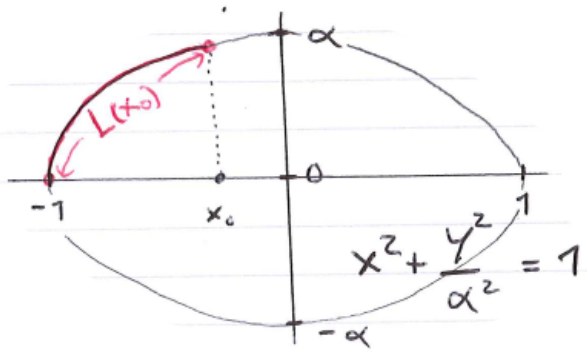
23MAC260 Elliptic Curves: Week 1

Last updated February 6, 2024

Introduction



Around 1750, Fagnano and Euler considered the problem of computing arc length on an ellipse. In the picture above, the length of the red arc is given by the integral

$$L(x_0) = \int_{-1}^{x_0} \frac{1-\beta^2x^2}{\sqrt{(1-x^2)(1-\beta^2x^2)}}\,dx \tag{1}$$

where $\beta^2 = 1-\alpha^2$.

If we put

$$y = \sqrt{(1-x^2)(1-\beta^2x^2)}$$

then Equation (1) becomes

$$L(x_0) = \int_{-1}^{x_0} \frac{1-\beta^2x^2}{y}\,dx$$

This naturally leads to considering the polynomial equation

$$y^2 = (1-x^2)(1-\beta^2x^2)$$

By an appropriate change of coordinates, this can be put in the form

$$y^2 = f(x) \tag{2}$$

where now $f(x)$ is a **cubic** polynomial.

Equation (2) defines an "algebraic curve" which is a subset of the plane. A curve defined by an equation of this special form is called an **elliptic curve**.

In this module, we will see that elliptic curves have many special properties in terms of geometry, complex analysis, and arithmetic. For this reason, they are among the most interesting and best-studied objects in modern mathematics. They have also found applications in areas such as cryptography; time permitting we will discuss these at the end of the module.

Reading List

Here are some resources that you might find useful in addition to the module materials. Bear in mind that they will all contain much more material than we cover in the module.

1. J.S. Milne, *Elliptic Curves*. Available from `https://www.jmilne.org/math/Books/ectext6.pdf`
2. J. Silverman and J. Tate, *Rational Points on Elliptic Curves*. 2nd ed., Springer Undergraduate Texts in Mathematics (2015).
3. J. Silverman, *The Arithmetic of Elliptic Curves*. 2nd ed., Springer Graduate Texts in Mathematics (2009).

1 The Projective Plane

Let K be any field (for example the rationals $\mathbb{Q}$, the real numbers $\mathbb{R}$, the complex numbers $\mathbb{C}$...)

Definition 1.1. *The* **projective plane over** K *is defined as the set of equivalence classes*

$$\mathbb{P}^2_K = \{(X, Y, Z) \mid X,\ Y,\ Z \in K,\ \textit{not all } 0\} / \sim$$

where $\sim$ *is the equivalence relation defined by*

$$(X, Y, Z) \sim (X', Y', Z') \Leftrightarrow$$
$$\textit{there exists } \lambda \in K \setminus \{0\} \textit{ such that } X = \lambda X',\ Y = \lambda Y',\ Z = \lambda Z'.$$

An alternative way to think of the projective plane $\mathbb{P}^2_K$ is the set of lines through the origin in K. The relation with the previous definition is the following: the line through (X, Y, Z) corresponds to the equivalence class containing (X, Y, Z). Note that if (X, Y, Z) and (X', Y', Z') are two vectors in K^3 that lie on the same line through $(0, 0, 0)$, then there is a scalar $\lambda \in K \setminus \{0\}$ such that $X = \lambda X'$, $Y = \lambda Y'$, and $Z = \lambda Z'$. This shows that the corresponds described above is really a bijection.

Definition 1.2. *Let* $p \in \mathbb{P}^2_K$ *be a point of the projective plane. If* (X, Y, Z) *is any representative of the equivalence class* p*, then we write*

$$p = [X, Y, Z]$$

and call this a set of **homogeneous coordinates** *for the point* p*.*

Warning: Unlike usual Cartesian coordinates in K^n, the homogeneous coordinates of a point in the projective plane are **not unique!** That is, a given point in $\mathbb{P}^2_K$ generally has more than one set of homogeneous coordinates. Why?

The equivalence relation defined in Definition 1.1 has the property that

$$(X, Y, Z) \sim (\lambda X, \lambda Y, \lambda Z)$$

for any nonzero λ. So

$$[X, Y, Z] = [\lambda X, \lambda Y, \lambda Z] \quad \forall \lambda \neq 0.$$

Warning: The expression

$$[0, 0, 0]$$

does not define a point of $\mathbb{P}^2_K$. This is because the point $(0, 0, 0)$ is not in the set on the right-hand side of the equality in Defintion 1.1. (Alternatively, the point $(0, 0, 0)$ does not uniqely specify a line in K^3.)

Examples: Here are some examples showing how different sets of homogeneous coordinates refer to the same point in $\mathbb{P}^2_K$.

$$[1, 2, 3] = [2, 4, 6]$$
$$[0, 0, c] = [0, 0, 1] \quad \forall c \neq 0.$$

2 Equations in the Projective Plane

Since homogeneous coordinates work differently from usual coordinates, *equations* in the projective plane also work differently. To make sense of equations and their zero sets, we need to restrict to equations of a particular kind:

Definition 2.1. *A polynomial*

$$F(X,Y,Z) = \sum_{i,j,k} \alpha_{ijk} X^i Y^j Z^k \qquad (\alpha_{ijk} \in K)$$

is **homogeneous** *if the total degree* $i+j+k$ *is the same for each term.*

Examples:

(i) The polynomial

$$F(X,Y,Z) = X^3 - 2XY^2 + 3YZ^2$$

is homogeneous of degree 3.

(ii) The polynomial

$$G(X,Y,Z) = X^3 - 2XY$$

is not homogeneous. (Why not?)

The importance of homogeneous polynomials is that they define **curves** in $\mathbb{P}^2_K$, as we now explain.

Proposition 2.2. *Let* $F(X,Y,Z)$ *be a homogeneous polynomial. For any* $X, Y, Z \in K$*, not all* 0*, we have*

$$F(X,Y,Z) = 0 \Leftrightarrow F(\lambda X, \lambda Y, \lambda Z) = 0 \ \forall \lambda \neq 0.$$

Proof. Say

$$F(X,Y,Z) = \sum_{i,j,k} \alpha_{ijk} X^i Y^j Z^k$$

where $i+j+k = d$ in each term. Then for any λ, we have

$$\begin{aligned} F(\lambda X, \lambda Y, \lambda Z) &= \sum_{i,j,k} \alpha_{ijk} (\lambda X)^i (\lambda Y)^j (\lambda Z)^k \\ &= \sum_{i,j,k} \alpha_{ijk} \lambda^{i+j+k} X^i Y^j Z^k \\ &= \lambda^d \sum_{i,j,k} \alpha_{ijk} X^i Y^j Z^k \\ &= \lambda^d F(X,Y,Z). \end{aligned}$$

Hence if $\lambda \neq 0$ we have $F(X,Y,Z) = 0 \Leftrightarrow F(\lambda X, \lambda Y, \lambda Z) = 0$. □

Corollary 2.3. *If* F *is homogeneous then the set*

$$V(F) = \left\{ [X,Y,Z] \in \mathbb{P}^2_K \mid F(X,Y,Z) = 0 \right\}$$

is well-defined. It is called the **curve defined by** F *or the* **vanishing set of** F.

3 Affine and Projective Space

We can view the "usual" affine plane $\mathbb{A}^2_K = K^2$ as part of the projective plane by means of the injective (one-to-one) map

$$\begin{aligned} \iota\colon \mathbb{A}^2_K &\to \mathbb{P}^2_K \\ (x, y) &\mapsto [x, y, 1]. \end{aligned}$$

If $[x, y, z]$ is a point of $\mathbb{P}^2_K$ with $z \neq 0$, then

$$[x, y, z] = [x/z, y/z, 1] = \iota(x/z, y/z).$$

So this shows

$$\begin{aligned} \mathbb{P}^2_K &= \mathbb{A}^2_K \cup \{[x, y, 0] \mid x, y \in K, \text{ not both } 0\} \\ &= \mathbb{A}^2_K \cup \mathbb{P}^1_K. \end{aligned}$$

In this context, we call the subset $\{[x, y, 0]\} \subset \mathbb{P}^2_K$ the **line at infinity**.

Remark: Actually there is nothing special about the last homogeneous coordinate: we could equally well have mapped $\mathbb{A}^2$ to $\mathbb{P}^2$ via either of the maps

$$\begin{aligned} \iota_y\colon (x, z) &\mapsto [x, 1, z] \\ \iota_z\colon (y, z) &\mapsto [1, y, z]. \end{aligned}$$

Each map gives us a decomposition of $\mathbb{P}^2$ into the disjoint union of a copy of $\mathbb{A}^2$ and a "line at infinity" which is a copy of $\mathbb{P}^1$. As an optional exercise, you can check that the images of these 3 maps cover $\mathbb{P}^2$. So we can think of $\mathbb{P}^2$ as being formed by "glueing together" 3 overlapping copies of $\mathbb{A}^2$ in a certain way.

However, in this module we will stick to the first point of view: for us, the "line at infinity" will always refer to the subset $\{[x, y, 0]\}$.

We can go back and forth between subsets of $\mathbb{A}^2_K$ and $\mathbb{P}^2_K$ by "homogenising" and "dehomogenising", as follows:

Definition 3.1. *(a) Let $f(x, y)$ be a polynomial in 2 variables. Its* **homogenisation** *$f^h(X, Y, Z)$ is obtained by inserting powers of Z to raise all terms to degree equal to the degree of f.*

(b) Let $F(X, Y, Z)$ be a homogeneous polynomial in 3 variables. Its **dehomogenisation** *$F_d(x, y)$ is obtained by setting $Z = 1$.*

Examples:

(i) If

$$f(x, y) = y^2 - x^3 - x - 1$$

then

$$f^h(X, Y, Z) = Y^2Z - X^3 - XZ^2 - Z^3.$$

(ii) If

$$F(X, Y, Z) = X^3 + Y^3 + Z^3$$

then

$$F_d(x, y) = x^3 + y^3 + 1.$$

Proposition 3.2. *(a) For any polynomial* $f(x, y)$ *we have*

$$(f^h)_d = f.$$

(b) For any homogeneous polynomial $F(X, Y, Z)$ *which is not divisible by* Z*, we have*

$$(F_d)^h = F.$$

Proof. See Problem Sheet 1. □

Examples:

(i) If

$$f(x, y) = y^2 - x^3 - x - 1$$

then

$$\begin{aligned} f^h(X, Y, Z) &= Y^2 - X^3 - XZ^2 - Z^3 \\ (f^h)_d &= y^2 - x^3 - x - 1 \\ &= f \end{aligned}$$

(ii) If

$$F(X, Y, Z) = X^3 + Y^3 + Z^3$$

then

$$\begin{aligned} F_d(x, y) &= x^3 + y^3 + 1 \\ (F_d)^h &= X^3 + Y^3 + Z^3 \\ &= F \end{aligned}$$

On the other hand, if we start with a polynomial which is divisible by Z, such as

$$G(X, Y, Z) = Z\left(X^2 + Y^2\right)$$

then

$$\begin{aligned} G_d(x, y) &= x^2 + y^2 \\ (G_d)^h &= X^2 + Y^2 \\ &\neq G. \end{aligned}$$

The following proposition then tells us how the subsets of $\mathbb{P}^2$ and $\mathbb{A}^2$ defined by a homogeneous polynomial F and its dehomogenisation F_d are related.

Proposition 3.3. *Let $F(X, Y, Z)$ be a homogeneous polynomial and let $F_d(x, y)$ be its dehomogenisation. Let*

$$V(F) = \left\{[X, Y, Z] \in \mathbb{P}^2 \mid F(X, Y, Z) = 0\right\} \quad \textit{and}$$
$$V(F_d) = \left\{(x, y) \in \mathbb{A}^2 \mid F_d(x, y) = 0\right\}.$$

Then

$$V(F_d) = V(F) \cap \mathbb{A}^2.$$

Proof. Let p be a point of $\mathbb{A}^2$ with coordinates (x, y). Viewing it as a point of $\mathbb{P}^2$ via the map ι, the homogeneous coordinates of p have the form $p = [x, y, 1]$.

Then

$$\begin{aligned} p \in V(F) &\Leftrightarrow F(x, y, 1) = 0 \\ &\Leftrightarrow F_d(x, y) = 0 \\ &\Leftrightarrow p \in V(F_d). \end{aligned}$$

□

4 Elliptic Curves

Now we can give the formal definition of the objects we are interested in — **elliptic curves**. For the time being, we will take our field K to be the complex numbers $\mathbb{C}$.

Definition 4.1. *Let $F(X, Y, Z)$ be a homogeneous polynomial of the form*

$$F(X, Y, Z) = Y^2Z - G(X, Z) \tag{3}$$

where $G(X, Z)$ is homogeneous of degree 3 and its dehomogenesation $G_d(x)$ has 3 distinct roots.

Then the curve

$$C = V(F) = \left\{[X, Y, Z] \in \mathbb{P}^2 \mid F(X, Y, Z) = 0\right\}$$

is called an **elliptic curve**.

Remarks:

(i) The condition that $G_d(x)$ have 3 distinct roots implies that $G(X, Z)$ has the form

$$G(X, Z) = \alpha X^3 + \cdots$$

for some complex number $\alpha \neq 0$.

(ii) Why do we require 3 distinct roots? Consider the dehomogenisation

$$f(x, y) = F_d(x, y) = y^2 - G_d(x).$$

Then we find

$$\frac{\partial f}{\partial y} = 2y$$
$$\frac{\partial f}{\partial x} = -\frac{dG_d}{dx}.$$

So any point (x, y) such that $f(x, y) = \frac{\partial f}{\partial x}(x, y) = \frac{\partial f}{\partial y}(x, y) = 0$ must have $G_d(x, y) = \frac{dG_d}{dx}(x, y) = 0$: in other words, it must be a multiple root of G_d. Requiring that G_d have 3 distinct roots ensures there are no such points — hence the curve C has no **singular points**.

(iii) Why do we restrict our attention to cubic polynomials of the special form (3)? In fact, every irreducible homogeneous cubic polynomial $F(X, Y, Z)$ can be put into this form by an appropriate projective change of coordinates

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = M \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

where M is an invertible 3×3 matrix. So every irreducible cubic curve in $\mathbb{P}^2$ is "projectively equivalent" to one given by a polynomial of the form (3). Projective equivalence does not change any of the geometric properties of a curve; on the other hand, restricting to polynomials of the form (3) will simplify calculations when it comes to "adding points" on elliptic curves in Week 2.

Example

Let C be the curve defined by

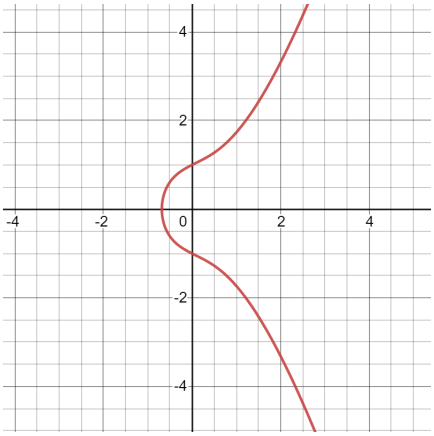
$$F(X, Y, Z) = Y^2Z - X^3 - XZ^2 - Z^3. \tag{4}$$

What does it look like?

- According to Proposition 3.3 the "affine part" of C, meaning the intersection $C \cap \mathbb{A}^2$, is exactly the set of zeroes of the dehomogenised equation $F_d(x, y) = 0$. This gives us the equation

$$y^2 = x^3 + x + 1.$$

We can draw a picture of the real number solutions of this equation:



- The part of C on the "line at infinity" $\{[X, Y, 0]\}$ can be found by putting $Z = 0$ in Equation (4): this gives $-X^3 = 0$, and hence $X = 0$. So we get a single point $[0, 1, 0]$.

  So the elliptic curve $C \subset \mathbb{P}^2$ is made up of the affine curve defined by $y^2 = x^3 + x + 1$, together with a single "point at infinity" $[0, 1, 0]$.

In fact, any elliptic curve intersects the "line at infinity" in a single point, just like in the previous example:

Proposition 4.2. *For any elliptic curve* C *given by an equation*

$$Y^2Z = G(X, Z)$$

the intersection $C \cap \{Z = 0\}$ *consists of the single point* $[0, 1, 0]$.

Proof. See Problem Sheet 1. □

Next time we will see how to **add together** points on C. We'll need the following result.

Proposition 4.3. *For any elliptic curve* C *as above, the intersection* $C \cap \{Z = 0\} = [0, 1, 0]$ *is a triple intersection point.*

Proof. The curve C is defined by an equation of the form

$$Y^2Z = \alpha X^3 + \beta X^2Z + \gamma XZ^2 + \delta Z^3 \quad (\alpha, \beta, \gamma, \delta \in \mathbb{C}, \alpha \neq 0).$$

To understand how the curve looks near the point $[0, 1, 0]$, we can dehomogenise with respect to Y: setting $Y = 1$ in the above equation we get

$$z = \alpha x^3 + \beta x^2 z + \gamma x z^2 + \delta z^3.$$

Now substitute the right-hand side into itself to get

$$\begin{aligned} z &= \alpha x^3 + \beta x^2(\alpha x^3 + \cdots) + \gamma x(\alpha x^3 + \cdots)^2 + \delta(\alpha x^3 + \cdots)^3 \\ &= \alpha x^3 + \text{ higher order terms} \end{aligned}$$

So $z(x)$ has a root of multiplicity 3 at $x = 0$, and therefore there is a triple intersection point here. □