb{F}\}.$$

Remark: In light of this definition, curves with points in $\mathbb{F}^2 = \mathbb{A}^2(\mathbb{F})$ are also called affine curves.

Projective Space I

The intuition behind projective space:

Projective Space = Affine Space + Intersection points of parallel lines

Remember from school: “Coplanar parallel lines intersect at infinity”.

Consequence: All coplanar parallel lines with a given direction supposedly meet in the same point (at infinity). $\longrightarrow$ See picture on the next slide.

Twist 1: We associate with every direction of parallel lines an intersection point (= point at infinity).

Twist 2: To properly distinguish between affine points and points at infinity we need to “step up” one dimension. $\longrightarrow$ For constructing projective n-space $\mathbb{P}^n$ we need to resort to $\mathbb{A}^{n+1}$.

Projective Space II

"Quick and dirty": from $\mathbb{A}^2(\mathbb{F})$ to $\mathbb{P}^2(\mathbb{F})$

- A point $(a_1, a_2) \in \mathbb{A}^2(\mathbb{F})$ from affine space is "encoded" as $(a_1, a_2, 1)$.
- An intersection point of parallel lines = point at infinity is "encoded" as $(a_1, a_2, 0)$.
- Two points at infinity $(a_1, a_2, 0)$, $(b_1, b_2, 0)$ are equal if they represent the same direction, i.e., if there is an element $\lambda \in \mathbb{F} \setminus \{0\}$ such that $a_i = \lambda b_i$ for all i.

Nota Bene: Points in $\mathbb{P}^2$ have three coordinates. $(0, 0, 0)$ is not an element of $\mathbb{P}^2$!

The formal way to construct projective n-space $\mathbb{P}^n$ is made explicit in the next definition.

Projective Space III

Definition

Let $\mathbb{F}$ be a field. Projective n-space over $\mathbb{F}$, denoted by $\mathbb{P}^n(\mathbb{F})$, is defined as the set of all $(n+1)$-tuples $(a_1, \dots, a_{n+1})$, with $a_i \in \mathbb{F}$ and not all a_i equal to zero, modulo the equivalence relation

$$(a_1, \dots, a_{n+1}) \sim (b_1, \dots, b_{n+1}) :\Leftrightarrow a_i = \lambda b_i \text{ for some } \lambda \in \mathbb{F} \setminus \{0\} \text{ and all i.}$$

In other words, we have

$$\mathbb{P}^n(\mathbb{F}) := \{[(a_1, \dots, a_n, a_{n+1})]_\sim : (a_1, \dots, a_n, a_{n+1}) \in \mathbb{F}^{n+1} \setminus \{0\}\}.$$

Instead of $[(a_1, \dots, a_{n+1})]_\sim$ one usually writes $[a_1 : \dots : a_{n+1}]$ and calls this homogeneous coordinates. All points of the form $[a_1 : \cdots : a_n : 0]$ are called points at infinity. Projective 2-space $\mathbb{P}^2$ is also called the projective plane.

Homogeneous Polynomials and Homogenisation I

Observation: If we ask for points on the curve defined by $f(X, Y) \in \mathbb{F}[X, Y]$ in the projective plane, we encounter an obstacle:

- Two representations of a zero of f in homogeneous coordinates needn't evaluate to the same value!

Example: The evaluation of $f(X, Y) = Y^2 - X^3 + 1 \in \mathbb{R}[X, Y]$ at the projective point P given in the form $[1 : 0 : 1]$ and $[2 : 0 : 2]$.

Resolution: We homogenise our defining polynomial f. But why does this help?

Remember: A homogeneous polynomial $f(X, Y, Z) \in \mathbb{F}[X, Y, Z]$ of degree d has the nice property that for every $\lambda \in \mathbb{F}$ it holds

$$f(\lambda X, \lambda Y, \lambda Z) = \lambda^d f(X, Y, Z).$$

Example: What is the evaluation of $F(X, Y, Z) = Y^2 Z - X^3 + Z^3$ at $[1 : 0 : 1]$ and $[2 : 0 : 2]$?

Homogeneous Polynomials and Homogenisation II

Definition

The homogenisation (with respect to Z) of a polynomial $f \in \mathbb{F}[X, Y]$ is the polynomial $F \in \mathbb{F}[X, Y, Z]$ given by

$$F(X, Y, Z) := Z^{\deg(f)} \cdot f(\frac{X}{Z}, \frac{Y}{Z}),$$

which is a homogeneous polynomial of degree $\deg(f)$. Moreover, if $F \in \mathbb{F}[X, Y, Z]$ is a homogeneous polynomial, then the polynomial $f \in \mathbb{F}[X, Y]$ with

$$f(X, Y) := F[X, Y, 1]$$

is called the dehomogenisation (with respect to Z) of F.

Example: Homogenisation (w.r.t. Z) of $f(X, Y) = X + Y^2 - 2$ and $g(X, Y) = X^3 - Y^3$?

Culmination: (Non-singular) Projective Cubic Curves I

With our previous observations, the definition of a projective cubic curve is straightforward.

Definition

A projective cubic curve over a field $\mathbb{F}$ is the set of all points $[a : b : c] \in \mathbb{P}^2(\mathbb{F})$ which satisfy a polynomial equation

$$F(x, y, z) = 0,$$

where $F(X, Y, Z) \in \mathbb{F}[X, Y, Z]$ is a homogeneous polynomial of degree 3 in three unknowns.

Example: The polynomial $Y^2 - X^3 + X \in \mathbb{R}$ defines an affine curve over $\mathbb{A}^2(\mathbb{R})$. What is the polynomial defining the corresponding projective curve?

Example: The polynomial $F(X, Y, Z) = Y^2Z - X^3 + XZ^2 + XY^2 + X^2Y$ defines a projective cubic, but the polynomial $G(X, Y, Z) = Y^2Z + XYZ + Y^2X^2 + Z^3$ doesn't (why?).

Culmination: (Non-singular) Projective Cubic Curves II

The definition of non-singular projective cubics is straightforward as well.

Definition

Let $\mathbb{F}$ be a field and $\mathcal{C}$ be a projective cubic curve with defining homogeneous polynomial $F \in \mathbb{F}[X, Y, Z]$. A point $P = [a : b : c] \in \mathcal{C}$ is said to be singular, if

$$F_X(a, b, c) = F_Y(a, b, c) = F_Z(a, b, c) = 0,$$

otherwise it is called non-singular (or regular or smooth). The curve $\mathcal{C}$ is called non-singular, if all points on the curve are non-singular. The set of points $[x : y : z] \in \mathbb{P}^2(\mathbb{F})$ satisfying the equation

$$F_X(a, b, c) \cdot (x - a) + F_Y(a, b, c) \cdot (y - b) + F_Z(a, b, c) \cdot (z - c) = 0$$

is called the projective tangent line to $\mathcal{C}$ at $P = [a : b : c]$.

Weierstrass Normal Form (WNF) I

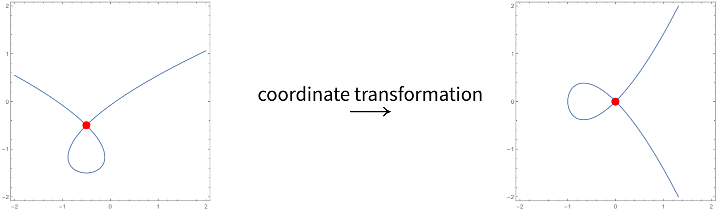
Observation: The most general equation of an affine cubic curve is given by

$$Ax^3 + Bx^2y + Cxy^2 + Dy^3 + Ex^2 + Fxy + Gy^2 + Hx + Iy + J = 0,$$

where $A, B, \ldots, J$ are coefficients in some field $\mathbb{F}$.

Question: Can we find a “nicer” equation (yielding the “same” curve) if we restrict our attention to non-singular curves?

Answer: Fortunately, yes!



Weierstrass Normal Form (WNF) II

Quintessence: The equation of a general affine cubic curve admits a **normal form** if we use the condition of non-singularity. This normal form is given by

$$y^2 + A'xy + B'y = x^3 + C'x^2 + D'x + E',$$

for some $A', B', \ldots E' \in \mathbb{F}$, and is called affine long Weierstrass (normal) form. We can even do better: if $\mathrm{char}(\mathbb{F}) \neq 2, 3$, we arrive at the so-called affine short Weierstrass (normal) form

$$y^2 = x^3 + A''x + B'',$$

for $A'', B'' \in \mathbb{F}$.

Nota Bene: We are not working out the details, but the idea behind transforming a general cubic into normal form is clear: it is just a certain change of coordinates.

Points at Infinity of Non-singular Cubic Curves

Question: By extending an affine non-singular cubic curve to projective space, how many points at infinity do we add to the curve?

Answer: There is exactly one! The justification is very easy, if we work with the Weierstrass form we've just discussed.

Sketch of the proof: We start with the homogeneous version of the long Weierstrass form

$$y^2z + A'xyz + B'yz^2 = x^3 + C'x^2z + D'xz^2 + E'z^3$$

and set $z = 0$ to obtain all intersection points at infinity. The only solution is $[0 : 1 : 0]$.

Teaser: Usually this unique point at infinity is used as the zero element for introducing the group law via the "chord-and-tangent-method" on a cubic curve.

Exercise: Check that the point at infinity $[0 : 1 : 0]$ we add to an affine non-singular cubic (in Weierstrass normal form) by extending it to projective space is non-singular as well.

Summary: Affine Curves vs. Projective Curves

Affine space $\mathbb{A}^2(\mathbb{F})$	projective completion $\rightleftarrows$ intersection	Projective space $\mathbb{P}^2(\mathbb{F})$
$\uparrow$		$\uparrow$
Affine curve over $\mathbb{F}$		Projective curve over $\mathbb{F}$
$\updownarrow$		$\updownarrow$
$f(X, Y) = 0$	homogenisation $\rightleftarrows$ dehomogenisation	$F(X, Y, Z) = 0$
$\updownarrow$		$\updownarrow$
Affine WNF		Projective WNF

Roundup II

What we have achieved so far

- ~ The line through two points on the curve needs to intersect the curve in a third point, and **nowhere else**.
- ✓ Every point on the curve needs to have a **unique tangent**.

Retardation: Intersection Points in Projective Space

Question: Can we be sure a line through two points on a curve always produces a unique third point of intersection on the curve?

Answer: Yes. But a rigorous proof involves some more concepts (like intersection multiplicity, algebraic closure, ...).

Intuitive justification: Let $\mathcal{C}$ be a projective curve over the field $\mathbb{F}$ with defining polynomial $F \in \mathbb{F}[X, Y, Z]$. The projective line through two points on $\mathcal{C}$ is described by an equation of the form

$$ax + by + cz = 0 \quad (a, b, c \in \mathbb{F}),$$

which we use to eliminate one variable in the curve equation $F(x, y, z) = 0$. Setting $z = 1$ (for affine intersections) or $z = 0$ (for intersections at infinity) yields a cubic equation in either x or y. Since we already know that two solutions lie in $\mathbb{F}$, the third one must lie in $\mathbb{F}$ (and not in the algebraic closure of $\mathbb{F}$) as well.

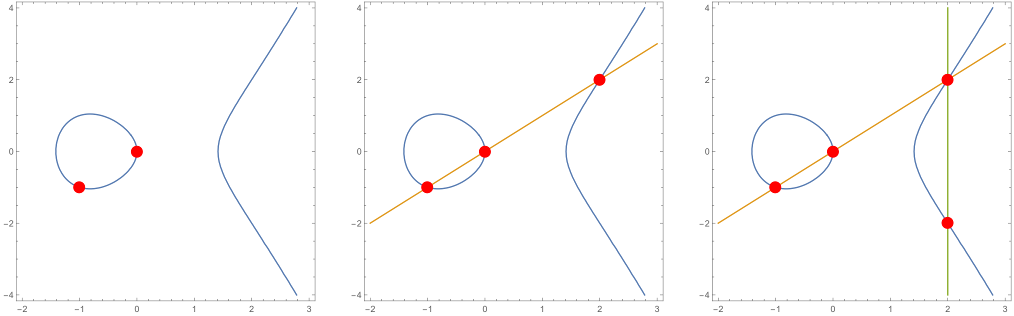
Roundup III

What we have achieved so far

- ✓ The line through two points on the curve needs to intersect the curve in a third point, and **nowhere else**.
- ✓ Every point on the curve needs to have a **unique tangent**.

“Chord-and-Tangent-Method": Revisited

All our preceding observations culminate in the following - and finally well-defined - group law on non-singular projective cubic curves.



Remark: We don’t prove the group law formally, but just to let you know: proving associativity via Weierstrass normal form is a real pain!

What’s Behind

- **How** and **why** we can calculate with points on cubic curves.
- A hands-on approach to **elliptic curves**.

Lysis: Elliptic Curves

Finally we state the following

Definition

An elliptic curve over $\mathbb{F}$ is a non-singular projective cubic curve with at least one point in $\mathbb{P}^2(\mathbb{F})$ on it.

Remarks

- We have discussed that every elliptic curve over $\mathbb{F}$ admits a long Weierstrass normal form
$$y^2 + Axy + By = x^3 + Cx^2 + Dx + E,$$
with coefficients in $\mathbb{F}$.
- Conversely, every such long Weierstrass normal form defines an elliptic curve if the coefficients A, B, C, D, E satisfy a certain condition ($\rightarrow$ discriminant of the equation).

Questions?

Questions for Self-Control

1. Explain the idea behind projective spaces. What is the main difference between affine space and projective space?
2. How is the tangent line to a point on an algebraic curve defined? How do tangent lines of real curves correlate with the taylor series expansion?
3. Sketch the group law on elliptic curves via the “chord-and-tangent-method”.
4. Which properties must hold for an algebraic curve to describe an elliptic curve? Discuss and motivate each property.
5. What is a (long) Weierstrass normal form and how is it related to elliptic curves?