CR 510
Crittosistemi Ellittici
Elliptic curves and applications to cryptography
Università degli Studi Roma 3
A.A. 2023/2024

Amos Turchet

December 2023 - Preliminary Version

Foreword

This notes are essentially based on the course material covered during the course on Elliptic Curves and applications to Criptography (CR 510) at the University of Roma Tre. These notes build on a previous version written by Prof. Laura Capuano during a previous instance of the same course. They also borrow insipiration and ideas from the excellent notes of Drew Sutherland at MIT [Sut23].

This course aims to give a friendly introduction to the theory of elliptic curves, and to discuss their applications in cryptography. Elliptic curves are one of the oldest and most central topic in Number Theory, going back in some form to Diophantus, and studied in the work of Fermat, Newton and Gauss among others (see [BM02] for some historical notes). In recent times they gain the spotlight for being a central tool in the proof by Wiles and Taylor of Fermat's Last Theorem [Wil95; TW95]. In a different direction, the advent of Elliptic Curve Crpyto-systems proposed by Koblitz [Kob87] and Miller [Mil86] independently, and their modern wide use for encryption algorithm in industry, has made the subject an incredible cross-roads between pure mathematical research and applications.

This notes are divided in four chapters:

1. **Algebraic curves and Elliptic Curves**: where we recall basic facts about the theory of algebraic curves and then focus on elliptic curves, their basic definitions and properties and the group law (see Chapter 1);
2. **Isogenies**: where we define morphisms of elliptic curves and isogenies, discuss their fundamental properties and consequences (see Chapter 2);
3. **Elliptic curves over finite fields**: where we discuss Hasse bounds, singular and supersingular elliptic curves, and properties that are need for the applications (see Chapter 3);
4. **Algorithmic aspects and cryptography**: where we focus on the algorithmic aspects of the theory of elliptic curves, and use this to describe the fundamental cryptosystems based on elliptic curves (see Chapter 4)

Contents

0 Introduction

0.1 Why Elliptic Curves

One of the central topics in Number Theory is the study of polynomial equations with integer (or rational) coefficients. One starts with a polynomial $f \in \mathbb{Z}[x_1, \dots, x_n]$ and want to study the set

$$V(f, R) := \{(z_1, \dots, z_n) \in \mathbb{Z}^n : f(z_1, \dots, z_n) = 0\}, \qquad \text{with } R = \mathbb{Z}, \mathbb{Q}.$$

Here by study we mean describe as much as possible $V(f, R)$, for example decide if the set is empty, finite or infinite, and in the case where is finite, list all of its element, or even give an algorithm that inputs f and returns $V(f, \mathbb{Z})$.

This very natural question has been the focus of mathematical works that goes back to the famous books of Diophantus of Alexandria (AD 200), so that these equations are commonly referred to as **Diophantine** equations. Famous examples include Fermat's Last Theorem, where $f = x_1^n + x_2^n - x_3^n$, Pell's equation where $f = x_1^2 - Dx_2^2 - 1$ among others. Untill the 1970 it was unknown whether an algorithm existed to determined, given f, whether $V(f, \mathbb{Z})$ was empty or not. This was the content of Hilbert's tenth problem (in his list of 1900 math problems), that asked:

> Given a Diophantine equation with any number of unknown quantities and with rational integral numerical coefficients: To devise a process according to which it can be determined in a finite number of operations whether the equation is solvable in rational integers.

In 1970 Matijasevič, building on works of Davis, Putnam and Robinson, proved in [Mat70] that such an algorithm does not exist in general. This however does not imply that it is possible to describe the set $V(f, \mathbb{Z})$ for certain types of polynomals. For example is not hard to see that this is indeed possible if the polynomial f is a polynomial in one variable $f(x)$. In this case the fundamental theorem of algebra tells us that the number of *complex* solutions is at most the degree of the polynomial f, so in particular $V(f, \mathbb{Q})$ is always a finite set. Moreover if $z = p/q$ is a solution with coprimes p and q, then there is only a finite number of possibilites for p and q, namely the divisors of the constant term and the leading coefficient of f. Thus in this case we can completely describe algorithmically $V(f, \mathbb{Q})$. For more details on this and much more we refer to [Zan09].

For the case of two variables the situation becomes more intricate; we will focus on the case $R = \mathbb{Q}$ for simplicity. So we are interested in study the rational solutions of the Diophantine equation $f(x, y) = 0$ with $f \in \mathbb{Z}[x, y]$. We start with the case in which

deg $f = 1$. In this case it is not hard to see that given *any* rational x there exists exactly one rational y such that $(x, y) \in V(f, \mathbb{Q})$. In particular there is a bijection of sets

$$V(f, \mathbb{Q}) \longrightarrow \mathbb{Q}.$$

We showed in particular that the set $V(f, \mathbb{Q})$ is always infinite if f is a linear polynomial in two variables.

The next case is the case of degree 2. In this case the set of real solutions to $f(x, y) = 0$ represents a conic $\mathcal{C}$ in the affine plane $\mathbb{R}^2$. There is effective argument, that goes back to Legendre, to determine whether the set $V(f, \mathbb{Q})$ is empty or not (see [Zan09, Supplements to Chapter 1]). So we assume that $V(f, \mathbb{Q}) \neq \emptyset$ and let $P = (a, b) \in V(f, \mathbb{Q})$. To any line $\ell : y = mx + q$ passing through P with rational slope we can associate the point $Q = \{\ell \cap \mathcal{C}\} \smallsetminus \{P\}$. The fact that the intersection is given by precisely two points can be read of the equation one obtains by substituting to y the expression $mx+q$ given by the line ℓ: such equation in one variable has one solution in $\mathbb{Q}$, namely a and therefore has exactly one other real solution. Moreover, since $m, q \in \mathbb{Q}$ and f has integral coefficients, it follows that also this other solution must belong to $\mathbb{Q}$. Therefore, $Q \in V(f, \mathbb{Q})$. Thus we obtain an infinite set of solutions by varying the line ℓ among all lines passing through P having rational slope. One can check that in fact, **every** point in $V(f, \mathbb{Q})$ is obtained in this way. We obtained again a bijection

$$V(f, \mathbb{Q}) \longrightarrow \mathbb{Q} \cup \{\infty\}.$$

Now the point ∞ corresponds to the case in which the line is parallel to the y-axis. We can summarize the discussion by saying that when f is a polynomial in two variables of degree ≤ 2, either $V(f, \mathbb{Q})$ is empty or it is infinite and parametrized by $\mathbb{P}^1_{\mathbb{Q}} = \mathbb{Q} \cup \{\infty\}$.

The next case, namely the one in which f has degree 3 is more complicated. Up to change of variables, in almost all cases, we can reduce to the case in which f has the following form:

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

for some rational numbers $A, B \in \mathbb{Q}$. This equation represents a cubic curve in $\mathbb{R}^2$ that, in the case when $x^3 + Ax + B$ has no multiple roots, is (the affine part of) an **elliptic curve**. More formally we will define an elliptic curve as a non-singular projective curve of genus 1 with a rational point (and explain in the sequel what all these terms mean). In this case, it is much harder to describe the set $V(f, \mathbb{Q})$. Already explicit examples show how to find any rational points is in fact hard.

Let us focus on an example of Diophantus (problem 24, Book IV): find two numbers whose sum is equal to a and their product is a cubic "minus its side": in other words find x, y such that

$$y(a - y) = x^3 - x.$$

The original solution of Diophantus sets $a = 6$ and substitutes $x = 3y - 1$: this cancels the linear and constant terms in y allowing to find the value $y = 26/27$ and thus

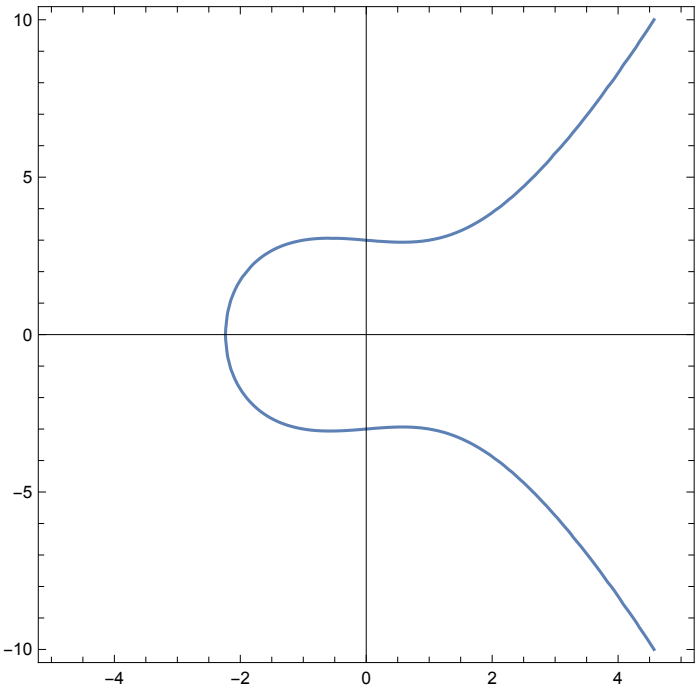

Figure 0.1: The real points of the curve $y^2 = x^3 - x + 9$

$x = 17/9$. Note that we can change the variables (e.g. subtracting 9 to both sides and update $y \to y - 3, x \to -x$) to obtain the previous form

$$E : y^2 = x^3 - x + 9.$$

A plot of the real points of E is given in Figure 0.1. Note that the polynomial $x^3 - x + 9$ has no double root, and we have the solution $P = (-1, 3)$ (that corresponds to the solution $(1, 0)$ before the change of variables) so that E represents an elliptic curve. It turns out that the point corresponding to $(17/9, 26/27)$ is $(-17/9, -55/27)$ that can be obtained geometrically as follows: one takes the tangent line to E at the point $(-1, 3)$ that intersects the cubic in a third point, namely $(-17/9, 55/27)$. If we reflect this point with respect to the x-axis we obtain exactly $(-17/9, -55/27)$. This apparently random process is in fact the geometric description of an operation defined on the points of the elliptc curve E. In other words, implicitly what Diophantus did was to compute the point $2P$.

The above discussion focus on a fundamental property of the set $E(\mathbb{Q}) := V(f, \mathbb{Q})$, that was proven by Poincaré but already discovered since Newton:

Theorem (Mordell)**.** *The set $E(\mathbb{Q})$ is a finitely generated abelian group.*

We will discuss the operation on the set of rational points and its properties in the next sections. The fact that $E(\mathbb{Q})$ is a group was originally observed much earlier than the statement that it is a finitely generated group, which was proven only in 1929 by

Mordell. A consequence of the Theorem is that the group $E(\mathbb{Q})$ is isomorphic to $\mathbb{Z}^r \oplus T$ where T is a finite group. The integer r is called the **rank** of the elliptic curve E and governs the size of the group (i.e. if $r = 0$ then the group is finite). At the present we lack an algorithm to compute the rank, or even to decide basic properties such us its boundedness. Even constructing examples with large rank is by no means an easy task: the present record is a curve constructed by Elkies of rank 28. A conjectural characterization of the rank is the core of one of the millennium problems, namely the *Birch–Swinnerton-Dyer Conjecture* that relates the rank to the order of vanishing of the so-called L-series associated to the elliptic curve.

On the other hand the group T (of torsion points) is quite well understood: a theorem of Mazur [Maz77] shows that T has cardinality at most 16 and it is the product of at most two cyclic groups in a finite list of possibilities.

To summarize, Elliptic Curves represent the smallest degree curves for which we have at the same time a rich structure and the presence of difficult and still unsolved problems. This makes them very interesting from the mathematical point of view but also suitable to cryptographic applications.

For sake of completeness we finish this section briefly discussing the situation for higher degree polynomials. If $f \in \mathbb{Q}[x, y]$ has degree ≥ 4 and no zeroes in common with its two derivatives (this technical condition will be explained later, it is related to the absence of "singular points", and it is satisfied by almost all the polynomials of fixed degree) then rational points become harder to find. Faltings showed in [Fal83], proving a conjecture of Mordell, that in fact there will be always at most finitely many.

Theorem (Faltings, 1983)**.** *For every irreducible polynomial $f \in \mathbb{Q}[x, y]$ of degree ≥ 4 with no zeroes in common with both its derivatives, the set $V(f, \mathbb{Q})$ is a finite set.*

`Finite Fields and Cryptography.` For applications one considers an elliptic curve E defined over a finite field $\mathbb{F}_p$, (or more generally $\mathbb{F}_q$ where $q = p^k$) where p is a prime number (which for simplicity we take different from 2 and 3). This is an equation of the form

$$E : y^2 = x^3 + Ax + B \qquad \text{with } A, B \in \mathbb{F}_p,$$

where we want to study the set

$$E(\mathbb{F}_p) = \{(x, y) \in \mathbb{F}_p^2 : y^2 = x^3 + Ax + B\}.$$

By construction, since $\mathbb{F}_p$ is finite, the set $E(\mathbb{F}_p)$ is a finite set. Since (roughly) half of the elements in $\mathbb{F}_p$ are squares we get that there is 50% chance that $x^3 + Ax + B$ for a random $x \in \mathbb{F}_p^*$. For each of these values we get to points corresponding to the two square roots. In particular we expect that the number of points we get is roughly $p + 1$ (where we count the point "at infinity"). This can be made more precise in the following theorem due to Hasse.

Theorem (Hasse)**.** *Given an elliptic curve $E : y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{F}_p$ one has*

$$|\#E(\mathbb{F}_p) - (p + 1)| \leq 2\sqrt{p}.$$

In the same way as in the characteristic zero case, one can show that the set $E(\mathbb{F}_p)$ is an abelian group. The interest in the group of points of an elliptic curve over a finite field comes from the fact that we can use it to implement cryptographic schemes based on the discrete logarithmic problem. In this instance the problem is the following:

Let $P \in E(\mathbb{F}_p)$ and let Q be a multiple of P. Find k such that $Q = [k]P$.

We will see that this problem is considered hard enough to be the core of many schemes to pass information in a secured way between to parties (similar to the RSA scheme). The advantage of using the group of points of an elliptic curve is that the size of the quantities involved, to make sure that the system is secure from computational attacks, is sensibly smaller than for those system that are based on the discrete logarithmic problem over a finite field. This will be discussed in length in the part of the course that focus on cryptography.

1 Algebraic curves and Elliptic Curves

In this chapter we discuss the basic theory of algebraic curves and the main properties of elliptic curves.

1.1 Algebraic curves

In this section we work over a field K which we assume separable. This in particular covers all the essential examples that will be important for us, namely $\mathbb{Q}, \mathbb{R}, \mathbb{C}, \mathbb{F}_p$ and $\mathbb{F}_q$. By the affine space over K, written $\mathbb{A}^n(K)$, we mean the set K^n with its K-vector space structure. In particular the affine plane $\mathbb{A}^2$ over K is just $K \times K$. The principal ideal domain $K[x, y]$ is the domain of all polynomials in two variables with coefficients in K. We say that a polynomial $f \in K[x, y]$ is **absolutely irreducible** if f is irreducible inside $\overline{K}[x, y]$ (where $\overline{K}$ is the algebraic closure of K).

1.1.1 Affine Curves

Definition 1.1.1. An **affine plane curve** $\mathcal{C}$ over K is the set of zeroes of a polynomial $f \in K[x, y]$. For every finite extension $L \supset K$ the set of L-points of $\mathcal{C}$, denoted $\mathcal{C}(L)$ is the set

$$\mathcal{C}(L) = \{(x, y) \in L : f(x, y) = 0\}.$$

We say that the curve $\mathcal{C}$ is **irreducible** (resp. **absolutely irreducible**) if f is irreducible (resp. absolutely irreducible).

We stress a crucial point: a curve is not "just" its set of points over a field (which might be the natural way one thinks about conics in high school) but rather a "gadget" that to every field L, that contains the field K where the coefficients live, associates the set $\mathcal{C}(L)$ of its L-rational points. These sets can have very different properties, e.g. if $\mathcal{C}$ is a conic then $\mathcal{C}(\mathbb{Q})$, $\mathcal{C}(\mathbb{R})$ and $\mathcal{C}(\mathbb{C})$ are very different! This point of view leads to the modern theory of schemes in algebraic geometry; while it is a very interesting mathematical setting, in order to keep the technicalities to a minimum, we will stick to the most elementary point of view in these notes.

Example 1.1.2. Consider $f(x, y) = y^2 - x^3 - x$. The associated affine plane curve $\mathcal{C}_1$ over $\mathbb{Q}$ is absolutely irreducible since f is irreducible in $\mathbb{C}[x, y]$.

The curve $\mathcal{C}_2$ defined by $g(x, y) = x^2 - 2y^2$ is irreducible over $\mathbb{Q}$ but is not absolutely irreducible since $g(x, y) = (x - \sqrt{2}y)(x + \sqrt{2}y)$ over $\mathbb{R}$ (or even over $\mathbb{Q}(\sqrt{2})$). Over $\mathbb{R}$ we can consider the two curves $\mathcal{C}_2'$ and $\mathcal{C}_2''$ defined by $(x - \sqrt{2}y)$ and $(x + \sqrt{2}y)$ respectively: they are absolutely irreducible and we call them the irreducible components of $\mathcal{C}$.

When K is an algebraically closed field, every non-constant polynomial defines a curve $\mathcal{C}$ that has infinitely many K-points. We want to characterize the points for which a tangent line to the curve is well-defined.

Definition 1.1.3. Given an affine plane curve $\mathcal{C}$ defined by a polynomial $f \in K[x, y]$ and a point $P \in \mathcal{C}(K)$, we say that P is a **smooth** (or non-singular) point of $\mathcal{C}$ if at least one among $\partial f/\partial x(P)$ and $\partial f/\partial y(P)$ is non zero. We say that the point P is **singular** if it is not smooth. An affine plane curve $\mathcal{C}$ defined over K is called **non-singular** if it has no singular point in $\overline{K}$.

Note that if the point $P = (x_P, y_P)$ is smooth there is well defined **tangent line** to $\mathcal{C}$ at P, namely the curve defined by the polynomial

$$\frac{\partial f}{\partial x}(P)(x - x_P) + \frac{\partial f}{\partial y}(y - y_P).$$

Example 1.1.4. Consider the curve E defined over $\mathbb{Q}$ defined by

$$E : f(x, y) = y^2 - x^3 - Ax - B.$$

If we set $g(x) = x^3 + Ax + B$ we see that

$$\frac{\partial f}{\partial x} = -\frac{\partial g}{\partial x} \qquad \frac{\partial f}{\partial y} = 2y.$$

Therefore a singular point $P = (x_0, y_0) \in E(\mathbb{Q})$ satisfies $y_0 = 0$ and $g'(x_0) = 0$. Moreover, since $P \in E(\mathbb{Q})$ and $y_0 = 0$ we also know that $g(x_0) = 0$. This implies that a singular point of E is necessarily a double root of the polynomial $g(x)$ (being a common zero of g and its derivative). Since g is a cubic polynomial, it has multiple roots if and only if its discriminant $4A^3 + 27B^2 = 0$([1]). In particular if $4A^3 + 27B^2 \neq 0$ then the curve E has no singular points (this argument works for any field K such that $\operatorname{char} K \neq 2$).

It is not hard to show that if $f = f_1 \cdot f_2$ is a product of two polynomials, then every point in the intersection of f_1 and f_2 is a singular point for $\mathcal{C}$ defined by f.

One can describe better singular points by "measuring" the singularity as follows.

Definition 1.1.5. Let $P = (a, b) \in \mathcal{C}(K)$ for an affine plane curve $\mathcal{C}$ defined by an irreducible polynomial $f \in K[x, y]$. Since $f(P) = 0$ we can use the Taylor expansion of f at P and write

$$f(x, y) = f_1(x - a, y - b) + f_2(x - a, y - b) + \cdots + f_n(x - a, y - b),$$

where $f_i(x - a, y - b)$ is a homogeneous polynomial of degree i in $x - a$ and $y - b$. Then P is a non singular point if and only if $f_1(P) \neq 0$ (and $f_1 = 0$ is the equation of the

[1] Recall that for a *monic* cubic polynomial with roots x_1, x_2, x_3 over $\mathbb{C}$, its discriminant is defined as $\Delta = (x_1 - x_2)^2(x_1 - x_3)^2(x_2 - x_3)^2$.

tangent line at $\mathcal{C}$ in P). On the other hand P is singular if and only if $f_1(P) = 0$. In this case we call the **multiplicity** of P the least integer $m \geq 2$ such that $f_m(P) \neq 0$.

Note that in this case

$$f(x, y) = f_m(x - a, y - b) + \text{ higher order terms.}$$

If $m = 2$ the point P is called a **double point**.

Let for simplicity $P = (0, 0)$. Then in $\overline{K}[x, y]$ we can factor

$$f_m(x, y) = \prod_i L_i^{r_i}(x, y)$$

where L_i are homogeneous linear polynomials. We call the L_i's the **tangent lines** at $\mathcal{C}$ in P, and r_i is the **multiplicity** of L_i.

A point P is called an **ordinary singularity** if $m \geq 2$ and $r_i = 1$ for every i. An ordinary double point is called a **node**.

Remark 1.1.6. Note that any **homogeneous** polynomial in two variables over K splits as product of linear polynomials in $\overline{K}$. In fact, given $f_m(x, y)$ homogeneous of degree m, there exists a polynomial $g \in K[x/y]$ such that $y^m g(x/y) = f(x, y)$. Since g splits as a product of linear terms in $\overline{K}$ we get the corresponding factorization for f.

Example 1.1.7. Consider the curve given by $y^2 = x^3 - ax^2$ over $\mathbb{Q}$; then $P = (0, 0)$ is a singular point. We can write

$$f(x, y) = y^2 - x^3 - ax^2 = (y^2 - ax^2) + (-x^3) = f_2(x, y) + f_3(x, y),$$

which shows that P is a double point since $m = 2$. If we look at f_2 we see that we have two cases:

$$f_2(x, y) = \begin{cases} y^2 & \text{if } a = 0, \\ (y - \sqrt{a}\, x)(y + \sqrt{a}\, x) & \text{if } a \neq 0. \end{cases}$$

which shows that P is a node (ordinary double point) if $a \neq 0$, since it has distinct tangent lines, and has a unique tangent line of multiplicity 2 if $a = 0$. In the latter case we say that P is a **cusp**.

1.1.2 Projective Curves

The projective plane over a field K is defined as

$$\mathbb{P}^2(K) = \{(x, y, z) \in K^3 : (x, y, z) \neq (0, 0, 0)\}/_\sim,$$

where $(x, y, z) \sim (x', y', z')$ if and only if there exists $\lambda \in K^\times$ such that $(x', y', z') = (\lambda x, \lambda y, \lambda z)$. The equivalence class is usually denoted by $(x : y : z)$. Note that given $P \in \mathbb{P}^2(K)$ the set of triples in the equivalence class of P lie in a single line $L(P) \in \mathbb{A}^3(K)$ passing through the origin. The map $P \mapsto L(P)$ defines a bijection between the lines

passing through the origin in K^3 and $\mathbb{P}^2(K)$. The same definition can be used to define $\mathbb{P}^n(K)$.

There are several subsets of $\mathbb{P}^2(K)$ that are in bijection with $\mathbb{A}^2(K)$. For convenience we will make a choice of a distinguish subset that we will call the **affine plane** inside $\mathbb{P}^2$. We let $U_2 = \{(x : y : z) : z \neq 0\}$ and its complement $L_\infty = \{(x : y : z) : z = 0\}$. Then, we have bijections

$$\mathbb{A}^2 \ni (x, y) \mapsto (x : y : 1) \in U_2,$$
$$\mathbb{P}^1 \ni (x : y) \mapsto (x : y : 0) \in L_\infty.$$

Moreover we can write $\mathbb{P}^2(K)$ as the disjoint union $U_2 \sqcup L_\infty$ of the affine plane and the **line at infinity** L_∞.

Definition 1.1.8. A **projective plane curve** $\mathcal{C}$ over K is defined by a *homogeneous* polynomial $f \in K[x, y, z]$. For every $L \supset K$ its L-**rational points** are the set

$$\mathcal{C}(L) = \{(x : y : z) \in \mathbb{P}^2(L) : f(x, y, z) = 0\}.$$

The **degree** of the curve $\mathcal{C}$ is the degree of the polynomial f defining $\mathcal{C}$.

Note that, since f is homogeneous we have that $f(cx, cy, cz) = c^{\deg f} f(x, y, z)$. This implies that the even if the *value* of $f(x : y : z)$ is not well defined, it does make sense to say whether f is zero at $(x : y : z)$.

Example 1.1.9. Let $f(x, y, z) := y^2z - x^3 - Axz^2 - Bz^3$ and let E be the corresponding curve over $\mathbb{Q}$. Is a projective plane curve of degree 3. In the affine plane U_2 we recover the curve $y^2 - x^3 - Ax - B$, i.e. $E \cap U_2$ is the affine plane curve given by the polynomial $f(x, y, 1)$. On the other hand $L_\infty \cap E = \{(0 : 1 : 0)\}$. So we can think of E has the affine plane cubic "plus" a unique point at infinity.

The previous example suggests that every projective curve is in fact build from affine patches. If we consider the two subsets

$$U_0 = \{(x : y : z) : x \neq 0\} \qquad U_1 = \{(x : y : z) : y \neq 0\}$$

then, they can both be naturally identified with $\mathbb{A}^2$ via the maps

$$(y, z) \mapsto (1 : y : z) \qquad (x, z) \mapsto (x : 1 : z).$$

Given that at least one among x, y, z is different from zero we have $\mathbb{P}^2(K) = U_0 \cup U_1 \cup U_2$. This implies that a projective plane curve $\mathcal{C}$ can be written as the union of three affine curves

$$\mathcal{C} = (\mathcal{C} \cap U_0) \cup (\mathcal{C} \cap U_1) \cup (\mathcal{C} \cap U_2).$$

Under the above identifications, if $\mathcal{C}$ is defined by the polynomial $f(x, y, z)$ then the three curves are defined by the polynomials $f(1, y, z)$, $f(x, 1, z)$ and $f(x, y, 1)$ respectively.

Using these patches we can define the notions of tangent line, multiplicity etc, by noting that each point $P \in \mathcal{C}(K)$ will lie on at least one of the affine curves whose union is $\mathcal{C}$.

Example 1.1.10. Consider the curve

$$\mathcal{C} : y^2z = x^3 + Axz^2 + Bz^3.$$

The associated three affine curves are

$$\mathcal{C}_0 : y^2z = 1 + Az^2 + Bz^3 \qquad \mathcal{C}_1 : z = x^3 + Axz^2 + Bz^3 \qquad \mathcal{C}_2 : y^2 = x^3 + Ax + B.$$

Given any of the three curves we can recover $\mathcal{C}$ by "homogenizing" the equation: given any polynomial in two variables, e.g. $g(x, y) = y^2 - x^3 - Ax - B$ we can associate a unique homogeneous polynomial $\tilde{g}(x, y, z)$ by multiplying each monomial by a suitable power of z in order to obtain a homogeneous polynomial of the same degree. In our example $\tilde{g}(x, y, z) = y^2z - x^3 + Axz^2 - Bz^3$.

1.1.3 Intersection Numbers and Bezout Theorem

The goal of this section is to state a theorem that describes quantitatively the intersection of two plane projective curves. In order to do that we have to first describe how to define the intersection number of two curves at a common point. For simplicity we will restrict to the case in which the intersection point is the origin.

Let $\mathcal{F}(K)$ be the set of polynomials $f, g \in K[X, Y]$ having no common factor $h \in K[X, Y]$ such that $h(0, 0) = 0$. Let $\mathcal{C}_f$ and $\mathcal{C}_g$ be the corresponding plane affine curves associated to f and g respectively. The next proposition shows that there is only one reasonable way to define the intersection number $I(f, g)$ of $\mathcal{C}_f$ and $\mathcal{C}_g$ at the origin.

Proposition 1.1.11. *There is a unique map $I : \mathcal{F}(K) \to \mathbb{N}$ such that*
(a) $I(x, y) = 1$;
(b) $I(f, g) = I(g, f)$ for all $(f, g) \in \mathcal{F}(K)$;
(c) $I(f, gh) = I(f, g) + I(f, h)$ for all ;
(d) $I(f, g + hf) = I(f, g)$ for all ;
(e) $I(f, g) = 0$ if $g(0, 0) \neq 0$.

Since if $L \supset K$ we have $\mathcal{F}(L) \subset \mathcal{F}(K)$ (this follows from the theory of resultants), hence we can replace K with its algebraic closure when defining $I(f, g)$. We let

$$K[x, y]_{(0,0)} = \left\{ \frac{h_1}{h_2} : h_1, h_2 \in K[x, y] : h_2(0, 0) \neq 0 \right\},$$

the localization at the maximal ideal $(0, 0)$. Then one can show that the quotient ring $K[x, y]_{(0,0)}/(f, g)$ by the ideal generated by f and g is a finite dimensional K-vector space. Then we set

$$I(f, g) = \dim_K K[x, y]_{(0,0)}/(f, g).$$

Let us see some examples.

Example 1.1.12. Let $f = y^2 - x^3 - x^2$ and $g = x$. Then

$$I(f, g) = I(y^2 - x(x^2 - x), x) = I(y^2, x) = 2.$$

Note that the intersection number is greater than 1 even if the y-axis is not tangent to the curve defined by f at the origin. This can be explained since such curve is singular at the origin.

Definition 1.1.13. Given two curves $\mathcal{C}_f$ and $\mathcal{C}_g$ in $\mathbb{A}^2(K)$ defined by two polynomials f, g we say that a point $P \in \mathcal{C}_f(K) \cap \mathcal{C}_g(K)$ is **isolated** if there is no common irreducible component passing through P. In this case there is no common factor h of f and g such that $h(P) = 0$. We can then define the intersection number of $\mathcal{C}_f$ and $\mathcal{C}_g$ at $P = (a, b)$ to be

$$I(P, \mathcal{C}_f \cap \mathcal{C}_g) := I(f(x + a, y + b), g(x + a, y + b)).$$

Example 1.1.14. Let $\mathcal{C}$ be the curve defined by the polynomial $y^2 = x^3$ and let $L : y = 0$ be the tangent line at the origin $P = (0, 0)$. Then

$$I(P, \mathcal{C} \cap L) = I(y^2 - x^3, y) = I(x^3, y) = 3$$

We can use the intersection number to state a fundamental result that describes how to projective curve intersect.

Theorem 1.1.15 (Bezout)**.** *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ two projective curves over a field K of degree d and e respectively. Then*

$$\sum_{P \in \mathcal{C}_1(\overline{K}) \cap \mathcal{C}_2(\overline{K})} I(P, \mathcal{C}_1 \cap \mathcal{C}_2) = d\, e.$$

In other words $\mathcal{C}_1$ and $\mathcal{C}_2$ intersect over $\overline{K}$ in exactly de points, counting multiplicities.

As an application we can use Bezout Theorem to compute the intersection of the projective curve $\mathcal{C}$ given by $y^2z - x^3 - Axz^2 - Bz^3$ with the line at infinity $L_\infty : z = 0$. We have already seen that the intersection is a single point $P = (0 : 1 : 0)$. Since the curve has degree 3 and the line has degree 1, Bezout's Theorem implies that the intersection has multiplicity 3. This can be computed explicitly as follows: we can work in the chart $y \neq 0$ in which the point P becomes the origin. Then,

$$I(P, L_\infty \cap \mathcal{C}) = I(z, z - x^3 - Axz^2 - Bz^3) = I(z, x^3) = 3.$$

A non singular point P con a curve $\mathcal{C}$ is called an **inflection point** if the intersection multiplicity with the tangent line at P is ≥ 3.

1.2 The group structure on an Elliptic Curve

In this section we will use Bezout Theorem to show that the group of K-points in an elliptic curve form a group. To show this we will start with a non singular projective plane curve of degree 3 over a field K (and one of the most interesting cases for us will be when $K = \mathbb{Q}$) such that $\mathcal{C}(K) \neq \emptyset$. Fix a point $\mathcal{O} \in \mathcal{C}(K)$ and pick two points $P, Q \in \mathcal{O}$; the line ℓ_{PQ} through P and Q is a projective curve of degree 1. Therefore by Bezout Theorem (Theorem 1.1.15) the intersection $\mathcal{C} \cap \ell$ consists of three points counted with multiplicities. In practice this means that the set-theoretic intersection between the two curves can consist of either one, two or three points.

Let us consider first the case in which $P \neq Q$ and $\ell_{PQ} \cap \mathcal{C}$ contains a third point R different from P and Q. Note that since $P, Q \in \mathcal{C}(K)$ the third point of intersection has also coordinates in the field K. To see this consider an affine patch that contain the three points (for example $z \neq 0$ if none of the points is a point at infinity for $\mathcal{C}$): then to find the x-coordinate of R we can substitute in the polynomial defining $\mathcal{C}$ an expression of the form $y = mx + q$, where the latter is the equation of the line ℓ_{PQ}. We obtain a cubic polynomial in the variable x with coefficients in K and with two roots that lie in K (namely the x coordinates of P and Q[2]). But then the polynomial has a third root which also lie in K, showing that both the x and y coordinate of the third point of intersection R are in K, i.e. $R \in \mathcal{C}(K)$. The same argument applies in the case in which $P = Q$, so ℓ_{PP} is the *tangent line* at $\mathcal{C}$ in P or $R \in \{P, Q\}$.

Then, we can define the following operation:

$$\begin{aligned} \oplus : E(K) \times E(K) &\to E(K) \\ (P, Q) &\mapsto P \oplus Q \end{aligned}$$

where $P \oplus Q$ is the third point of intersection of $\ell_{\mathcal{O}R} \cap \mathcal{C}$ where R is as above. If we take $\mathcal{C}$ defined by the polynomial $f(x, y, z) = y^2z - x^3 - Axz^2 - Bz^3$, $\mathcal{O} = (0 : 1 : 0)$, and we draw the affine curve $\mathcal{C}$ in the patch $z \neq 0$ then we can visualize the operation as follows: given P and Q, distinct and neither of them lying in the tangent line to the other, the point $P \oplus Q$ is the reflection with respect to $y = 0$ of the point R, the third point of intersection of ℓ_{PQ}. See Figure 1.1 for a picture of this situation.

An analogue description of the $\oplus$ operation is that if $\{P, Q, R\}$ are the three points of intersection of $\ell_{PQ} \cap \mathcal{C}$ then $(P \oplus Q) \oplus R$ is the identity element of the group. To ease the notation we will denote by $P * Q$ the third point of intersection $\ell_{PQ} \cap \mathcal{C}$.

Theorem 1.2.1. *Given a non singular projective curve $\mathcal{C}$ of degree three defined over a field K, and a point $\mathcal{O} \in \mathcal{C}(K)$, the operation $\oplus$ defines a group $(\mathcal{C}(K), \oplus)$ such that*

1. *the group is commutative, i.e. $P \oplus Q = Q \oplus P$ for every $P, Q \in \mathcal{C}(K)$;*
2. *the point $\mathcal{O}$ is the identity element, i.e. $P \oplus \mathcal{O} = P$ for every $P \in \mathcal{C}(K)$;*
3. *every point has its inverse, i.e. for every $P \in \mathcal{C}(K)$ there exists a point $-P \in \mathcal{C}(K)$ such that $P \oplus (-P) = \mathcal{O}$;*

[2] unless we are in the case in which the two coordinates coincide. But in this case the line would be of the form $x =$ const and we can run the same argument switching the roles of x and y.

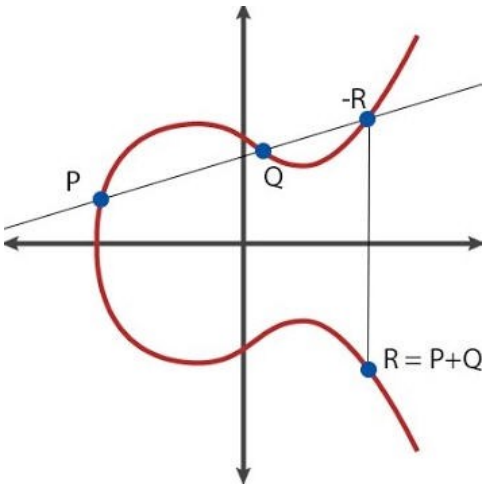

Figure 1.1: The addition law on an elliptic curve

4. *the operation* $\oplus$ *is associative, i.e. for every* $P, Q, R \in \mathcal{C}(K)$ $(P \oplus Q) \oplus R = P \oplus (Q \oplus R)$.

We will see later that, up to an invertible linear change of variable, we can always assume that $\mathcal{O}$ is the point $(0 : 1 : 0)$ which is an inflection point for $\mathcal{C}$.

Proof. From the definition of $\oplus$ we can see that it is commutative, since the line ℓ_{PQ} and ℓ_{QP} coincide. This proves (1).

Let us compute $P \oplus \mathcal{O}$: by definition we define $P * Q$ to be the third point of intersection $\ell_{P\mathcal{O}} \cap \mathcal{C}$, and then $P \oplus \mathcal{O}$ is $R * \mathcal{O}$, i.e. the third point of intersection $\ell_{(P*Q)\mathcal{O}} \cap \mathcal{C}$. But by construction $P, P * Q$ and $\mathcal{O}$ are collinear, i.e. $\ell_{P\mathcal{O}} = \ell_{(P*Q)\mathcal{O}}$. Therefore $P \oplus \mathcal{O} = P$ as wanted (note that the argument did not require $P * Q$ to be distinct from P or $\mathcal{O}$). This proves (2).

To prove that every point P has an inverse we will show that $-P$ is $P * (\mathcal{O} * \mathcal{O})$, i.e. is the third point of intersection $\ell_{P(\mathcal{O}*\mathcal{O})} \cap \mathcal{C}$, where $\mathcal{O} * \mathcal{O}$ is the third point of intersection of the tangent line at $\mathcal{C}$ in $\mathcal{O}$ (we do not exclude that $\mathcal{O} * \mathcal{O} = \mathcal{O}$). Now, to compute $P \oplus (-P)$ we need the line $\ell_{P(-P)}$ that by construction coincide with the line $\ell_{P(\mathcal{O}*\mathcal{O})}$, since the points $P, \mathcal{O} * \mathcal{O}, (-P)$ are collinear. But then $P \oplus (-P)$ will be the third point of intersection $\ell_{(\mathcal{O}*\mathcal{O})\mathcal{O}} \cap \mathcal{C}$, and, by construction, the line $\ell_{(\mathcal{O}*\mathcal{O})\mathcal{O}}$ is the tangent line at $\mathcal{C}$ in $\mathcal{O}$ so that $P \oplus (-P) = \mathcal{O}$ as wanted. This proves (3).

Property (4) is the hardest to show. There are three main ways to prove associativity: the first one is by hand by applying repetitively the definition of the operation. The second one uses the Riemann-Roch Theorem, and gets the result as an immediate corollary of a different characterization of the group operation. Finally, the third way, which we will present here, uses a more geometric approach.

The goal is to show that $(P \oplus Q) \oplus R = P \oplus (Q \oplus R)$. It is then sufficient to show that $(P \oplus Q) * R = P * (Q \oplus R)$ i.e. the third point of intersection of $\ell_{(P\oplus Q)R} \cap \mathcal{C}$ and $\ell_{P(Q\oplus R)} \cap \mathcal{C}$ coincide. Let us define the following six lines:

$$\begin{aligned} \ell_1 &= \text{ line through } P, Q & m_1 &= \text{ line through } P, Q \oplus R \\ \ell_2 &= \text{ line through } Q \oplus R, \mathcal{O} & m_2 &= \text{ line through } Q, Q * R \\ \ell_3 &= \text{ line through } P \oplus Q, R & m_3 &= \text{ line through } P * Q, \mathcal{O} \end{aligned}$$

Then $(P \oplus Q) * R = \ell_3 \cap \mathcal{C}$ and $P * (Q \oplus R) = m_1 \cap \mathcal{C}$. Therefore it suffices to show that $\ell_3 \cap m_1 \in \mathcal{C}$.

Let $g(x, y, z)$ the cubic polynomial defined by the product of the lines $\ell_1\ell_2\ell_3$, and let similarly $h(x, y, z) = m_1m_2m_3$. By construction these two cubics have eight points in common, namely $\mathcal{O}, P, Q, R, P * Q, Q * R, P \oplus Q, Q \oplus R$. To simplify the argument let as assume that these eight points are in general position, i.e. if we denote by $(x_i : y_i : z_i)$ the coordinates of the eight points we assume that the vectors

$$(x_i^3, x_i^2y_i, x_i^2z_i, \ldots, z_i^3) \qquad i = 1, \ldots, 8,$$

where the components are all possible monomials of degree three in x_i, y_i, z_i are linearly independent.

Then the subspace of $K[x, y, z]_3$, the vector space of homogeneous polynomials of cubics in three variables, formed by all the cubics that pass through the eight point has dimension 2. In particular every such cubic is of the form $\lambda F + \mu G$ for some cubics F, G. By Bezout Theorem (Theorem 1.1.15) F and G have a ninth zero in common, and therefore every curve of the form $\lambda F + \mu G$ passes through this ninth point.

Finally we observe that $\mathcal{C}$ is one of these cubics, since it passes through all the eight point and therefore has to pass also through the ninth, which is $\ell_3 \cap m_1$ as wanted.

To handle the general case one uses the following Lemma:

Lemma 1.2.2. *Let C, C', C'' three cubic curves in $\mathbb{P}^2$ with C irreducible. Suppose that $C \cap C'$ consists of non singular points of C and eight of them, counted with multiplicity, are also points of intersections of $C \cap C''$. Then the ninth point of $C \cap C'$ coincide with the ninth point of $C \cap C''$.*

□

1.2.1 Divisors and the Riemann-Roch Theorem

We will now see a different approach to the group structure on the set of K-point on an elliptic curve. With this aim we introduce some definitions. For the moment we will work with an algebraically closed field K.

Definition 1.2.3. Let $\mathcal{C}$ be an affine plane curve over K, defined by an irreducible polynomial $f(x, y)$. Then any polynomial $g(x, y)$ defines a function

$$C(K) \to K \qquad (a, b) \mapsto g(a, b).$$

that we call a **regular function** on $\mathcal{C}$.

Any multiple of f defines the zero function on $\mathcal{C}(K)$ and Hilbert's Nullstellensatz implies that the converse holds. In particular any two polynomials $g_1, g_2 \in K[x,y]$ defines the same regular function if $g_1 - g_2 \in (f)$. Then we have a bijection

$$\frac{K[x,y]}{(f(x,y))} \to \{ \text{ regular function on } \mathcal{C}\}.$$

Definition 1.2.4. The **coordinate ring** of the curve $\mathcal{C}$ is

$$K[\mathcal{C}] := \frac{K[x,y]}{(f(x,y))}$$

Its elements are the regular functions on $\mathcal{C}$ and they are polynomials in the **coordinate functions** $P \mapsto \overline{x}(P)$ and $P \mapsto \overline{y}(P)$, where $\overline{x}, \overline{y}$ are the image of x, y in the coordinate ring.

Note that a regular function on $\mathcal{C}$ has only finitely many zeros on $\mathcal{C}$: this follows from the fact that given a polynomial $g(x,y)$ the curve $g(x,y) = 0$ intersects $\mathcal{C}$ only in finitely many points, unless $f \mid g$.

Since (f) is a prime ideal in $K[x,y]$ the coordinate ring $K[\mathcal{C}]$ is an integral domain.

Definition 1.2.5. We denote by $K(\mathcal{C})$ the field of fraction of the coordinate ring $K[\mathcal{C}]$. An element $\varphi = g/h \in K(\mathcal{C})$ defines a function

$$\mathcal{C}(K) \smallsetminus \{ \text{ zeros of } h\} \to K \qquad (a,b) \mapsto \frac{g(a,b)}{h(a,b)}.$$

We call φ a **rational function** on $\mathcal{C}$, and we say that φ is **regular** on the complement of the zeros of h.

Example 1.2.6. If $\mathcal{C} : y = 0$ then

$$K[\mathcal{C}] = \frac{K[x,y]}{(y)} \cong K[x] \qquad \text{and} \qquad K(\mathcal{C}) = K(x).$$

If $\mathcal{C} : y^2 = x^3 + Ax + B$ then

$$K[\mathcal{C}] = \frac{K[x,y]}{(y^2 = x^3 + Ax + B)} = K[\overline{x}, \overline{y}].$$

In this case $\overline{x}, \overline{y}$ satisfy the relation $\overline{y}^2 = \overline{x}^3 + A\overline{x} + B$ and any regular function on $\mathcal{C}$ is a polynomial in $\overline{x}, \overline{y}$.

We can mimic the same definition for projective curves, keeping in mind that in general, for a homogeneous polynomial $G(x,y,z)$, its value at a projective point is not well defined. On the other hand, if we have *two* homogeneous polynomials G, H of the same degree the function

$$(a : b : c) \mapsto \frac{G(a,b,c)}{H(a,b,c)}$$

is a well defined function for every $(a : b : c)$ such that $H(a,b,c) \neq 0$. We then give the following definition:

Definition 1.2.7. Let $\mathcal{C}$ be a plane projective curve defined over K defined by an irreducible homogeneous polynomial $F(x,y,z)$. For every couple of homogeneous polynomials of the same degree $G(x,y,z), H(x,y,z)$ such that H is not a multiple of F the function

$$\mathcal{C}(K) \to K \qquad (a:b:c) \mapsto \frac{G(a,b,c)}{H(a,b,c)}$$

is a well defined function on the complement in $\mathcal{C}(K)$ of the finite set of zeros of H. We denote by $K(\overline{x},\overline{y},\overline{z})$ the fraction field of $K[\overline{x},\overline{y},\overline{z}] = K[x,y,z]/(f)$. We define the **function field** of the curve $\mathcal{C}$ to be

$$K(\mathcal{C}) := \left\{ \frac{g}{h} \in K(\overline{x},\overline{y},\overline{z}) : g,h \in K[\overline{x},\overline{y},\overline{z}]_d \text{ for some } d \right\}.$$

Here $K[\overline{x},\overline{y},\overline{z}]_d$ is the set of elements having a representative in $K[x,y,z]$ homogeneous of degree d.

Then $K(\mathcal{C})$ is a sub field of $K(\overline{x},\overline{y},\overline{z})$ and its elements are called **rational functions** on $\mathcal{C}$. Given a point $P \in \mathcal{C}(K)$ we say that a rational function is **regular** at P if there exists a representative G/H such that $H(P) \neq 0$.

Example 1.2.8. Let $\mathcal{C}$ be the projective line $\mathbb{P}^1$, say given by $y = 0$. Then the rational functions are of the form

$$(a:0:b) = (a:b) \mapsto \frac{G(a,b)}{H(a,b)}$$

where $G(x,z), H(x,z)$ are homogeneous polynomials of the same degree.

Example 1.2.9. Let $\mathcal{C}$ be the curve defined by $zy^2 - x^3 - z^2x = 0$ and let $P = (0:0:1) \in \mathcal{C}(K)$. Consider the rational function $\alpha(x:y:z)$ associate to the expression $3xz/y^2 \in K(\mathcal{C})$. Since $y(P) = 0$ we cannot directly evaluate $\alpha(P)$: on the other hand we have that in $K(\mathcal{C})$ the relation $\overline{z}\overline{y}^2 = \overline{x}^3 + \overline{z}^2\overline{x}$ we can rewrite

$$\frac{3xz}{y^2} = \frac{3xz^2}{y^2z} = \frac{3xz^2}{x^3+z^2x} = \frac{3z^2}{x^2+z^2}.$$

Hence there is a representative of α for which $\alpha(P) = 3$. In particular α is regular at the point P.

Note that the elements of $K(\mathcal{C})$ are equivalence classes using **two** different relations: the first one is the one coming from the quotient $K[x,y,z]/(F)$ (so numerator and denominator are defined up to multiples of F). The second one comes from the definition of fractions: $\alpha \in K(\mathcal{C})$ can be written in different ways (as we did in the example) since $p/q = p'/q'$ if $pq' - qp'$ is zero in $K[x,y,z]/(f)$.

We note that the function field of an affine curve $\mathcal{C}$ in fact coincide with the function field of its projective closure $\overline{\mathcal{C}}$.

For what follows we will assume that K is *algebraically closed*. We want to understand the rational functions on the curve $\mathcal{C}$ using their zeros and poles.

Definition 1.2.10. The **group of divisors** $\mathrm{Div}(\mathcal{C})$ on $\mathcal{C}$ is the free abelian group on the set $\mathcal{C}(K)$. In particular, an element of $\mathrm{Div}(\mathcal{C})$ is a formal sum

$$D = \sum n_p[P] \qquad \text{where } n_p \in \mathbb{Z} \quad P \in \mathcal{C}(K).$$

We define the **degree** of D as above as $\sum n_P$. The **support** of a divisor $D = \sum n_p P$ is the set $\{P \in \mathcal{C}(K) : n_p \neq 0\}$.

To every rational function $\alpha \in K(\mathcal{C})$ we can associate a divisor as follows: let $\alpha = G/H$ where G, H are of the same degree m, and let us assume that $\alpha \neq 0$: this implies that F does not divide G (and by definition it does not divide H). We can then apply Bezout theorem and get

$$(\deg F) \cdot m = \sum_{P:F(P)=G(P)=0} I(P, C \cap \{G = 0\})$$
$$(\deg F) \cdot m = \sum_{P:F(P)=H(P)=0} I(P, C \cap \{H = 0\}).$$

Definition 1.2.11. In the notation above we define the **divisor of** α to be

$$\mathrm{div}(\alpha) = \sum_{P:F(P)=G(P)=0} I(P, C \cap \{G = 0\})[P] - \sum_{P:F(P)=H(P)=0} I(P, C \cap \{H = 0\})[P]$$

The $[P]$ in the support of $\mathrm{div}(\alpha)$ which appear with positive coefficient are called **zeros** of α and those occurring with negative coefficients are its **poles**. By definition the degree of $\mathrm{div}(\alpha)$ is zero.

The divisor associated to a rational function $\alpha = G/H$ is well defined: in particular for every choice of presentation of α its divisor will be the same. Note that a priori changing the numerator or the denominator might change the set $\{P : F(P) = G(P) = 0\}$ or the set $\{P : F(P) = H(P) = 0\}$: the fact that the divisor is well-defined means that in final sum, the two divisor will be equal.

Example 1.2.12. Let us compute the divisor associated to $\alpha = 3xz/y^2$ using the two presentation $3xz/y^2 = 3z^2/x^2 + z^2$ in the curve $\mathcal{C}$ defined by $F = y^2z - x^3 - xz^2 = 0$.

$\boxed{\alpha = 3xz/y^2}$ Let us start by computing the points where the numerator $3xz$ or the denominator y^2 vanish on the curve $\mathcal{C}$. If $z = 0$ then we have only the point $P_1 = [0 : 1 : 0]$. On the other hand if $z \neq 0$ then the equation becomes $y^2 - x^3 - x = 0$ where the only point with $x = 0$ is the point $P_2 = [0 : 0 : 1]$. Finally, if $y = 0$ then $x(x^2 + 1) = 0$ which gives either the point P_2 or the two points $Q_1 = [i : 0 : 1]$ and $Q_2 = [-i : 0 : 1]$. This shows that in this presentation the divisor of α is given by

$$\begin{aligned}\mathrm{div}(\alpha) =& I(P_1, 3xz \cap F)P_1 + I(P_2, 3xz \cap F)P_2 + \\ & - I(P_2, y^2 \cap F)P_2 - I(Q_1, y^2 \cap F) - I(Q_2, y^2 \cap F).\end{aligned}$$

In order to compute the intersection numbers we start with the point P_1 that lives in the chart $y \neq 0$, where using the properties of the intersection numbers we get

$$\begin{aligned} I(P_1, 3xz \cap F) &= I(P_1, x \cap z - x^3 - xz^2) + I(P_1, z \cap z - x^3 - xz^2) \\ &= I(P_1, x \cap z) + I(P_1, z \cap x^3) = 4. \end{aligned}$$

Similarly for the point P_2 in the chart $z \neq 0$ we get

$$I(P_2, 3xz \cap F) = I(P_2, x \cap y^2 - x^3 - x) = I(P_2, x \cap y^2) = 2.$$

For the denominator we get

$$\begin{aligned} I(P_2, y^2 \cap F) &= 2I(P_2, y \cap y - x^3 - x) = 2I(P_2, y \cap -x(x^2+1)) \\ &= 2I(P_2, y \cap -x) + 2I(P_2, y \cap x^2 + 1) = 2. \end{aligned}$$

For the point Q_1 we also use the chart $z \neq 0$ and get

$$\begin{aligned} I(Q_1, y^2 \cap F) &= 2I(Q_1, y \cap -x(x^2-1)) = 2I(Q_1, y \cap x^2 + 1) \\ &= 2I(Q_1, y \cap (x-i)(x+i)) = 2I(Q_1, y \cap x - i) = 2. \end{aligned}$$

The analogue computation for Q_2 yields $I(Q_2, y^2 \cap F) = 2$. Thus the divisor of α in this presentation is

$$\operatorname{div}(\alpha) = 4P_1 + 2P_2 - 2P_2 - 2Q_1 - 2Q_2 = 4P_1 - 2Q_1 - 2Q_2.$$

We leave as an exercise to compute $\operatorname{div}(\alpha)$ in for the second presentation and to verify that one obtains the same expression for the divisor.

Given two divisors $D_1, D_2 \in \operatorname{Div}(\mathcal{C})$ we can define a partial order by setting $D_1 \geq D_2$ as follows: let $D_1 = \sum n_P P$ and $D_2 = \sum n_Q Q$: we can add formally points to D_1 and D_2 with zero coefficients until their supports coincide so that we can write $D_1 = \sum n_{1P} P$ and $D_2 = \sum n_{2P} P$. Then $D_1 \geq D_2$ if $n_{1P} \geq n_{2P}$ for every P in the support of D_1. For example we say that $D \geq 0$ if all the coefficients of D are non-negative.

Definition 1.2.13. Given a divisor $D \in \operatorname{Div}(\mathcal{C})$ we define

$$L(D) := \{\varphi \in K(\mathcal{C}) : \operatorname{div}\varphi + D \geq 0\} \cup \{0\}.$$

Since we added the element 0 we have that $L(D)$ is a K-vector space of finite dimension. We let $\ell(D) = \dim_K L(D)$.

The importance of the vector space $L(D)$ is summarized in the following two fundamental theorems.

Theorem 1.2.14. *There exists a non-negative integer g, called the genus of $\mathcal{C}$ such that, for all divisors $D \in \operatorname{Div}(\mathcal{C})$:*

$$\ell(D) \geq \deg D + 1 - g.$$

Theorem 1.2.15 (Riemann-Roch)**.** *There exists a divisor $K_{\mathcal{C}} \in \mathrm{Div}(\mathcal{C})$, called the* ***canonical divisor*** *of $\mathcal{C}$, such that for every divisor $D \in \mathrm{Div}(\mathcal{C})$ we have*

$$\ell(D) - \ell(K_{\mathcal{C}} - D) = \deg D + 1 - g.$$

Example 1.2.16. Given m-points $P_1, \ldots, P_m$ in $\mathbb{A}^1(K) \subset \mathbb{P}^1(K)$ and positive integers $r_1, \ldots, r_m$ we let $D = \sum r_i[P_i] \in \mathrm{Div}(\mathbb{P}^1)$. By definition

$$L(D) = \{\varphi \in K(\mathcal{C}) : \mathrm{div}\,\varphi + D \geq 0\} \cup \{0\}.$$

In particular if $\varphi \in L(D)$ then the set of poles of φ is contained in $\{P_1, \ldots, P_m\}$ and each of the pole with the order of at most r_i. Such rational functions are of the form

$$\varphi = \frac{f(x)}{(x-a_1)^{s_1}(x-a_2)^{s_2}\cdots(x-a_m)^{s_m}}$$

where $f(x) \in k[x]$ of degree $s_1 + \ldots s_m$ and for every i $s_i \leq r_i$, and $P_i = (a_i : 1)$. Note that the condition $\deg f \leq \sum_i r_i$ implies that f has no pole at $(1:0)$. In particular one can compute directly that the dimension $\ell(D) = \dim L(D) = r_1 + \cdots + r_m = \deg D$. Hence using Theorem 1.2.14 we conclude that the genus of $\mathbb{P}^1$ is zero.

In order to "use" Theorem 1.2.15 one needs to know how to compute the genus g of the curve $\mathcal{C}$. There are various definition that help to compute the integer, but for *non-singular projective plane curves* one can relate the genus to the degree of the polynomial defining the curve. More precisely, if $\mathcal{C}$ is a non singular plane projective curve of degree d then the genus g of $\mathcal{C}$ satisfies

$$g = \frac{(d-1)(d-2)}{2}.$$

In particular curves of degree one and two have genus 0, and cubic curves have genus 1 (note that this is not true for *singular* curves).

The other ingredient that appear in Theorem 1.2.15 is $\ell(K_{\mathcal{C}} - D)$. One can formally define the canonical divisor using the theory of Kähler differentials. For our applications it is sufficient to know that in the case of an elliptic curve the canonical divisor satisfies the following two properties:

- $\ell(K) = 1$;
- $\deg K_{\mathcal{C}} = 0$.

Note that, from the definition it follows that, if $\deg D < 0$ then $\ell(D) = 0$. Therefore the previous properties of the canonical divisor on a curve of genus 1 give the following corollary of Theorem 1.2.15.

Corollary 1.2.17. *Let $\mathcal{C}$ be a non singular projective curve of genus 1. Then for every divisor D with $\deg D \geq 1$ one has*

$$\ell(D) = \deg D.$$

Proof. We apply Theorem 1.2.15 with $g = 1$ and the fact that $\ell(K_{\mathcal{C}} - D) = 0$ since $\deg(K_{\mathcal{C}} - D) = \deg K_{\mathcal{C}} - \deg D < 0$. □

This gives an interesting applications to the shape of divisors of rational function on an elliptic curve

Corollary 1.2.18. *Let $\mathcal{C}$ be a non singular projective curve of genus 1, and let $f \in K(\mathcal{C})$. If $\operatorname{div} f = [P] - [Q]$ then f is constant, and hence $P = Q$.*

Proof. Since $\operatorname{div} f = [P] - [Q]$ it follows that $f \in L([Q])$. Applying Corollary 1.2.17 we know that $\ell([Q]) = \deg[Q] = 1$. On the other hand the constant functions are all in $L([Q])$, which, since it is a K-vector space of dimension 1, contains only the constant functions. This implies that f is constant and in particular $P = Q$. □

We will use this fact in a different proof that the group of K-points in an elliptic curve form a group. With this goal in mind, we introduce the following definitions.

Definition 1.2.19. Given a projective non singular curve $\mathcal{C}$ we call a divisor of the form $\operatorname{div} \varphi \in \operatorname{Div}(\mathcal{C})$ a **principal divisor**.

We say that two divisors $D_1, D_2 \in \operatorname{Div}(\mathcal{C})$ are **linearly equivalent**, and we write $D_1 \sim D_2$ if the difference $D_1 - D_2$ is a principal divisor. In other words, $D_1 \sim D_2$ if there exists $\varphi \in K(\mathcal{C})$ such that $D_1 - D_2 = \operatorname{div} \varphi$.

Recall that, by our definition of $\operatorname{div} f$, for a rational function $f \in K(\mathcal{C})$, one has $\deg \operatorname{div} f = 0$, since a rational function over an algebraically closed field has the same number of zeros and poles counted with multiplicities (thanks to Bezout's Theorem).

Definition 1.2.20. We denote by $\operatorname{Div}^0(\mathcal{C})$ the subgroup of $\operatorname{Div}(\mathcal{C})$ of divisors of degree 0. We denote by $\operatorname{Princ}(\mathcal{C})$ the subgroup of $\operatorname{Div}^0(\mathcal{C})$ of principal divisors. Since all the groups involved are abelian we can form the quotients: we denote by $\operatorname{Pic}(\mathcal{C}) = \operatorname{Div}(\mathcal{C})/\operatorname{Princ}(\mathcal{C})$, called the **Picard group** of $\mathcal{C}$, and by $\operatorname{Pic}^0(\mathcal{C}) = \operatorname{Div}^0(\mathcal{C})/\operatorname{Princ}(\mathcal{C})$.

Proposition 1.2.21. *Let $\mathcal{C}$ be a non singular projective curve of genus 1 and let $\mathcal{O} \in \mathcal{C}(K)$. Then the map*

$$\begin{aligned} \mathcal{C}(K) &\to \operatorname{Pic}^0(\mathcal{C}) \\ P &\mapsto [P] - [\mathcal{O}] \end{aligned}$$

is bijective.

Proof. First we show that the map is injective. Assume that there are two points $P, Q \in \mathcal{C}(K)$ such that $[P] - [\mathcal{O}] = [Q] - [\mathcal{O}]$ in $\operatorname{Pic}^0(\mathcal{C})$: this implies that $[P] - [\mathcal{O}] \sim [Q] - [\mathcal{O}]$, i.e. there exists a rational function $f \in K(\mathcal{C})$ such that $\operatorname{div} f = [P] - [\mathcal{O}] - ([Q] - [\mathcal{O}]) = [P] - [Q]$. By Corollary 1.2.18 we conclude that $[P] = [Q]$ as wanted.

Let us fix a (class of a) degree zero divisor $D \in \operatorname{Pic}^0(\mathcal{C})$. By the Riemann-Roch Theorem for genus 1 curves, Corollary 1.2.17, since $D + [\mathcal{O}] > 0$ and $\deg(D + [\mathcal{O}]) = 1$ we have that $\ell(D + [\mathcal{O}]) = 1$. We choose a nonzero element $f \in L(D + [\mathcal{O}])$, which is

therefore a basis for the vector space. Note that, since $f \in K(\mathcal{C})$, $\deg \operatorname{div} f = 0$, and f satisfies $\operatorname{div} f + D + [\mathcal{O}] \geq 0$. In particular $\operatorname{div} f + D + [\mathcal{O}]$ is a positive divisor of degree 1, so is of the form $[P]$ for some $P \in \mathcal{C}(K)$. But then

$$[P] = \operatorname{div} f + D + [\mathcal{O}] \Rightarrow D \sim [P] - [\mathcal{O}],$$

so the map is surjective. □

We can use the bijection of Proposition 1.2.21 to define a group structure on $\mathcal{C}(K)$ as follows: given $P, Q, R \in \mathcal{C}(K)$ we have $P + Q = R$ if and only if

$$[P] - [\mathcal{O}] + [Q] - [\mathcal{O}] = [R] - [\mathcal{O}]$$

using the operation in $\operatorname{Pic}^0(\mathcal{C})$. In fact we can show that this operation is the same we have defined using the chord and tangent construction in Theorem 1.2.1. To see this let $P, Q, R \in \mathcal{C}(K)$ such that $P \oplus Q = R$. Recall that this means that $R = (P * Q) * \mathcal{O}$, i.e. R is the third point of intersection of $\ell_{(P*Q)\mathcal{O}} \cap \mathcal{C}$, where $P * Q$ is the third point of intersection of $\ell_{PQ} \cap \mathcal{C}$. Denote $\ell_1 = \ell_{PQ}$ and $\ell_2 = \ell_{(P*Q)\mathcal{O}}$. Then we can define $\varphi = \ell_1/\ell_2 \in K(\mathcal{C})$, where we identify the lines with their equations. Since we know the intersections $\ell_1 \cap \mathcal{C}$ and $\ell_2 \cap \mathcal{C}$ we can compute the divisor associated to φ as

$$\operatorname{div} \varphi = [P] + [Q] + [P * Q] - [P * Q] - [\mathcal{O}] - [R] = [P] + [Q] - [R] - [\mathcal{O}].$$

In particular $[P] + [Q] \sim [R] - [\mathcal{O}]$ so that $P + Q = R$ using the operation induced by $\operatorname{Pic}^0(\mathcal{C})$.

Note that, using Proposition 1.2.21, we do not need to verify that the operation satisfies all the axiom of the group, since *by definition*, $\operatorname{Pic}^0(\mathcal{C})$ is a group! On the other hand the above construction has been made only for algebraically closed fields: it can be modified to apply to more general fields, e.g. perfect fields, but we will not to this here.

1.2.2 Elliptic curves and equations

In this subsection we use the theory developed in the last subsection to show that every elliptic curve has a special defining equation in the plane. Recall that we have defined an elliptic curve E over a (perfect) field K to be a non singular projective curve of genus 1 defined over K with a K-rational point $\mathcal{O} \in E(K)$. In the literature one finds usually four different definitions:

(a) A non singular projective plane curve E over K of degree 3 together with a point $\mathcal{O} \in E(K)$;

(b) A non singular projective plane curve E over K of degree 3 together with an inflection point $\mathcal{O} \in E(K)$;

(c) A non singular projective plane curve E over K defined by an equation of the form

$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3 \tag{1.1}$$

(d) A non singular projective curve E over K of genus 1 together with a point $\mathcal{O} \in E(K)$.

We will now show that all these definition are in fact equivalent. Before stating the first proposition we introduce the concept of morphism between curves.

Definition 1.2.22. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be plane projective curves defined over a field K. A **rational map** $\phi : \mathcal{C}_1 \to \mathcal{C}_2$ is a projective triple $(\phi_x : \phi_y : \phi_z) \in \mathbb{P}^2(K(\mathcal{C}))$ such that for every point $P \in \mathcal{C}_1(\overline{K})$ where $\phi_x(P), \phi_y(P), \phi_z(P)$ are defined and not all zero, then the projective point $[\phi_x(P) : \phi_y(P) : \phi_z(P)]$ lies in $\mathcal{C}_2(\overline{K})$.

We say that the map ϕ is **regular**, or defined, at P if there exists $g \in K(\mathcal{C})^\times$ such that $g\phi_x, g\phi_y$ and $g\phi_z$ are all defined at P and not all zero. We say that ϕ is a **morphism** (or regular) if it is regular at every $P \in \mathcal{C}_1(K)$. A morphism $\phi : \mathcal{C}_1 \to \mathcal{C}_2$ is called an **isomorphism** if there exists a morphism $\phi^{-1} : \mathcal{C}_2 \to \mathcal{C}_1$ such that $\phi \circ \phi^{-1}$ and $\phi^{-1} \circ \phi$ are the identity maps.

Note that the function g in the previous definition in general depends on P. The idea is that the ratios ϕ_x/ϕ_y, ϕ_x/ϕ_z and ϕ_y/ϕ_z have the same values, where they are defined, even when we multiply each function by g (at points where g is not zero). Therefore, when they are both defined, the values $(\phi_x : \phi_y : \phi_z)(P)$ and $(g\phi_x : g\phi_y : g\phi_z)(P)$ coincide. In fact one can show that at a smooth point P, a rational map with domain a curve is always regular at P.

Lemma 1.2.23. *Let P be a smooth point of a projective curve $\mathcal{C}_1$ and let $\Phi : \mathcal{C}_1 \to \mathcal{C}_2$ a rational map. Then Φ is regular at P. In particular if $\mathcal{C}_1$ is non singular then Φ is a morphism.*

Morphism of projective curves have very strict behaviour as illustrated by the following Theorem.

Theorem 1.2.24. *A morphism between projective curves is either constant or surjective.*

Example 1.2.25. Consider the two curves $\mathcal{C}_1 : x = 0$ and $\mathcal{C}_2 : y = 0$. We define the following rational map:

$$\Phi : \mathcal{C}_1 \to \mathcal{C}_2 \qquad \Phi = \left[\frac{x}{y} : 0 : \frac{z}{y}\right] \in \mathbb{P}^2(\mathbb{C}(\mathcal{C}_1)).$$

It is not hard to see that, by construction, for every $P \in \mathcal{C}_1(\mathbb{C})$ in which the map is defined $\Phi(P) \in \mathcal{C}_2(\mathbb{C})$ (in fact everything we will say will not depend on the field $\mathbb{C}$). Given a point $P = (0 : b : c) \in \mathcal{C}_1(\mathbb{C})$ if $bc \neq 0$ then $\Phi(P) = [0 : 0 : 1]$ and P is regular at P. If $b = 0$, i.e. $P = (0 : 0 : 1)$ then multiplying each component of Φ by $g = y/z \in \mathbb{C}(\mathcal{C}_1)$ we get that $g\Phi = [x/z : 0 : 1]$ and $\Phi(P) = (0 : 0 : 1)$. Finally it is sufficient to notice that, as elements of $\mathbb{C}(\mathcal{C}_1)$ one has $x/z = 0$ since $x \cdot 1 - z \cdot 0 \in (x)$ and hence $\Phi(0 : 1 : 0) = (0 : 0 : 1)$.

In particular the morphism Φ is constant, even if its expression is given by non constant rational functions in $K(\mathcal{C}_1)$.

We end this discussion by citing a criterion for a morphism to be an isomorphism. Note that any non constant rational map $\phi : \mathcal{C}_1 \to \mathcal{C}_2$ induces, by composition, an inclusion of function fields $\phi^* : K(\mathcal{C}_2) \subset K(\mathcal{C}_1)$. This extension of fields is a finite extension, and we define the **degree** of the morphism ϕ to be the degree $[K(\mathcal{C}_1) : \phi^*(K(\mathcal{C}_2))]$.

Lemma 1.2.26. *Let $\Phi : \mathcal{C}_1 \to \mathcal{C}_2$ be a non constant map of non singular projective curves. Assume that the field extension $K(\mathcal{C}_1) \supset \phi^*(K(\mathcal{C}_2))$ is separable[3]. Then for all but finitely many $Q \in \mathcal{C}(K)$*

$$\#\Phi^{-1}(Q) = \deg \Phi.$$

Moreover, if $\deg \Phi = 1$ then Φ is an isomorphism.

We can now show the equivalence of definitions (c) and (d).

Theorem 1.2.27. *Let E be a non singular projective curve of genus 1 over a field K with a point $\mathcal{O} \in E(K)$ then there is a isomorphism between E and $\mathcal{C}$ where $\mathcal{C}$ is defined by*

$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3.$$

Proof. Since E has genus 1, Corollary 1.2.17, shows that $\ell(m[\mathcal{O}]) = m$ for every $m \geq 1$. Since the constant functions lie in $L([\mathcal{O}])$ we have that $L([\mathcal{O}]) \equiv K$, and $\{1\}$ is a basis. Since $\ell(2[\mathcal{O}]) = 2$ there exists a non constant function $x \in K(E)$ such that $\{1, x\}$ is a basis of $L(2[\mathcal{O}])$. Similarly we can choose a function $y \in L(3[\mathcal{O}])$ so that $\{1, x, y\}$ are a basis of $L(3[\mathcal{O}])$. Then, $\{1, x, x^2, y\}$ is a basis of $L(4[\mathcal{O}])$, and $\{1, x, x^2, y, X\}$ is a basis of $L(5[\mathcal{O}])$.

The set $\{1, x, x^2, y, X, x^3, y^2\}$ is contained in $L(6[0])$, which is a K-vector space of dimension 6. This implies that the seven functions have to be linearly dependent, i.e. there exists $a_i \in K$ such that

$$a_0y^2 + a_1X + a_3y = a_0'x_3 + a_2x^3 + a_4x + a_6,$$

and the equality holds as functions, regular outside of the only pole $\mathcal{O}$. Note that, since $\{1, x, x^2, y, X\}$ are linearly independent, a_0 and a_0' are both different from zero (at least one has to be for dimension reason, and if exactly one was zero, the other will give a linear combination of functions in $L(5[\mathcal{O}])$ equal a function that has a pole of order strictly larger that 5 at $\mathcal{O}$). Therefore, after replacing y with a_0y/a_0' and x with a_0x/a_0', and multiplying all the terms by $a_0'^2/a_0^3$ we can suppose $a_0 = a_0' = 1$. We define the map $E \smallsetminus \{\mathcal{O}\} \to \mathcal{C}$ given by $P \mapsto (x(P), y(P))$ which maps the affine curve $E \smallsetminus \{\mathcal{O}\}$ into the plane affine curve $\mathcal{C}$ with equation

$$y^2 + a_1X + a_3y = x_3 + a_2x^2 + a_4x + a_6.$$

We now show that the map $P \mapsto (x(P), y(P))$ is an isomorphism. First we consider the rational function $x \in K(E)$: by construction it has a single pole of order exactly two at $\mathcal{O}$, and therefore only two zeros. We claim that the composition

$$E \smallsetminus \{\mathcal{O}\} \to \mathcal{C} \to \mathbb{A}^1 \qquad P \mapsto (x(P), y(P)) \mapsto x(P)$$

is 2:1. So in particular, by Lemma 1.2.26, $\deg x = 2$. This follows from the fact that given any $c \in K$ the equation of $\mathcal{C}$ with x replaced by c has almost always 2 solution in a

[3]Recall that a field extension $L \supset K$ is separable if every element of L has separable minimal polynomial over K.

finite extension L of K. Similarly the map defined using the rational function $y \in K(E)$ has $\deg y = 3$ since it is almost always $3:1$.

To conclude is enough to observe that the degree of the map $P \mapsto (x(P), y(P))$ has to divide both $\deg x$ and $\deg y$ which implies that it is a map of degree 1 and hence, by Lemma 1.2.26, an isomorphism. □

Now we show the equivalence between (b) and (c).

Proposition 1.2.28. *Let $\mathcal{C}$ be a non singular cubic projective plane curve over K and $\mathcal{O} \in \mathcal{C}(K)$ a point of inflection. Then*

1. *After an invertible linear change of variables with coefficients in K, the point $\mathcal{O}$ will have coordinates $(0:1:0)$ and the tangent line to $\mathcal{C}$ at $\mathcal{O}$ will be $L : z = 0$;*
2. *If $\mathcal{O} = (0:1:0)$ with tangent line $L : z = 0$ then the equation of $\mathcal{C}$ is of the form (1.1):*
$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3.$$

Proof. Let $(r:s:t) \in \mathbb{P}^2(K)$, after possibly interchanging y and z we can assume $s \neq 0$. Then the regular map

$$(x:y:z) \mapsto (sx - ry : sy : sz - ty) : \mathbb{P}^2 \to \mathbb{P}^2$$

sends $(r:s:t)$ to $(0:s^2:0) = (0:1:0)$ and it is an isomorphism. Thus we can suppose $\mathcal{O} = (0:1:0)$. Let

$$L : ax + by + cz = 0$$

the tangent line at $(0:1:0)$ with $a, b, c \in K$ and not all zero. Let $A = (a_{ij})$ be any invertible 3×3 matrix whose first two columns are orthogonal to (a, b, c). Define a change of variables by

$$A \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

The equation of L in the variables x', y', z' becomes

$$0 = (a, b, c) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = (a, b, c)\, A \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = (0, 0, d) \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = dz'.$$

Moreover $d \neq 0$ and so the equation of L becomes $Z' = 0$. This shows the first assertion.

For the second assertion let us start with the general cubic form defining $\mathcal{C}$, i.e. a polynomial of the form

$$F(x, y, z) = c_1x^3 + c_2x^2y + c_3x^2z + c_4xy^2 + c_5xyz + c_6xz^2 + c_7y^3 + c_8y^2z + c_9yz^2 + c_{10}z^3.$$

Since $\mathcal{C}$ is non singular the polynomial F is absolutely irreducible.

Because $\mathcal{O} = (0:1:0) \in \mathcal{C}(K)$ then $\boxed{c_7 = 0}$. In the affine patch $U_2 : y = 1$ the equation of $\mathcal{C}$ becomes:

$$F(x, y, z) = c_1x^3 + c_2x^2 + c_3x^2z + c_4x + c_5xz + c_6xz^2 + c_8z + c_9z^2 + c_{10}z^3.$$

Since the tangent line at $(0:0)$ is given by $c_4x + c_8z = 0$ and it has to equal $L_\infty : z = 0$ we get $\boxed{c_4 = 0}$ and $\boxed{c_8 \neq 0}$.

Moreover $\mathcal{O}$ is an inflection point, so the intersection number with the line L_∞ has to be ≥ 3. On the other hand

$$I(\mathcal{O}, L_\infty \cap \mathcal{O}) = I(z, F(x,1,z)) = I(z, c_1x^3 + c_2x^2)$$

which is ≥ 3 only if $\boxed{c_2 = 0}$.

Thus we get that our curve has an equation of the form

$$F(x,y,z) = c_1x^3 + c_3x^2z + c_5xyz + c_6xz^2 + c_8y^2z + c_9yz^2 + c_{10}z^3.$$

Moreover $c_1 \neq 0$ since otherwise the polynomial is divisible by z (and hence not irreducible). We can therefore divide by c_1 and then replace z by $-c_1z/c_8$ to obtain an equation of the form (1.1) as wanted. □

The equation of the form (1.1) is called a **Weierstrass equation** for an elliptic curve. If the characteristic of the field is $\neq 2, 3$ then we can further simplify the form of the equation.

Proposition 1.2.29. *Let K be a field of characteristic $\neq 2, 3$. Then every elliptic curve E with inflection point $\mathcal{O}$ is isomorphic to a curve of the form*

$$E_{A,B} : y^2z = x^3 + Axz + Bz^3$$

for some B, where the inflection point is $(0:1:0)$. Moreover every curve $E_{A,B}$ is nonsecular if and only if $4A^3 + 27B^2 \neq 0$.

Proof. We already showed that the condition $4A^3 + 27B^2 \neq 0$ guarantees that the curve $E_{A,B}$ is non singular and has an inflection point in $(0:1:0)$ so it is an elliptic curve.

For the converse Proposition 1.2.28, any elliptic curve is isomorphic to a curve of the form

$$y^2z + a_1xyz + a_3yz^2 = x^3 + a_2x^2z + a_4xz^2 + a_6z^3.$$

We begin by completing the square on the left hand side, substituting

$$x' = x \qquad y' = y + \frac{a_1}{2}x \qquad z' = z$$

we obtain an equation in x', y', z' without the xyz monomial. Similarly one can check the following change of variables eliminates the yz^2 and x^2z monomials

$$x'' = x' + \frac{a_2}{3} \qquad y'' = y' + \frac{a_3}{2} \qquad z' = z.$$

We thus get an equation of the form

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

□

We call an equation of the form $y^2 = x^3 + Ax + B$ a **short Weierstrass form**. The previous Proposition shows that every elliptic curve $\mathcal{C}$ over K, where $\operatorname{char} K \neq 2, 3$, is isomorphic to a curve of the form $E_{A,B}$ and we say that the equation of $E_{A,B}$ is a short Weierstrass form for $\mathcal{C}$. The short Weierstrass from allows us to study explicitly isomorphisms between elliptic curves of the form $E_{A,B}$ and fully describe them.

Lemma 1.2.30. *Let $E_{A,B}$ and $E_{A',B'}$ be two elliptic curves in short Weierstrass form. If*

$$\varphi : E_{A',B'} \to E_{A,B} \qquad \varphi(\mathcal{O}) = \mathcal{O},$$

is an isomorphism, then there exists $c \in K^\times$ such that $A' = c^4 A$, $B' = c^6 B$ and the map φ is given by

$$(x : y : z) \mapsto (c^2 x : c^3 y : z).$$

Conversely, if $A' = c^4 A$ and $B' = c^6 B$ for some $c \in K^\times$, then the map $\varphi(x : y : z) = (c^2 x : c^3 y : z)$ is an isomorphism between the curve $E_{A',B'}$ and $E_{A,B}$ that sends $\mathcal{O}$ to $\mathcal{O}$.

Proof. Consider the functions $x \circ \varphi$ and $y \circ \varphi$. Since φ is an isomorphism then $x \circ \varphi \in L(2\mathcal{O})$ and $y \circ \varphi \in L(3\mathcal{O})$ (see the proof of Theorem 1.2.27). But since $\{1, x'\}$ and $\{1, x', y'\}$ are basis for the two vector spaces we know that there exist $u_1, u_2 \in K^\times$ and $r, s, t \in K$ such that

$$x \circ \varphi = u_1 x' + r \qquad y \circ \varphi = u_2 y' + s x' + t.$$

On the other hand φ induces a ring homomorphism

$$\varphi^* : \frac{k[x, y]}{(y^2 - x^3 - Ax - B)} \to \frac{k[x', y']}{(y'^2 - x'^3 - A'x' - B')}$$

defined by $\varphi^*(f) = f \circ \varphi$. Since $y^2 - x^3 - Ax - B = 0$ the fact that φ^* is an homomorphism implies that

$$(u_2 y' + s x' + t)^2 = (u_1 x' + r)^2 + A(u_1 x + r) + B.$$

But, by definition, any combination of powers of x', y' that is zero is a multiple of the equation $E_{A',B'}$. This implies that $r = s = t = 0$ and $u_2^2 = u_1^3$. Moreover we get $A' = c^4 A$ and $B' = c^6 B$ with $c = u_1/u_2$. In particular we get $x = u_1 x' = c^2 x'$ and $y = u_2 y' = c^3 y'$. □

Given an elliptic curve in Weierstrass form we can define an invariant that plays an important role in the study of elliptic curves.

Definition 1.2.31. Given an elliptic curve over a field K, with $\operatorname{char} K \neq 2, 3$, with Weierstrass equation $E_{A,B}$ the **j-invariant** of the elliptic curve is defined as

$$j(E) := \frac{1728(4A^3)}{4A^3 + 27B^2} = \frac{1728(4A^3)}{\Delta},$$

where $\Delta = 4A^3 + 27B^2$ is the **discriminant** of $E_{A,B}$ (sometimes referred to the discriminant of E).

Theorem 1.2.32. *Given two elliptic curves E_1, E_2 over a field K of* $\operatorname{char} K \neq 2, 3$*, E_1 is isomorphic to E_2 over $\overline{K}$ if and only if $j(E_1) = j(E_2)$; Moreover, given $j_0 \in \overline{K}$, there exists E_0 elliptic curve over $\overline{K}$ such that $j(E_0) = j_0$.*

Proof. If E_1 and E_2 are isomorphic, then we can apply Lemma 1.2.30 and conclude that the two j-invariants are equal.

Conversely suppose that the two j-invariants are equal. Let us fix Weierstrass equation for the two curves:

$$E_1 : y^2 = x^3 + A_1 x + B_1 \qquad E_2 : y^2 = x^3 + A_2 x + B_2.$$

The hypothesis $j(E_1) = j(E_2)$ yields

$$4(A_1A_2)^3 + 27A_1^3B_2^2 = 4(A_1A_2)^3 + 27A_2^3B_1^2,$$

so in particular we get $A_1^3B_2^2 = A_2^3B_1^2$. Together with the fact that the discriminant of both equations is difference from zero, it follows that $A_1 = 0$ if and only if $A_2 = 0$ and $B_1B_2 \neq 0$, and similarly $B_1 = 0$ if and only if $B_2 = 0$ and $A_1A_2 \neq 0$.

By Lemma 1.2.30, we know that to give an isomorphism between E_1 and E_2 it is sufficient to determine an element $c \in \overline{K}^\times$ such that $A_2 = c^4A_1$ and $B_2 = c^6B_1$. We distinguish three cases:

$\boxed{A_1 = 0}$ As noted above this implies $A_2 = 0$ and $B_1B_2 \neq 0$ (note that the only value of j in this case is $j = 0$). If we had an isomorphism $E_1 \to E_2$ this will send $x \mapsto c^2x$ and $y \mapsto c^3y$. This maps the curve E_1 into the curve of equation

$$y^2 = x^3 + \frac{B_1}{c^6}.$$

If we choose c such that $c^6 = B_1/B_2$ then we obtain the equation of E_2 as wanted.

$\boxed{B_1 = 0}$ The same argument as above shows that we can choose c such that $c^4 = A_1/A_2$.

$\boxed{A_1B_1 \neq 0}$ In this case the substitution yields

$$y^2 = x^3 + \frac{A_1}{c^4}x + \frac{B_1}{c^6}.$$

In this case we want $c^4 = A_2/A_1$ and $c^6 = B_2/B_1$. Such a $c \in K^\times$ exists since the condition $j(E_1) = j(E_2)$ implies

$$\left(\frac{A_2}{A_1}\right)^3 = \left(\frac{B_2}{B_1}\right)^2.$$

□

Note that in general elliptic curves over a field K with the same j-invariant are not necessarily isomorphic over K, i.e. there might not exists $c \in K^\times$ that induces an isomorphism. As an example we can consider the following two curves defined over $\mathbb{Q}$:

$$E : y^2 = x^3 + 1 \qquad E' : y^2 = x^3 + 8.$$

They have both j-invariant equal to 0 but the isomorphism requires $c = \sqrt[3]{2} \notin \mathbb{Q}$.

Legendre Form

Sometimes it is useful to have an equation for an elliptic curve that depends only on one parameter. We have seen in the previous paragraph that up to isomorphism an elliptic curve over $\overline{K}$ is identified via the j-invariant. We introduce now another useful form for an equation of an elliptic curve.

Definition 1.2.33. An Elliptic curve defined over a field K is in **Legendre form** if its equation have the form

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

for some $\lambda \in K$.

Proposition 1.2.34. *Let us assume that* $\operatorname{char} K \neq 2$. *Then every elliptic curve is isomorphic (over* $\overline{K}$*) to an elliptic curve in Legendre form*

$$E_\lambda : y^2 = x(x-1)(x-\lambda)$$

for some $\lambda \in \overline{K} \setminus \{0,1\}$. *Moreover we have the following two properties*

(a) the j-invariant of E_λ can be computed as follows:

$$j(E_\lambda) = \frac{2^8(\lambda^2-\lambda+1)^3}{\lambda^2(\lambda-1)^2};$$

(b) The map $j : \lambda \mapsto j(E_\lambda)$ is surjective, and for every $j_0 \neq \{0, 1728\}$, there are exactly 6 elliptic curves in Legendre form with j-invariant equal to j_0. Moreover, if $\operatorname{char} K \neq 3$ *we have $\#j^{-1}(0) = 2$ and $\#j^{-1}(1728) = 3$, while if* $\operatorname{char} K = 3$ *$\#j^{-1}(0) = \#j^{-1}(1728) = 3$.*

Proof. Using the proof of Theorem 1.2.27 and the proof of Proposition 1.2.29 we know that, when $\operatorname{char} K \neq 2$, any elliptic curve is isomorphic to a curve with equation

$$y^2 = x^3 + b_2x^2 + b_4x + b_6$$

for some $b_2, b_4, b_6 \in \overline{K}$. Therefore we can find $e_1, e_2, e_3 \in \overline{K}$ such that the equation of the curve becomes

$$y^2 = (x-e_1)(x-e_2)(x-e_3).$$

Moreover, since the curve is not singular we have that $e_i \neq e_j$ for every $i \neq j$. The goal is to show that there exists an invertible change of coordinates such that $\{e_1, e_2, e_3\}$ gets send to $\{0, 1, \lambda\}$ for some $\lambda \in \overline{K}$. We use the following transformation:

$$\begin{cases} x = (e_2-e_1)x' + e_1 \\ y = (e_2-e_1)^{3/2}y' \end{cases}$$

Then the three linear factors become

$$\begin{aligned} x - e_1 &= (e_2-e_1)x' \\ x - e_2 &= (e_2-e_1)(x'-1) \\ x - e_3 &= (e_2-e_1)\left(x' - \frac{e_3-e_1}{e_2-e_1}\right) \end{aligned}$$

This shows that with this change of variables we get an equation in Legendre form E_λ with $\lambda = (e_3 - e_1)/(e_2 - e_1)$.

Part (a) follows from a direct computation using the formulas for the j-invariant.

The previous discussion (and the results on the Weierstrass equation) shows that the map j in part (b) is surjective. Let us now show that the map is 6 to 1. To do this we discuss when to elliptic curves E_λ, E_μ in Legendre form satisfy $j(E_\lambda) = j(E_\mu)$. Since they have the same j-invariant by Theorem 1.2.32 there is an isomorphism $E_\lambda \to E_\mu$.

Recall from the proof of Lemma 1.2.30 that any isomorphism (that might not necessarily preserve the Weierstrass form) has the following shape:

$$\begin{cases} x = c^2 x' + r \\ y = c^3 y' + s x' + t \end{cases}$$

On the other hand, this isomorphism sends the Legendre form into a Legendre form. Thus $s = t = 0$, and $x = c^2 x' + r$ and $y = c^3 y'$. Substituting this into the equation of E_λ we get

$$y'^2 = \left(x' + \frac{r}{c^2}\right)\left(x' + \frac{r-1}{c^2}\right)\left(x' + \frac{r-\lambda}{c^2}\right).$$

This equation is the equation of E_μ if the right hand side correspond with $x(x-1)(x-\mu)$. This is equivalent to the fact that

$$\{0, 1, \mu\} = \left\{\frac{r}{c^2}, \frac{r-1}{c^2}, \frac{r-\lambda}{c^2}\right\}.$$

There are exactly 6 bijection of the two sets. For example $r/c^2 = 0$ implies $r = 0$, and with this case $\lambda = c^2$ therefore $\mu = 1/\lambda$. Computing all possible options one gets that $j(E_\lambda) = j(E_\mu)$ if and only if

$$\mu \in \left\{\lambda, \frac{1}{\lambda}, 1-\lambda, \frac{\lambda-1}{\lambda}, \frac{1}{1-\lambda}, \frac{\lambda}{\lambda-1}\right\}.$$

This shows that the map j is 6 to 1 if the six values in the set are distinct. Let us discuss the various cases

1. If $\lambda = 1/\lambda$ then $\lambda^2 = 1$. Since $\lambda \neq 1$ this implies $\lambda = -1$ (which corresponds to $j = 1728$). In this case the six options reduce to $\{-1, 2, 1/2\}$.
2. If $\lambda = 1 - \lambda$ then $\lambda = 1/2$ and we are in the previous case.
3. If $\lambda = \lambda/\lambda - 1$ then $\lambda = 2$ (since $\lambda \neq 0$), and we are again in the previous case.
4. If $\lambda = 1/1 = -\lambda$, then $\lambda^2 - \lambda + 1 = 0$. Then $\lambda = (1 \pm \sqrt{-3})/2$ and in this case the six options reduce to $\{(1 + \sqrt{-3})/2, (1 - \sqrt{-3})/2\}$. In this case $j = 0$.
5. If $\lambda = \lambda - 1/\lambda$ we are in the previous case.

Finally in the case char $K = 3$ all the previous cases coincide (note that $1720 \equiv 0 \bmod 3$) and all the values collapse into a single solution $\mu = -1$. □

Group Law in Algebraic Terms We want to describe the group structure on an elliptic curve in Weierstrass from with explicit equation. More precisely we start with an elliptic curve E over a field K (of characteristic different from 2 and 3) given by

$$E : y^2 = x^3 + Ax + B$$

and two point $P = (x_1, y_2)$ and $Q = (x_2, y_2)$ in $E(K)$. We want to explicitly write the coordinates of $R = P \oplus Q = (x_3, y_3)$ in terms of x_1, x_2, y_1, y_2. Since $\mathcal{O}$ is the identity of the group we can assume that $P \neq \mathcal{O}$ and $Q \neq \mathcal{O}$, which explain why we can write P and Q in the affine chart where $\mathcal{O}$ is the point at infinity. We distinguish two cases

$\boxed{x_1 \neq x_2}$ In this case the line $\overline{PQ}$ has slope

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

so that the equation of $\overline{PQ}$ is $y - y_1 = m(x - x_1)$. By definition of the group operation the third point of intersection $(\overline{PQ} \cap E) \smallsetminus P, Q$ is $-R = (x_3, -y_3)$. Therefore, the coordinate x_3 is a solution of

$$(m(x - x_1) + y_1)^2 = x^3 + Ax + B.$$

This is a cubic equation in x of the form

$$g(x) = x^3 - m^2 + \cdots = 0.$$

We know that the polynomial g has two roots in K, namely x_1, x_2 and therefore it factors in $K[x]$ as a product of three linear terms

$$g(x) = (x - x_1)(x - x_2)(x - x_3)$$

We can compare the term of degree two with our previous expression to obtain $m^2 = x_1 + x_2 + x_3$, i.e. $x_3 = m^2 - x_1 - x_2$. We can compute the coordinate $-y_3$ using the equation of the line $\overline{PQ}$. Then, we obtain:

$$\begin{aligned} m &= \frac{y_2 - y_1}{x_2 - x_1} \\ x_3 &= m^2 - x_1 - x_2 \\ y_3 &= m(x_1 - x_3) - y_1 \end{aligned}$$

We have expressed the coordinates of $R = P \oplus Q$ as rational function of the coordinates of P and Q as wanted. Ignoring additions and subtractions (which costs is usually negligible) we have to perform 3 multiplications and one inversion in the field K.

$\boxed{x_1 = x_2}$ In this case we have $y_1 = \pm y_2$. If $y_1 = -y_2$ then $Q = -P$ and $R = P \oplus Q = \mathcal{O}$. Therefore we can assume $y_1 = y_2$ and $Q = P$ so that $R = 2P$ and the line $\overline{PQ}$ is

the tangent line at E in P. One computes the slope of the line (for example with implicit differentiation) and obtains

$$m = \frac{3x_1^2 + A}{2y_1}.$$

Using the same strategy as in the previous case one then computes equations for x_3, y_3 in terms of x_1, y_1. In this case one needs to perform 4 multiplications and one inversion.

At the following link one gets an implementation in SageMath (CoCalc) of the above formulas and algorithms that check the associativity property of $\oplus$:

CoCalc Worksheet - credit to Andrew Sutherland

2 Isogenies

In almost every branch of mathematics, when one studies a certain category of mathematical objects with prescribed properties, the maps between objects play a crucial role. In the case of groups or rings we have homomorphisms, for vector spaces we have linear maps and for topological spaces we have continuous functions. For elliptic curves (and more in general for abelian varieties), the structure-preserving maps are called *isogenies*. First of all, we note that the elliptic curves has essentially two structures: an algebraic one (because they have an abelian group structure) and a geometric structure (since they are smooth projective curves), so we want to consider maps which are both group homomorphisms and morphisms of algebraic curves.

We have already define morphism (and more general rational maps) between algebraic curves, but in general a morphism between two elliptic curves does not have to preserve the group structure, i.e. it might not be a homomorphism of the group of points. We will now define and characterize those maps that preserve this structure and study their properties.

2.1 Isogenies of elliptic curves

Structure preserving maps between elliptic curves which will play a key role in the course (both in theory and applications).

Definition 2.1.1. An *isogeny* $\phi : E_1 \to E_2$ between elliptic curves over K is a non-constant morphism of curves defined over K such that $\phi(\mathcal{O}_1) = \mathcal{O}_2$. The elliptic curves E_1 and E_2 are said to be *isogenous* over K.

Recall that a morphism of curves is said to be defined over K if it can be represented by rational maps whose coefficients lie in K. In general, if E_1 and E_2 are elliptic curves defined over K and L/K is an algebraic extension, we say that the two elliptic curves are isogenous over L if there exists an isogeny $\phi : E_1 \to E_2$ defined over L.[1]

Notice that, with our definition, the zero-map (i.e. the morphism sending everything to $\mathcal{O}_2$ is not an isogeny); this is a general convention which requires isogenies to preserve dimensions, and depends on the book you are considering (for example, in [Sil09] the authors includes the zero-morhpism in the definition). We will still have the occasion to refer to the zero-morphism, but we will not call it an isogeny.

Let us notice moreover, by Theorem 1.2.24, every non-constant morphism between elliptic curves is surjective, hence we are considering only surjective maps. The following

[1] We will see in an example that it can happen that curves defined over K are isogenous over a bigger field but not over K.

theorem ensures that these are exactly the structure-preserving maps that we want to consider.

Theorem 2.1.2. *Any morphism of elliptic curves $\phi : E_1 \to E_2$ that preserves the identity element induces a group homomorphism, i.e. for every $P_1, P_2 \in E_1(\overline{K})$, we have $\phi(P_1 + P_2) = \phi(P_1) + \phi(P_2)$.*[2]

For a proof of this theorem see [Sil09, Theorem III.4.8].

Definition 2.1.3. A morphism from an elliptic curve E/K to itself that fixed the distinguished element is called an *endomorphism*. An endomorphism that is also an isomorphism[3] is called *automorphism*.

Except for the zero-morphism, every endomorphism is an isogeny. We will see later in the course that the set of endomophisms of an elliptic curves has a natural ring structure.

2.2 Examples of isogenies

We are now ready to see some examples of isogenies which are also endomorphisms of an elliptic curve E/K defined by a short Weierstrass equation $y^2 = x^3 + Ax + B$ (assume $\mathrm{char}(K) \neq 2, 3$).

2.2.1 The negation map

Let us consider the map $\phi_1 : E \to E$ sending each point P in $-P$. In projective coordinates, this is given by

$$[x : y : z] \mapsto [x : -y : z],$$

hence it is clearly a rational map. Moreover, it is defined at every projective point, hence is a morphism. Finally, it fixes the zero element $\mathcal{O} = [0 : 1 : 0]$ and it is not constant, hence it is an isogeny. Notice that this is also an automophism of the elliptic curve, since it is an endomorphism and an isomorphism (it is its own inverse).

2.2.2 The multiplication-by-2 map

Let E/K be an elliptic curve defined by $y^2 = x^3 + Ax + B$, and let $\phi : E \to E$ be the duplication map, i.e. sending $P \mapsto 2P$. This is obviously a non-constant group homomorphism; let us now show that this is actually a morphism of projective curves.

[2] This is true in more general setting, namely for abelian varieties.

[3] Recall that an isomorphism between two algebraic varieties $\phi : V_1 \to V_2$ is a morphism such that there exists another morphism $\phi' : V_2 \to V_1$ such that their compositions $\phi \circ \phi'$ and $\phi' \circ \phi$ are the identities.

Recall that, if $P = (x_1, y_1)$ in affine coordinates, the formulas for double the points (which holds in the case $y_1 \neq 0$, i.e. $2P \neq \mathcal{O}$) gives

$$\lambda = \frac{3x_1^2 + A}{2y_1}, \qquad \mu = \frac{-x_1^3 + Ax_1 + 2B}{2y_1} = y_1 - \lambda x_1$$

and

$$x_3 = \lambda^2 - 2x_1 \qquad y_3 = -\lambda x_3 - \mu,$$

which gives

$$2P = \left(\frac{(3x_1^2 + A)^2 - 8x_1y_1^2}{4y_1^2}, \frac{12x_1y_1^2(3x_1^2 + A) - (3x_1^2 + A)^3 - 8y_1^4}{8y_1^3} \right).$$

Homogenizing these and clearing the denominators we get the rational map $\phi = (\psi_x : \psi_y : \psi_z)$, where

$$\begin{aligned}
\psi_x(x, y, z) &= 2yz((3x^2 + Az^2)^2 - 8xy^2z), \\
\psi_y(x, y, z) &= 12xy^2z(3x^2 + Az^2) - (3x^2 + Az^2)^3 - 8y^4z^2, \\
\psi_z(x, y, z) &= 8y^3z^3.
\end{aligned}$$

Notice that, if $y = 0$, then $\phi(P) = [0 : 1 : 0]$, giving what expected also on the points of order 2. Now, if $y \neq 0$, then $(3x^2 + Az^2) \neq 0$, (because the curve $y^2z - x^3 - Axz^2 - B = 0$ is non-singular), hence the only point where ϕ_x, ϕ_y and ϕ_z simultaneously vanish is the point $\mathcal{O} = [0 : 1 : 0]$, so the map ϕ is defined everywhere except for the point $\mathcal{O}$. As a rational map of smooth projective curves, we know that ϕ is a morphism (see Theorem **??**), hence defined everywhere, so there must be an alternative representation of ϕ that we can evaluate at the point $\mathcal{O}$. On the other hand, we know a priori that $\phi(\mathcal{O}) = \mathcal{O}$ since $2 \cdot \mathcal{O} = \mathcal{O}$, but let us verify it explicitely.

Let us now find that there is another representation of the rational map ϕ which we can use to evaluate it in $\mathcal{O}$. Indeed, in projective coordinates the equation of the curve is $f(x, y, z) = y^2z - x^3 - Axz^2 - Bz^3 = 0$. Recall that we can add to any of of the ϕ_i a suitable multiple of f without changing the rational function ϕ they define. Let us then replace the function ψ_x with $\psi_x + 18xyzf$ and ϕ_y with $\psi_y + (27f - 18y^2z)f$, and remove the common factor z^2 to obtain

$$\begin{aligned}
\psi_x(x, y, z) &= 2y(xy^2 - 9Bxz^2 + A^2z^3 - 3Ax^2z), \\
\psi_y(x, y, z) &= y^4 - 12y^2z(2Ax + 3Bz) - A^3z^4 + 27Bz(2x^3 + 2Axz^2 + Bz^3) \\
&\quad + 9Ax^2(3x^2 + AAz^2), \\
\psi_z(x, y, z) &= 8y^3z.
\end{aligned}$$

This is another representation of the rational map ϕ and we can use this representation of ϕ to evaluate

$$\phi(\mathcal{O}) = [\psi_x(0, 1, 0) : \psi_y(0, 1, 0) : \psi_z(0 : 1 : 0)] = [0 : 1 : 0] = \mathcal{O},$$

as expected.

Having seen how messy things can get even with relatively simple isogeny $P \mapsto 2P$, in the future we will be happy to omit such verifications and rely on the fact that, if we have a rational map that we know represents an isogeny ϕ, then $\phi(\mathcal{O}) = \mathcal{O}$ must hold. For elliptic curves in Weierstrass form, this means we only have to worry about evaluating isogenies at affine points, which allows us to simplify the equations by fixing $z = 1$.

2.2.3 The Frobenius endomorphism

Let $\mathbb{F}_p$ be a finite field of prime order p. We recall that the *Frobenius automorphism*

$$\pi : \overline{\mathbb{F}_p} \to \overline{\mathbb{F}_p}$$

is the map $x \mapsto x^p$. It is easy to check that π is a field automorphism; indeed, $0^p = 0$, $1^p = 1$, $(-a)^p = -a^p$, $(a^{-1})^p = (a^p)^{-1}$, $(ab)^p = a^p b^p$ and $(a+b)^p = a^p + b^p$. Notice that, if $f(x_1, \ldots, x_k)$ is any rational function with coefficients in $\mathbb{F}_p$, then

$$f(x_1, \ldots, x_k)^p = f(x_1^p, \ldots, x_k^p)$$

since all the coefficients of f are fixed by π, which acts trivially on $\mathbb{F}_p$.

Notice now that every n-th power $\pi^n = \pi \circ \cdots \circ \pi$ is also an automorphism of $\overline{\mathbb{F}}_p$, and the fixed field of π^n, i.e.

$$\{\ \alpha \in \overline{\mathbb{F}}_p \mid \alpha^{p^n} = \alpha\ \}$$

is exactly the finite field $\mathbb{F}_{p^n}$ with p^n elements. For $q = p^n$ we call the map $x \mapsto x^q$ the *q-power* Frobenius map, and we denote it by π_q.

Definition 2.2.1. Let E be an elliptic curve defined over a finite field $\mathbb{F}_q$; the *Frobenius endomorphism* of E is the map

$$\begin{aligned} \pi_E : E &\longrightarrow E \\ [x : y : z] &\longmapsto [x^q : y^q : z^q]. \end{aligned}$$

First, let us prove that this defines a morphism from E to itself; assume that the curve E is defined by a Weierstrass equation of the form $f(x, y, z) = 0$, where $f(x, y, z) \in \mathbb{F}_q[x, y, z]$; then, for any point $P = [x_0 : y_0 : z_0] \in E(\overline{F}_q)$ we have $f(x_0, y_0, z_0) = 0$; now, if we rise both the members of the equation to $q = p^n$, we have

$$0 = (f(x_0, y_0, z_0))^q = f(x_0^q, y_0^q, z_0^q)$$

since π_q acts trivially on the coefficients of f; thus $[x_0^q : y_0^q : z_0^q] \in E(\overline{\mathbb{F}}_q)$. Notice that this is define by a rational function, hence it is an endomorphism. In this case, one can also prove directly that π_E is a group homomorphism; indeed, recall that the group law on E is defined by rational functions whose coefficients lie in $\mathbb{F}_q$; consequently, since these coefficients are invariant under the q-power map, we have that $\pi_E(P + Q) = \pi_E(P) + \pi_E(Q)$ for every $P, Q \in E(\overline{\mathbb{F}}_p)$.

2.3 A standard form for isogenies

To facilitate our work with isogenies, it is convenient to put them in a standard form; in order to do so, we will assume to work with elliptic curves in short Weierstrass form $y^2 = x^3 + Ax + B$, i.e. being on a field of characteristic $\neq 2, 3$.[4] Let us prove the following important lemma.

Lemma 2.3.1. *Let E_1 and E_2 be elliptic curves defined over K in short Weierstrass form and let $\alpha : E_1 \to E_2$ an isogeny. Then, α can be defined by an affine rational map of the form*

$$\alpha(x, y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)} y \right),$$

where $u, v, s, t \in K[x]$ and $(u, v) = (s, t) = 1$.

Example 2.3.2. We saw that the multiplication-by-2 map can be expressed on the affine plane as

$$(x, y) \mapsto \left(\frac{(3x^2 + A)^2 - 8xy^2}{4y^2}, \frac{12xy^2(3x^2 + A) - (3x^2 + A)^3 - 8y^4}{8y^3} \right);$$

notice that the first coordinate depends only on y^2, hence using the relation $y^2 = x^3 + Ax + B$ we get

$$\frac{(3x^2 + A)^2 - 8xy^2}{4y^2} = \frac{(3x^2 + A)^2 - 8x(x^3 + Ax + B)}{4(x^3 + Ax + B)} = \frac{x^4 - 2Ax^2 - 8Bx + A^2}{4(x^3 + Ax + B)};$$

for the second coordinate, if we multiply numerator and denominator by y and use again that $y^2 = x^3 + Ax + B$, we get

$$\begin{aligned} &\frac{12xy^2(3x^2 + A) - (3x^2 + A)^3 - 8y^4}{8y^3} = \frac{12xy^2(3x^2 + A) - (3x^2 + A)^3 - 8y^4}{8y^4}\, y \\ &= \frac{12x(x^3 + Ax + B)(3x^2 + A) - (3x^2 + A)^3 - 8(x^3 + Ax + B)^2}{8(x^3 + Ax + B)^2} y \\ &= \frac{x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - A^3 - 8B^2}{8(x^3 + Ax + B)^2} y, \end{aligned}$$

giving the standard form.

We will use the following lemma.

Lemma 2.3.3. *Let E be an elliptic curve over a field K in short Weierstrass form*

$$E : y^2 = x^3 + Ax + B$$

and let $f \in K(E)$ be a rational function on E. We say that f is even if $f(P) = f(-P)$ for every $P \in E(\overline{K})$. Then f is even if and only if $f \in \overline{K}(x)$.

[4] It is easy to see that all the arguments of these section apply readily to curves of the form $y^2 = f(x)$, so we need only to assume that $\mathrm{char}(K) \neq 2$.

Proof. We know by the property of the group operation that if $P = (x_0, y_0)$ then $-P = (x_0, y_0)$ so that every element of $\overline{K}(x)$ is even. Let us now suppose that $f \in \overline{K}(E)$ is even. Using the Weierstrass equation we can write f as

$$f(x, y) = g(x) + h(x)y \qquad \text{with } g, h \in \overline{K}(x).$$

Then the condition $f(P) = f(-P)$ implies that

$$\begin{aligned} f(x, y) &= f(x, -y) \\ g(x) + h(x)y &= g(x) - h(x)y \\ 2yh(x) &= 0 \end{aligned}$$

for every x, y. Hence if $\operatorname{char} K \neq 2$ this holds only if h is the zero polynomial. □

Proof. Assume that α is defined by the rational map $[\alpha_x : \alpha_y : \alpha_z]$; then, for any affine point $[x : y : 1] \in E_1(\overline{K})$ we can write

$$\alpha(x, y) = (r_1(x, y), r_2(x, y)),$$

where $r_1(x, y) = \frac{\alpha_x(x,y,1)}{\alpha_z(x,y,z)}$ and $r_2(x, y) = \frac{\alpha_y(x,y,1)}{\alpha_z(x,y,z)}$. By repeatedly using the equation $y^2 = x^3 + Ax + B$ to replace y^2 with a polynomial in x, we can assume that both r_1 and r_2 have degree at most 1 in y. We then have

$$r_1(x, y) = \frac{p_1(x) + p_2(x)y}{p_3(x) + p_4(x)y},$$

for some $p_1, p_2, p_3, p_4 \in K[x]$. We now multiply the numerator and denominator of $r_1(x, y)$ by $p_3(x) - p_4(x)y$ and use the curve equation to get rid of y^2 in the denominator, and we have

$$r_1(x, y) = \frac{q_1(x) + q_2(x)y}{q_3(x)}$$

for some $q_1, q_2, q_3 \in K[x]$. We now use the property that α is a group homomorphism, hence $\alpha(-P) = -\alpha(P)$ for any $P \in E(\overline{K})$. Recall that that the inverse of an affine point $P = (x, y)$ on a curve in short Weierstrass form is $(x, -y)$; thus, if $-\alpha(x, y) = \alpha(x, -y)$, we have

$$(r_1(x, -y), r_2(x, -y)) = (r_1(x, y), -r_2(x, y)).$$

Consequently, $r_1(x, y)$ is a **even** rational function in $K(E)$. By Lemma 2.3.3 we conclude that $q_2 = 0$. After eliminating any common factor between q_1 and q_3 we get $r_1(x, y) = \frac{u(x)}{v(x)}$ with $(u, v) = 1$. The same argument for $r_2(x, y)$, where we use now $r_2(x, -y) = -r_2(x, y)$ to show that $q_1 = 0$ gives the rest of the proof. □

We will refer to the expression $\alpha(x, y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ as the *standard form* of an isogeny $\alpha : E_1 \to E_2$. Notice that the fact that the rational functions are in lowest terms implies tha the polynomials u, v, s and t are uniquely determined up to a scalar in $K^\times$.

Lemma 2.3.4. *Let $E_1 : y^2 = f_1(x)$ and $E_2 : y^2 = f_2(x)$ be elliptic curves over K and let $\alpha(x, y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny from E_1 to E_2 written in standard form. Then, v^3 divides t^2 and t^2 divides $v^3 f_1$. Moreover, $v(x)$ and $t(x)$ have the same set of roots in $\overline{K}$.*

Remark 2.3.5. This is the case in Example 2.3.2; indeed in this case the curve is defined by the equation $y^2 = x^3 + Ax + B$, $v(x) = 4(x^3 + Ax + B)$ and $t(x) = 8(x^3 + Ax + B)^2$; hence we have that, up to scalars that $v^3 \mid t^2$, $t^2 \mid v^3 f$ and t and v have the same set of roots.

Proof. First, the image of α must lie in E_2; hence, substituting $(\frac{u}{v}, \frac{s}{t}y)$ in the equation for E we have

$$\left(\frac{s}{t}y\right)^2 = f_2\left(\frac{u}{v}\right),$$

and using the equation for E_1 to express y^2 in terms of x gives

$$\left(\frac{s}{t}\right)^2 f_1(x) = \left(\frac{u}{v}\right)^3 + A_2\left(\frac{u}{v}\right) + B_2$$

as an identity involving polynomials $f_1, s, t, u, v \in K[x]$. Let us denote by $w = u^3 + A_2uv^2 + B_2v^3$; hence, clearing denominators we have

$$v^3 s^2 f_1 = t^2 w.$$

Now, the fact that $(u, v) = 1$ implies that $(v, w) = 1$; consequently, we have that $v^3 \mid t^2$; on the other hand, using the fact that $(s, t) = 1$, we get that $t^2 \mid v^3 f_1$. This also implies directly that t and v have the same roots in $\overline{K}$; indeed, every root of v is a root of t since $v^3 \mid t^2$. On the other hand, every root of t is a double root of t^2; but $t^2 \mid v^3 f_1$, and f_1 does not have any double root (because the curve $y^2 = f_1(x)$ is non-singular), hence x_0 is a root of v, concluding the proof. □

As any isogeny is a group homomorphism, it makes sense to speak about the kernel of the isogeny. If $\phi : E_1 \to E_2$, then

$$\ker(\phi) = \{P \in E_1(\overline{K}) \mid \phi(P) = \mathcal{O}_2\}.$$

Notice that $\mathcal{O}_1 \in \ker(\phi)$ for every isogeny; the following result allows to compute the other elements of the kernel starting from the standard form of the affine part of ϕ.

Corollary 2.3.6. *Let $\alpha(x, y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny $E_1 \to E_2$ in standard form; then, the affine points $[x_0 : y_0 : 1] \in E_1(\overline{K})$ in the kernel of α are exactly those for which $v(x_0) = 0$.*

Proof. If $v(x_0) \neq 0$, then $t(x_0) \neq 0$ and $\alpha(x_0, y_0) = \left(\frac{u(x_0)}{v(x_0)}, \frac{s(x_0)}{t(x_0)}y_0\right)$ is an affine point and therefore not $\mathcal{O}$, hence $(x_0, y_0) \notin \ker(\alpha)$. By homogenizing and putting α into projective form, we can write α as

$$\alpha = [ut : vsy : vt],$$

where ut, vst and vt are now homogenous polynomials of equal degree ($s, t, u, v \in K[x, z]$). Suppose $y_0 \neq 0$; by the previous lemma, if $v(x_0, 1) = 0$, then $t(x_0, 1) = 0$ and since $v^3 \mid t^2$, the multiplicity of $(x_0, 1)$ as a root of t is strictly greater than its multiplicity as a root of v. This implies that, working over $\overline{K}$, we can renormalize α by dividing by a suitable power of $x - x_0 z$ so that α_y does not vanish at $[x_0 : y_0 : 1]$ but α_x, α_z both do. Then, $\alpha(x_0 : y_0 : 1) = (0 : 1 : 0) = \mathcal{O}$, hence $(x_0 : y_0 : 1) \in \ker(\alpha)$ as claimed. If $y_0 = 0$, then x_0 is a root of the cubic polynomial f_1 appearing in the equation of $E_1 : y^2 = f_1(x)$, and it is not a double root since E_1 is non-singular. In this case, we normalize $\alpha = [ut : vsy : vt]$ by multiplying by yz and then replacing $y^2 z$ by $f_1(x, z)$ to obtain

$$\alpha = [utyz : vsf_1 : vtyz].$$

Notice now that $(x_0, 1)$ is a root of multiplicity 1 of $f_1(x, z)$; hence, using the fact that $v^3 \mid t^2$, we have that the multiplicity of $(x_0, 1)$ as a root of vf_1 is not greater than the multiplicity as a root of t; hence, we can again normalize α by dividing by a suitable power of $x - x_0 z$ so that α_y does not vanish at $(x_0 : 0 : 1)$ but α_x and α_z do. Thus, $[x_0 : 0 : 1]$ is again in the kernel of α, concluding the theorem. □

This corollary implies that, when we have an isogeny $\phi : E_1 \to E_2$ in standard form, we know exactly what to do whenever we get a zero in the denominator when we try to compute $\alpha(P)$: these are exactly the case in which $\alpha(P) = \mathcal{O}$. Consequently, we have that

$$\ker(\alpha) = \{[x_0 : y_0 : 1] \in E_1(\overline{K}) \mid v(x_0) = 0\} \cup \{[0 : 1 : 0]\}.$$

We obtain also the following corollary.

Corollary 2.3.7. *Let $\alpha : E_1 \to E_2$ be an isogeny of elliptic curves defined over a field K; then, the kernel of α is a finite subgroup of $E_1(\overline{K})$.*

This fact is true in general, but we prove it assuming that we can put α in standard form (hence $\mathrm{char}(K) \neq 2$).

Proof. Let us assume that α is in standard form $(u/v, sy/t)$; then, the polynomial v has at most $\deg v$ distinct roots in $\overline{K}$, each of them appearing as the x-coodinate of at most 2 points on the elliptic curve E_1. □

We conclude this section with an exercise.

Exercise 2.3.8. Without using Theorem 1.2.24, prove that, if $\alpha : E_1 \to E_2$ is an isogeny, then it is a surjective map from $E_1(\overline{K})$ to $E_2(\overline{K})$.

2.4 Degree and separability

We now define two important invariants of an isogeny that can be easily determined when it is in standard form.

Definition 2.4.1. Let $\alpha(x,y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be an isogeny written in standard form. The *degree* of α is $\deg\alpha = \max\{\deg u, \deg v\}$. Moreover, we say that α is separable if the derivative of u/v is nonzero; otherwise we say that α is inseparable.

Notice that, since the polynomials u, v, s, t are uniquely determined up to a scalar factor, the degree and the separability of α are intrinsic properties that do not depend on its representation as a rational map.

Remark 2.4.2. The degree and separability of an isogeny can be defined equivalently in a more intrinsic way in terms of the function fields. Indeed, if $\alpha : E_1 \to E_2$ is an isogeny of function fields defined over K, then it induces an injective map

$$\begin{aligned} \alpha^* : K(E_2) &\to K(E_1) \\ f &\mapsto f \circ \alpha. \end{aligned}$$

In this way, $\alpha^*(K(E_2))$ is a subfield of $K(E_1)$. The degree of α is then the degree of the extension $K(E_1)/\alpha^*(K(E_2))$. Notice that this is a finite extension, because both the extensions are finite extensions of a purely transcendental extension of K (for a proof of this, see [Sil09, Theorem 2.4]). Moreover, the isogeny is said to be *separable* if this field extension is seperable[5] (and α is said to be inseparable otherwise). This definition is equivalent to ours, but we will not prove it here. Let us notice that this definition has the virtue of being more general, but it is not easy to apply it explicitely.

Let us finally return to the three examples we saw earlier.

- The standard form of the negation map is $\alpha(x,y) = (x,-y)$; this is separable and it has degree 1;
- The standard form of the multiplication-by-2 isogeny is

  $$\alpha(x,y) = \left(\frac{x^4 - 2Ax^2 - 8Bx + A^2}{4(x^3+Ax+B)}, \frac{x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - A^3 - 8B^2}{8(x^3+Ax+B)^2}y\right);$$

  it is separable and it has degree 4;
- The Frobenius endomorphism is given by $\pi_E : E \to E$

  $$[x:y:z] \mapsto [x^q : y^q : z^q];$$

  hence. on the affine part of E, we have

  $$\pi_E(x,y) = (x^q, y^q).$$

  Recall that q is odd (since we are in characteristic $\neq 2$), hence $q-1$ is even; we can use the Weierstrass equation to transform y^q and we get

  $$\pi_E(x,y) = (x^q, (x^3+Ax+B)^{(q-1)/2}y).$$

  Notice that the degree is q and the map is inseparable, since $(x^q)' = qx^{q-1} = 0$ in $\mathbb{F}_q$.

[5] Recall that an algebraic extension L/F is said to be separable if, for every element $\beta \in L$, its minimal polynomial $\mu_\beta \in K[x]$ has no repeated roots in any extension field.

2.5 Isogeny kernels

Given an isogeny $\alpha : E_1 \to E_2$ between two elliptic curves, its kernel plays an important role in the study of the isogeny. Moreover, many important subgroups can be seen as the kernel of an isogeny. we start with an example.

Example 2.5.1. Let n be an integer ≥ 1 and consider the multiplication-by-n map $\alpha : E \to E$ given by $P \mapsto nP := P + \cdots + P$ n times; this is an isogeny, because it is a group homomorphism defined by a non-constant rational map. The kernel is exactly the n-torsion subgroup

$$E[n] = \{P \in E(\overline{K}) \mid nP = \mathcal{O}\}.$$

Torsion subgroups play an important role in the theory of elliptic curves; in particular, when $K = \mathbb{F}_q$, the finte abelian group $E(\mathbb{F}_q)$ is completely determined by its intersections with the n-torsion subgroups $E[n]$ (in fact, its intersections with $E[\ell^e]$ for the prime powers that divide $\#E(\mathbb{F}_q)$). Understanding the structure of $E[n]$ will be the key ingredient to understand the structure of $E(\mathbb{F}_q)$.

In the first part of this section we want to prove a result on the cardinality of the kernel of an isogeny; more specifically, we are going to prove that, for separable isogenies, the cardinality of $\ker(\alpha)$ is equal to the degree of α. Before doing this, we will show that every isogeny can be decomposed into the composition of a separable isogeny and a suitable power of the Frobenius morphism (which has trivial kernel). First, we prove the following lemma.

Lemma 2.5.2. *Let u, v be relatively prime polynomials in $K[x]$; then,*

$$\left(\frac{u}{v}\right)' = 0 \quad \iff \quad u' = v' = 0 \quad \iff \quad u = f(x^p) \text{ and } v = g(x^p),$$

where f and g are polynomials in $K[x]$ and p is the characteristic of K (which may be 0).

Proof. Assume first that $\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2} = 0$; then, $u'v = uv'$. Recall that the polynomials u and v have no common roots in $\overline{K}$, therefore evry root of u' must be also a root of v and viceversa; but $\deg u' < \deg u$, so this is possible only if $u' = 0$, and by the same argument we must also have that $v' = 0$. This proves the first equivalence.

Now, let $u(x) = \sum a_n x^n$; if $u'(x) = \sum n a_n x^{n-1} = 0$, then $n a_n = 0$ for every n, which means that n must be a multiple of p over every nonzero a_n; hence, we can write

$$u(x) = \sum_m a_{p^m} (x^p)^m = f(x^p),$$

where $f = \sum_m a_m x^m$. Similarly, $v(x) = g(x^p)$ for some $g \in K[x]$. The converse is trivial. □

As an important corollary we have the following result.

Corollary 2.5.3. *If K is a field of characteristic 0, every isogeny defined over K is separable.*

We know show that, if K is a field of characteristic p, every inseparable isogeny arises as the composition of a separable isogeny and some p-power Frobenius map $\pi : [x : y : z] \to [x^p : y^p : z^p]$.

Proposition 2.5.4. *Let $\alpha : E_1 \to E_2$ be an inseparable isogeny of elliptic curves*

$$E_1 : \ y^2 = x^3 + A_1x + B_1, \quad E_2 : \ y^2 = x^3 + A_2x + B_2$$

defined over a field K of characteristic $p > 2$; then, α can be written in the form

$$\alpha(x, y) = (r_1(x^p), r_2(x^p)y^p)$$

for some rational functions $r_1, r_2 \in K(x)$.

Proof. Let $\alpha(x, y) = \left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be in standard form; since by assumption α is inseparable, by Lemma 2.5.2 it follows that $\frac{u(x)}{v(x)} = r_1(x^p)$ for some rational function $r_1 \in K(x)$; we are left then to show that $\frac{s(x)}{t(x)}y$ can be put in the form $r_2(x^p)y^p$. As in the proof of Lemma 2.3.4, we can substitute $\left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ into the equation of E_2 and then use the equation defining E_1 to eliminate y^2; doing this we obtain the equality

$$v^3s^2f = t^2w,$$

where $f(x) = x^3 + A_1x + B_1$ and $w(x) = u^3 + A_2uv^2 + B_2v^3$. We can rewrite this equality as

$$\frac{w}{v^3} = \frac{s^2f}{t^2}.$$

Since α is inseparable, we have $u' = v' = 0$, hence also $w' = 0$ and consequently $\left(\frac{w}{v^3}\right)' = \left(\frac{s^2f}{t^2}\right)' = 0$. This implies that $s(x)^2f(x) = g(x^p)$ and $t(x)^2 = h(x^p)$ for some polynomials g, h. Since $(t(x)^2)'$ is $2t(x)t(x)'$ and it is equal to zero, it follows that $t(x)' = 0$ (recall that $\operatorname{char} K \neq 2$), hence $t(x) = h_1(x^p)$ for some polynomial h_1.

Notice that every root of $g(x^p)$ has multiplicity equal to a multiple of p, hence since f has all distict roots and p is odd, we can write $s(x)^2f(x) = g(x^p) = s_1(x)^2f(x)^p$, where $s_1(x) = g_1(x^p)$ for some polynomial g_1. This follows from the fact that $0 = g(x^p)' = 2s_1(x)s_1(x)'f(x) + ps_1^2(x)f_1(x)^{p-1}f_1(x)' = 2s_1(x)s_1(x)'f_1(x)$ implies that $s_1(x)' = 0$ and we can apply again Lemma 2.5.2.

We then have

$$(s(x)y)^2 = s(x)^2f(x) = g_1(x^p)^2f(x)^p = g_1(x^p)^2y^{2p},$$

wehre we used that $y^2 = f(x)$ on E_1. Thus we have

$$\left(\frac{s(x)}{t(x)}y\right)^2 = \left(\frac{g_1(x^p)}{h_1(x^p)}y^p\right)^2 = (r(x^p)y^p)^2,$$

with $r(x) = g_1(x)/h_1(x)$. It follows that $\frac{s(x)}{t(x)}y = r_2(x^p)y^p$ with $r_2 = \pm r$, since two rational functions that agree up to a sign at infinitely many points can only differ in sign. □

Corollary 2.5.5. *Let α be an isogeny of elliptic curves over a field K of characteristic $p > 0$; then,*

$$\alpha = \alpha_{\text{sep}} \circ \pi^n,$$

where α_{sep} is a separable isogeny, $n \geq 0$ is an integer and π is the p-power endomorhpism $\pi[x : y : z] = [x^p : y^p : z^p]$. In particular, we have $\deg(\alpha) = p^n \deg \alpha_{\text{sep}}$.

Proof. This property holds for any characteristic, but we will prove it assuming $p > 3$ for the sake of simplicity. If α is separable, then $\alpha = \alpha_{\text{sep}}$ and $n = 0$. Assume that α is inseparable; then we can apply the previous proposition and we may write $\alpha(x, y) = (r_1(x^p), r_2(x^p)y^p)$ with $r_1, r_2 \in K(x)$. We than have $\alpha = \alpha_1 \circ \pi$, where $\alpha_1(x, y) = (r_1(x), r_2(x)y)$. We now have to choices: either α_1 is separable, hence we are done, or α_1 is inseparable, and we can apply the same argument to α_1. We can hence apply this argument recursively and obtain $\alpha = \alpha_n \circ \pi^n$, where α_n is a separable isogeny (notice that this procedure has to stop because the degree of α is finite and every step reduce the degree by a factor of p). Moreover, $\deg(\alpha) = \deg(\alpha_{\text{sep}} \circ \pi^n) = \deg(\alpha_{\text{sep}}) \deg(\pi)^n = p^n \deg(\alpha_{\text{sep}})$, concluding the proof. □

Remark 2.5.6. Notice that the isogeny α_{sep} does not necessarily have the same domain as $\alpha : E_1 \to E_2$, since in principle the image of π^n is not necessarily E_1 (it is whenever E_1 is defined over $\mathbb{F}_{p^n}$). Notice also that, if K is a field of characteristic zero or a finite field, then we can also decompose α as $\alpha = \pi^n \circ \tilde{\alpha}_{\text{sep}}$, where $\tilde{\alpha}_{\text{sep}}$ and α_{sep} have the same degree.

If we write $\alpha = \alpha_{\text{sep}} \circ \pi^n$, the degree of α_{sep} is called the *separable degree* of α (denoted by $\deg_s(\alpha)$), and p^n is called the *inseparable degree* of α (denoted by $\deg_i(\alpha)$). It follows from the previous corollary that the degree of α is always the product of its separable and inseparable degrees, i.e.

$$\deg(\alpha) = \deg_s(\alpha) \deg_i(\alpha).$$

Notice that the isogeny π^n has separable degree 1; such isogenies are said to be *purely inseparable*. The degree of a purely inseparable isogeny is always a power of p, but the converse is not true as we shall see later in the course.

Notice that $\ker(\pi) = \{[0 : 1 : 0]\}$; indeed, we can have $[x^p : y^p : z^p] = [0 : 1 : 0]$ if and only if $[x : y : z] = [0 : 1 : 0]$. We can now prove this very important result.

Theorem 2.5.7. *The cardinality of the kernel of an isogeny is equal to its separable degree.*

Proof. Let $\alpha = \alpha_{\text{sep}} \circ \pi^n$; then, $\#\ker(\alpha) = \#\ker(\alpha_{\text{sep}})$ since the kernel of π (and hence of π^n) is trivial. It is thus sufficient to prove the result for separable isogenies. Assume now that α is separable and let $\left(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y\right)$ be its standard form. By definition,

$\ker(\alpha) = \alpha^{-1}([0:1:0])$, and since α is a group homomorphism, the cardinality of the preimage of a point is always the same (if this set is not empty), hence it will be sufficient to compute $\#\alpha^{-1}(a,b)$ for any $(a,b) \in \alpha(E_1(\overline{K}))$. Let us choose point $(a,b) \in \alpha(E_1(\overline{K}))$ with $a, b \neq 0$ and such that a is not equal to the ratio of the leading coefficients of u and v.[6] Call the set $S(a,b) := \alpha^{-1}(a,b)$. We require moreover that, for every $(x_0, y_0) \in S(a,b)$, $(u/v)'(x_0) \neq 0$, which is again possible since α is a separable isogeny, hence $(u/v)'$ is not identically zero, hence it has only finitely many roots. Let us now compute the cardinality of $S(a,b)$. If $(x_0, y_0) \in S(a,b)$, then

$$\frac{u(x_0)}{v(x_0)} = a, \quad \frac{s(x_0)}{t(x_0)} y_0 = b.$$

We must have $t(x_0) \neq 0$, since α is defined at (x_0, y_0), and $b \neq 0$ implies $s(x_0) \neq 0$. It follows that $y_0 = \frac{t(x_0)}{s(x_0)} b$ is uniquely determined by x_0. Thus, to compute $\#S(a,b)$ it suffices to count the number of distinct values of x_0 that occour among the points in $S(a,b)$. We now let $g(x) = u(x) - av(x)$, so that $\alpha(x_0, y_0) = (a,b)$ if and only $g(x_0) = 0$. Since a is not equal to the ratio of the leading coefficients of u and v, then $\deg(g) = \max\{\deg(u), \deg(v)\} = \deg(\alpha)$, hence the cardinality of $S(a,b)$ is equal to the number of *distinct* roots of g. We are left to prove that g has all distinct roots.[7] Any $x_0 \in \overline{K}$ is a multiple root of g if and only if $g(x_0) = g'(x_0) = 0$, equivalently, $av(x_0) = u(x_0)$ and $av'(x_0) = u'(x_0)$. If we multiply opposite sides of these equations and cancels the a's we get $u'(x_0)v(x_0) = u(x_0)v'(x_0)$. But by construction for every $(x_0, y_0) \in S(a,b)$ we have that $(u/v)'(x_0) \neq 0$, hence $u'(x_0)v(x_0) - u(x_0)v'(x_0) \neq 0$, which implies that g has no multiple roots, concluding the proof since

$$\#\ker(\alpha) = \#S(a,b) = \deg(g) = \deg(\alpha).$$

□

We now state two more corollaries.

Corollary 2.5.8. *Every purely inseparable isogeny has trivial kernel.*

Corollary 2.5.9. *For any composition of isogenies $\alpha = \beta \circ \gamma$ we have*

$$\deg(\alpha) = \deg(\beta)\deg(\gamma), \quad \deg_s(\alpha) = \deg_s(\beta)\deg_s(\gamma), \quad \deg_i(\alpha) = \deg_i(\beta)\deg_i(\gamma).$$

Proof. Since the degree of an isogeny is the product of its separable degree and its inseparable degree, it sufficies to prove the last two equalities. The fact that γ is a surjective group homomorphism implies that $\ker(\alpha) = \gamma^{-1}(\ker(\beta))$, and this is exactly the union of $\#\ker(\gamma)$ cosets of $\ker(\beta)$[8]; this implies that $\#\ker(\alpha) = \#\ker(\beta)\#\ker(\gamma)$, hence by Theorem 2.5.7 we have that $\deg_s(\alpha) = \deg_s(\beta)\deg_s(\gamma)$ as wanted. To prove the second part, let us use Corollary 2.5.5 to write $\alpha = \beta \circ \gamma$ as

$$\alpha_{\text{sep}} \circ \pi^a = (\beta_{\text{sep}} \circ \pi^b) \circ (\gamma_{\text{sep}} \circ \pi^c).$$

[6] Such a point exists because $\alpha(E_1(\overline{K}))$ is infinite.
[7] This is trivial if $\text{char}K = 0$.
[8] Notice that every element $a \in \gamma^{-1}(\ker(\beta))$ is of the form $b+c$ with $b \in \ker(\beta)$ and $c \in \gamma^{-1}(\mathcal{O}) = \ker(\gamma)$.

Let us now call $\delta := \pi^b \circ \gamma_{\text{sep}}$; since π has trivial kernel, this isogeny has the same kernel as γ_{sep}, hence the same separable degree of γ and inseparable degree equal to β, hence we can apply again Corollary 2.5.5 to write $\delta = \delta_{\text{sep}} \circ \pi^b$. We then have

$$\alpha_{\text{sep}} \circ \pi^a = \beta_{\text{sep}} \circ \delta_{\text{sep}} \circ \pi^{bc},$$

and so

$$\deg_i(\alpha) = p^a = \deg_i\left(\beta_{\text{sep}} \circ \delta_{\text{sep}} \circ \pi^{bc}\right) = p^{bc} = \deg_i(\beta)\deg_i(\gamma),$$

since the composition of two separable isogenies is still separable.[9] This conclude the proof of the corollary. □

2.6 Isogenies from kernels

We have seen in the previous section that, for every isogeny $\alpha : E \to E'$, the kernel of α is a finite subgroup of $E(\overline{K})$. It is then natural to ask whether the converse is also true, namely if, given a finite subgroup G of $E(\overline{K})$, there exists an isogeny from α to some elliptic curve E' having G as kernel? The answer is yes; moreover, when consider separable isogenies (and we should, since every isogeny can be written as the composition of a separable isogeny and a suitable power of π and π has trivial kernel), the isogeny α and the elliptic curve E' are uniquely determined up to isogeny. More precisely, the following theorem holds.

Theorem 2.6.1. *Let E/K be an elliptic curve and let G be a finite subgroup of $E(\overline{K})$; there exists an elliptic curve E' ans a separable isogeny $\phi : E \to E'$ with $\ker(\phi) = G$. The curve E' and the isogeny ϕ are defined over a finite extension of K and are unique up to isomorphism.*

There is more than a way to prove this theorem; we will give an idea of how to prove it from a geometric point of view; later, we will give explicit formulas for constructing α and E' from G due to Vélu [Vél71].

To do this, we need some important properties coming from geometry. Given $\mathcal{C}_1/K$ and $\mathcal{C}_2/K$ two agebraic, let $\phi : \mathcal{C}_1 \to \mathcal{C}_2$, let ϕ be a nonconstant rational function defined over K; then, the composition via ϕ induces an injection of function fields

$$\phi^* : K(\mathcal{C}_2) \to K(\mathcal{C}_1), \quad \phi^*(f) = f \circ \phi,$$

and this map in an inclusion of fields. This has the following important properties.

Proposition 2.6.2.

(a) Let $\phi : \mathcal{C}_1 \to \mathcal{C}_2$ be a nonconstant map defined over K; then $K(\mathcal{C}_1)$ is a finite extension of $\phi^* K(\mathcal{C}_2)$;

(b) let $\iota : K(\mathcal{C}_2) \to K(\mathcal{C}_1)$ be an injection of function fields fixing K; then, there exists a unique nonconstant map $\phi : \mathcal{C}_1 \to \mathcal{C}_2$ (defined over K) such that $\phi^* = \iota$;

[9] You can prove this as an exercise.

(c) Let $\mathbb{K} \subset K(\mathcal{C}_1)$ be a subfield of finite index containing K; then, there exist a smooth curve $\mathcal{C}'/K$, unique up to K-isomorphism, and a nonconstant map $\phi : \mathcal{C} \to \mathcal{C}'$ defined over K such that $\phi^*(K(\mathcal{C}')) = \mathbb{K}$.

For a proof of this, see [Sil09, Theorem 2.4]. The main idea is that a finitely generated extension $\mathbb{K}/K$ is generated by a finite set S of elements of $\mathbb{K}$ (both algebraic and trascendental over K). One gets a morphism of K algebras

$$K[x_1, \dots, x_n] \to K[S]$$

whose kernel is an ideal generated by a finite number of polynomials. These polynomials are the equations of an algebraic variety X whose function field is $\mathbb{K}$. Since $\mathbb{K}$ has trascendence degree 1, such X is a curve, and since $\mathbb{K} \cap \overline{K} = K$ the coefficients of the polynomials are in K.

The above result shows essentially the close connection between (smooth) curves and their function fields. This can be made precise by stating that the following map is an equivalence of categories:

$$\begin{bmatrix} \textit{Objects} : \text{smooth curves} \\ \text{defined over } K \\ \textit{Maps} : \text{nonconstant rational} \\ \text{maps}^{10} \text{ defined over } K \end{bmatrix} \rightsquigarrow \begin{bmatrix} \textit{Objects} : \text{finitely generated} \\ \text{extensions } \mathbb{K}/K \text{ of transcendence} \\ \text{degree one with } \mathbb{K} \cap \overline{K} = K \\ \textit{Maps} : \text{fields injections fixing } K \end{bmatrix}$$

$$\mathcal{C}/K \quad \rightsquigarrow \quad K(\mathcal{C})$$

$$\phi : \mathcal{C}_1 \to \mathcal{C}_2 \rightsquigarrow \quad \phi^* : K(\mathcal{C}_2) \to K(\mathcal{C}_1)$$

Next, we will need to consider the *translation map* τ_T over the elliptic curve. Fix a point $T \in E(\overline{K})$ and consider the map $\tau_T : E \to E$ sending $P \mapsto P + T$. This is a nonconstant rational map, but it is not an isogeny if $T \neq \mathcal{O}$. However, this map induces in general an automorphism of $\overline{K}(E)$ given by $\tau_T^* : \overline{K}(E) \to \overline{K}(E)$ sending $f \mapsto f \circ \tau_T$.

Assume now that we have an isogeny $\phi : E_1 \to E_2$ between two elliptic curves and, for every $T \in \ker \phi$, let us consider the map $\tau_T^* : \overline{K}(E_1) \to \overline{K}(E_1)$. We have the following result.

Proposition 2.6.3. *We have that ${\tau_T^*}_{|_{\phi^*(\overline{K}(E_2))}} = id$, i.e., for every $g \in \phi^*(\overline{K}(E_2))$, $\tau_T^*(g) = g$.*

Proof. Assume that $T \in \ker \phi$ and let $\tau_T : E_1 \to E_1$ the translation-by-T map; we want to show that, for every $g \in \phi^*(\overline{K}(E_2))$, $\tau_T^*(g) = g$. Now, by definition, if $g \in \phi^*(\overline{K}(E_2))$, then there exists $f \in \overline{K}(E_2)$ such that $g = f \circ \phi$. Hence we have that

$$\begin{aligned} \tau_T^*(g)(P) &= \tau_T^*(f \circ \phi)(P) = f(\phi(P * T)) \\ &= f(\phi(P) + \phi(T)) = f(\phi(P) + \mathcal{O}) = f(\phi)(P) = g(P), \end{aligned}$$

proving the claim. □

The previous proposition gives as a result that, given an isogeny $\phi : E_1 \to E_2$, then for every $T \in \ker \phi$, the map $\tau_T{}^* \in \mathrm{Aut}\left(\overline{K}(E_1)/\phi^*(\overline{K}(E_2))\right)$. Let us hence consider the map

$$\Phi : \ker \phi \to \mathrm{Aut}\left(\overline{K}(E_1)/\phi^*(\overline{K}(E_2))\right), \quad T \mapsto \tau_T^*.$$

We have the following.

Theorem 2.6.4. $\ker \phi \cong \mathrm{Aut}\left(\overline{K}(E_1)/\phi^*(\overline{K}(E_2))\right)$.

Proof. For simplicity let us prove the theorem for ϕ separable. The map is clearly a group homomorphism. Moreover, let us notice that it is enough to prove that Φ is injective; indeed, we have that $\#\,\mathrm{Aut}\left(\overline{K}(E_1)/\phi^*(\overline{K}(E_2))\right) \leq \deg \phi$, and by Theorem 2.5.7, $\#\ker \phi = \deg \phi$. Let us now prove that Φ is injective, i.e. that, if $\tau_{T_1}^* = \tau_{T_2}^*$, then $T_1 = T_2$. Notice that it is enough to prove that, if $\tau_T^* = \tau_{\mathcal{O}}^* = \mathrm{id}$, then $T = \mathcal{O}$. Indeed, assume that $\tau_T^* = \mathrm{id}$, i.e. $\forall f \in \overline{K}(E_2)$ and for all $P \in E_2(\overline{K})$, then $f(P+T) = f(P)$. But if we consider as f coordinate functions, this implies that $x(P+T) = x(P)$, $y(P+T) = y(P)$ and $z(P+T) = z(P)$ for every $P \in E_2(\overline{K})$, implying that $T = \mathcal{O}$, as wanted. □

Let us now finally prove Theorem 2.6.1.

Proof. Let $T \in E(\overline{K})$ be a point on our elliptic curve; we can consider the *translation-by-T* map $\tau_T : E(\overline{K}) \to E(\overline{K})$ sending $P \mapsto P + T$. Let us consider the induced map of the function field of E, i.e. $\tau_T^* : \overline{K}(E) \to \overline{K}(E)$ given by $f \mapsto f \circ \tau_T$, and this is an automorphism of $\overline{K}(E)$. Let us define

$$\overline{K}(E)^G := \{f \in \overline{K}(E) \mid \tau_T^*(f) = f \text{ for all } T \in G\}.$$

By applying Galois theory, we have that $\overline{K}(E)^G \subset \overline{K}(E)$ is a finite Galois extension with Galois group G.[11] Now, one can prove that, since the transcendence degree of $\overline{K}(E)^G$ over $\overline{K}$ is 1, there exists a unique (up to isomorhpisms) smooth curve $\mathcal{C}/\overline{K}$ and a finite morphism $\phi : E \to \mathcal{C}$ such that $\phi^*(\overline{K}(\mathcal{C})) = \overline{K}(E)^G$. To conclude, one has to show that

1. $\mathcal{C}$ is an elliptic curve, considered with neutral element $\phi(\mathcal{O})$;
2. $\ker \phi = G$ (this comes from the fact that $\ker \phi \cong \mathrm{Aut}(\overline{K}(E)/\phi^*(\overline{K}(\mathcal{C})))$);
3. $\phi : E \to \mathcal{C}$ is an isogeny, if we consider over $\mathcal{C}$ the group structure having $\phi(\mathcal{O})$ as neutral element.

To prove the first statement, we have to prove that the map ϕ is unramified, i.e. that $\forall Q \in \phi(E)$ then $\#\phi^{-1}(Q) = \deg \phi$. For the properties of finite morphisms of curves, it is always true that $\#\phi^{-1}(Q) \leq \deg \phi$, so let us prove $\geq$. To prove this, we will prove that, if $T \in G$, then $\phi(P+T) = \phi(P)$ for all $P \in E(\overline{K})$. Indeed, for every $f \in \overline{K}(\mathcal{C})$ one has that $f(\phi(P+T)) = (\tau_T^* \circ \phi^*)f(P)) = \phi^* f(P) = f(\phi(P))$, where we used the property that τ_T^* fixes every element of $\phi^*(\overline{K}(\mathcal{C}))$. This implies that $\phi(P+T) = \phi(P)$ for all P. Hence, if we choose a point $P \in E$ such that $P \in \phi^{-1}(Q)$, then $\phi^{-1}(Q) \subset \{P+T \mid T \in G\}$. Now,

[11] This comes from the fact that, if L is a field and G is a group acting faithfully (i.e. there is no element $g \in G \setminus \{\mathrm{id}\}$ such that $gx = x$ for all $x \in L$) on L, then K/K^G is a Galois extension with Galois group G.

the points in this set are all distinct, hence $\#\phi^{-1}(Q) \geq \#G = \deg\phi$. Knowing now that ϕ is unramified, an application of Riemann-Hurwitz formula (see [Sil09, Theorem 5.9]) gives you that if ϕ is unramified and E is an elliptic curve, then also $\mathcal{C}$ is an elliptic curve. The points 2 and 3 are then clear.

Notice that, if we have another separable isogeny $\phi' : E \to E'$ with the same kernel G, then we can view $\overline{K}(E')$ as a subfield of $\overline{K}(E)$ via the induced embedding $\phi'^* : \overline{K}(E') \to \overline{K}(E)$. Moreover $\phi'^*(\overline{K}(E')) \subset \overline{K}(E)^G$; indeed, for every $T \in G$ and every $g \in \overline{K}(E)$ $(\tau_T)^*(\phi'^*(g)) = (\tau_T)^*(g \circ \phi') = (g \circ \phi' \circ \tau_T)$ and $\phi' \circ \tau_T = \phi'$ if $T \in \ker(\phi')$. Since $\phi' : E \to E'$ is separable, we have $\deg(\phi') = [\overline{K}(E) : \phi'^*(\overline{K}(E'))] = \#G$, so $\phi'^*(\overline{K}(E'))$ must be isomorphic to $\overline{K}(E)^G$ (since they are contained one into the other). Consequently there exists an isomorphism $\iota : \mathcal{C} \to E'$ such that $\phi' = \iota \circ \phi$, hence the isogeny ϕ and the curve $\mathcal{C}$ are unique up to isomorphism. For a detailed proof of this result, see [Sil09, Proposition 4.12]. □

If G is a finite subgroup of $E(\overline{K})$ and $\phi : E \to E'$ is an isogeny with kernel G, we will denote the curve E' by E/G. As an application of this theorem, we can consider the following.

Exercise 2.6.5. An isogeny of composite degree can always be decomposed into a sequence of isogenies of prime degree.

Let $\alpha : E_1 \to E_2$ be an isogeny; if we are working in a field of characteristic $p > 0$, by writing α by $\alpha = \alpha_{\text{sep}} \circ \pi^n$, and we can decompose $\pi^n = \pi \circ \ldots \circ$ and each of them is an isogeny of prime degree p, hence it is sufficient to consider α separable. We procede by induction on the number of prime factors appearing in the decomposition of $\deg(\alpha)$. If $\deg(\alpha)$ is prime, there is nothing to prove, hence let us assume that $\deg(\alpha)$ is the product of (at least) two primes. Let us now consider $G = \ker(\alpha)$; since $\deg(\alpha) > 1$, we have that $G \neq \{e\}$, hence G contains a subgroup H of prime order. By Theorem 2.6.1, there exists a separable isogeny $\alpha_1 : E_1 \to E_3$ having H as kernel. Then, $\alpha_1(G)$ is a finite subgroup of $E_3(\overline{K})$ isomorphic to G/H; applying Theorem 2.6.1 again, there exists a separable isogeny $\alpha_2 : E_3 \to E_4$ having $\alpha_1(G)$ as its kernel. The kernel of the composition $\alpha_2 \circ \alpha_1$ is $G = \ker(\alpha)$ (indeed $\ker(\alpha_2 \circ \alpha_1) = \{e \in E_1 \mid \alpha_1(e) \in \ker(\alpha_2)\} = G$), so there exists an isomorphism $\iota : E_4 \to E_2$ such that $\alpha = \iota \circ \alpha_2 \circ \alpha_1$. We can then apply the induction to $\iota \circ \alpha_2$, which has smaller degree than α; hence, we obtain a sequence of separable isogenies of prime degree whose composition is equal to α as wanted.

This is very nice from an abstract point of view, but, given a finite subgroup G of $E(\overline{K})$, we would like to have a more explicit description of the curve E/G and of the isogeny $\phi : E \to E/G$ having G as kernel. These are due to Vélu [Vél71]. For simplicity of exposition, we restrict ourselves to the case when the elliptic curve E has an equation in short Weierstrass form $y^2 = x^3 + Ax + B$, and we consider the case in which either G has cardinality 2 or it is odd. Notice that this covers all separable isogenies of prime degree; then, by applying Corollary 2.5.5 we can cover any case by composing separable isogenies of prime degree and of the Frobenius morphism (if necessary).

Let us first treat the case $\#G = 2$.

Theorem 2.6.6. *Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve defined over K and let $x_0 \in \overline{K}$ be a root of $x^3 + Ax + B$. Define $t := 3x_0^2 + A$ and $w := x_0 t$. Then, the rational map*

$$\phi(x, y) = \left(\frac{x^2 - x_0 x + t}{x - x_0}, \frac{(x - x_0) - t}{(x - x_0)^2} y \right)$$

is a separable isogeny from E to $E' : y^2 = x^3 + A'x + B'$, where $A' = A - 5t$ and $B' = B - 7w$. Moreover, the kernel of ϕ is the group of order 2 generated by $(x_0, 0)$.

Proof. It is clear that ϕ is a separable isogeny of degree 2 with $(x_0, 0)$ in its kernel. The only thing to check is that E' is its image, which is an easy verification (just plug the formulas for $\phi(x, y)$ into the equation for E'). □

Remark 2.6.7. If $x_0 \in K$, then both ϕ and E' will be defined over K, but in general they will be defined over the extension field $K(x_0)$ which contains A' and B'.

Theorem 2.6.8. *Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve defined over K and let G be a finite group of $E(\overline{K})$ of odd order. For each $Q = (x_Q, y_Q) \in G$ not equal to the identity element, let us define*

$$t_Q := 3x_Q^2 + A, \quad u_Q := 2y_Q^2, \quad w_Q = u_Q + t_Q x_Q,$$

and let

$$t := \sum_{Q \in G_{\neq O}} t_Q, \quad w := \sum_{Q \in G_{\neq O}} w_Q, \quad r(x) := x + \sum_{Q \in G_{\neq \mathcal{O}}} \left(\frac{t_Q}{x - x_Q} + \frac{u_Q}{(x - x_Q)^2} \right).$$

The rational map

$$\phi(x, y) = (r(x), r'(x)y)$$

is a separable isogeny from E to $E' : y^2 = x^3 + A'x + B'$, where $A' = A - 5t$ and $B' = B - 7w$ with $\ker \phi = G$.

We will not prove the theorem; for a proof of the result see [Was08, Theorem 12.16]. Notice that the formulas for t, w, r sum over all nozero points in G; however, they depend only on the x-coordinates x_Q. Since $|G|$ is odd and $Q = (x_Q, y_Q) \in G$ if and only if $-Q = (x_Q, -y_Q) \in G$, it suffices to sum over just half of the points of $G_{\neq \mathcal{O}}$ and then double the result. Notice moreover that the elliptic curve E' and ϕ are defined over any extension L/K where G is defined.

Exercise 2.6.9. Let us consider the ellitic curve defined by $y^2 z - x^3 - 16z^3 = 0$ (assume char $K \neq 2, 3$, so that the curve is nonsingular). Consider $G =< [0 : 4 : 1] >$.

- Prove that $|G| = 3$.
- Compute us compute a separable isogeny $\phi : E \to E'$ to some elliptic curve E' with kernel G.

2.7 Division polynomials

The aim of this section is to study the behaviour of the multiplication-by-n map; indeed, we want to study the torsion subgroups

$$E[n] = \{P \in E(\overline{K}) \mid nP = \mathcal{O}\} = \ker([n]),$$

where $[n] : E \to E$ is the multiplication-by-n map. We already studied $E[2]$ and $E[3]$; we want to prove the following theorem.

Theorem 2.7.1. *Let E an elliptic curve over a filed K and let n be a positive integer. If the characteristic of K does not divide n or it is 0, then*

$$E[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

If the characteristic of K is $p > 0$ and $p \mid n$, write $n = p^r n'$ with $p \nmid n'$. Then,

$$E[n] \cong \mathbb{Z}/n'\mathbb{Z} \times \mathbb{Z}/n'\mathbb{Z} \quad \text{or} \quad \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n'\mathbb{Z}.$$

We will prove this theorem in the next section. We will say in characteristic p that a curve is *ordinary* if $E[p] \cong \mathbb{Z}/p\mathbb{Z}$, while is *supersingular* if $E[p] \cong \{\mathcal{O}\}$.

In order to prove the theorem, we need to describe the map on an elliptic curve given by the multiplication by an integer. This is of course an endomorphism of the elliptic curve, and can be described by rational functions. We shall give formulas for these functions.

Let us define the **division polynomials** $\psi_m \in \mathbb{Z}[x, y, A, B]$ by

$$\begin{aligned}
&\psi_0 := 0;\\
&\psi_1 := 1,\\
&\psi_2 := 2y;\\
&\psi_3 := 3x^4 + 6Ax^2 + 12Bx - A^2;\\
&\psi_4 := 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - 8B^2 - A^3);\\
&\psi_{2m+1} := \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3 \quad \text{for } m \geq 2;\\
&\psi_{2m} := (2y)^{-1}(\psi_m)(\psi_{m+2}\psi_{m-1}^3 - \psi_{m-2}\psi_{m+1}^3) \quad \text{for } m \geq 3.
\end{aligned}$$

Notice that in principle it is not clear by the recurrence definition that these are polynomial, so let us prove it by induction.

Lemma 2.7.2. *ψ_n is a polynomial in $\mathbb{Z}[x, y^2, A, B]$ when n is odd, and ψ_n is a polynomial in $2y\mathbb{Z}[x, y^2, A, B]$ when n is even.*

Proof. The lemma is true for $n \leq 4$. Assume, by induction, that it holds for all $n < 2m$; we may assume $2m > 4$, so that $m > 2$. In this case, $2m > m + 2$, so all the polynomials appearing in the definition of ψ_{2m} satisfy the induction assumption. If m is even, then $\psi_m, \psi_{m+2}, \psi_{m-2}$ are in $2y\mathbb{Z}[x, y^2, A, B]$, from which it follows that ψ_{2m} lies in $2y\mathbb{Z}[x, y^2, A, B]$. If m is odd, then ψ_{m-1} and ψ_{m+1} lies in $2y\mathbb{Z}[x, y^2, A, B]$, so again we find that ψ_{2m} is in $\mathbb{Z}[x, y^2, A, B]$. Therefore, the lemma holds for $n = 2m$. A similar argument shows that the same is true for $n = 2m + 1$, proving the statement. □

Let us also define for every $m \geq 1$ the polynomials

$$\phi_m = x\psi_m^2 - \psi_{m+1}\psi_{m-1},$$
$$\omega_m = (4y)^{-1}(\psi_{m+2}\psi_{m-1}^2 - \psi_{m-2}\psi_{m+1}^2).$$

Lemma 2.7.3. *We have that $\phi_n \in \mathbb{Z}[x, y^2, A, B]$ for all n; moreover, if n is odd, then $\omega_n \in y\mathbb{Z}[x, y^2, A, B]$, while if n is even, then $\omega_n \in \mathbb{Z}[x, y^2, A, B]$.*

Proof. If n is odd, then ψ_{n+1} and ψ_{n-1} lies in $y\mathbb{Z}[x, y^2, A, B]$, hence the product lies in $\mathbb{Z}[x, y^2, A, B]$; therefore, $\phi_n \in \mathbb{Z}[x, y^2, A, B]$. The proof is similar if n is even.

The fact that $\omega_n \in y\mathbb{Z}[x, y^2, A, B]$ for odd n and $\omega_n \in \frac{1}{2}\mathbb{Z}[x, y^2, A, B]$ for even n follows from the previous lemma, and this would be enough for the applications. However, for simplicity let us prove that $\omega_n \in \mathbb{Z}[x, y^2, A, B]$ for even n. Indeed, using induction treating separately the various possibilities for $m \bmod 4$, one can show that

$$\psi_n \equiv (x^2 + A)^{(n^2-1)/4} \pmod 2 \quad \text{when } n \text{ is odd},$$

and

$$(2y)^{-1}\psi_n \equiv \left(\frac{n}{2}\right)(x^2 + A)^{(n^2-4)/4} \pmod 2 \quad \text{when } n \text{ is even}.$$

A straightforward calculation then gives the lemma. □

We now consider the elliptic curve given by

$$E : y^2 = x^3 + Ax + B, \quad 4A^3 + 27B^2 \neq 0;$$

we do not specify for now the field of definition of A and B. Using that $y^2 = x^3 + Ax + B$, we regard polynomials in $\mathbb{Z}[x, y^2, A, B]$ as polynomials in $\mathbb{Z}[x, A, B]$; therefore, we can write $\phi_n(x)$ and $\psi_n^2(x)$. Note that ψ_n is not necessarily a polynomial only in x, but its square always is.

Next, we can compute the degrees of $\phi_n(x)$ and $\psi_n^2(x)$; indeed we have the following.

Lemma 2.7.4.

$$\phi_n(x) = x^{n^2} + \textit{lower degree terms}$$
$$\psi_n^2(x) = n^2x^{n^2-1} + \textit{lower degree terms}.$$

Proof. In fact, we claim that

$$\psi_n = \begin{cases} y(nx^{(n^2-4)/2}) + \cdots \text{ if } n \text{ is even} \\ nx^{(n^2-1)/2} + \cdots \text{ if } n \text{ is odd}. \end{cases}$$

This is proved by induction; let us give an example of one of the cases. For example, if $n = 2m + 1$ with m is even, then the leading term of $\psi_{m+2}\psi_m^3$ is equal to

$$(m + 2)m^3y^4x^{\frac{(m+2)^2-4}{2} + \frac{3m^2-12}{2}}.$$

Using that $y^4 = (x^3 + Ax + B)^2$, we have that the leading term becomes

$$(m+2)m^3x^6x^{\frac{(m+2)^2-4}{2}+\frac{3m^2-12}{2}} = (m+2)m^3x^{\frac{(2m+1)^2-1}{2}}.$$

Similarly, the leading term of $\psi_{m-1}\psi_{m+1}^3$ is

$$(m-1)(m+1)^3x^{\frac{(2m+1)^2-1}{2}}.$$

Recalling that the recursion formula gives

$$\psi_{2m+1} := \psi_{m+2}\psi_m^3 - \psi_{m-1}\psi_{m+1}^3,$$

hence subtracting the leading coefficients we get $(2m+1)x^{\frac{(2m+1)^2-1}{2}}$ as wanted. The other cases are treated similarly. □

We can now state the important result, which relates the division polynomials to the multiplication-by-n-map.

Theorem 2.7.5. *Let $P = (x_P, y_P)$ be a point on the elliptic curve defined by the equation $y^2 = x^3 + Ax + B$ (defined over some field of characteristic not 2 and 3); then,*

$$nP = \left(\frac{\phi_n(x)}{\psi_n^2(x)}, \frac{\omega_n(x,y)}{\psi_n(x,y)^3}\right).$$

We will not give a proof of this result here. There are several ways of proving this theorem; one is only algebraic, using induction, but it is highly computational. For a proof using the theory of complex numbers, see [Was08, Section 9.5]. This theorem has although an important corollary.

Corollary 2.7.6. *Let E be an elliptic curve; then the endomorphism of E given by the multiplication by $n \geq 1$ has degree n^2.*

Proof. By Lemma 2.7.4, we have that the maximum of the degrees of the numerator and the denominator of $\phi_n(x)/\psi_n^2(x)$ is n^2; therefore, the degree of the rational function is n^2 provided that this rational function is reduced, i.e. $\phi_n(x)$ and $\psi_n^2(x)$ has no common roots. We will show that this is the case. Suppose not, and call n the smallest index for which $\phi_n(x)$ and $\psi_n^2(x)$ have a common root. We distinguish two cases:

- assume $n = 2m$, i.e. n is even; then, a quick calculation shows that

  $$\phi_2(x) = x^4 - 2Ax^2 - 8Bx + A^2.$$

  Given a point $P = (x, y)$, let us compute the x-coordinate of $[2m]P$ in two steps, namely first multiplying by m and then by 2, and using the fact that

  $$\psi_2(x)^2 = 4y^2 = 4(x^3 + Ax + B),$$

we obtain

$$\begin{aligned}\frac{\phi_{2m}}{\psi_{2m}^2} &= \frac{\phi_2(\phi_m/\psi_m^2)}{\psi_2^2(\phi_m/\psi_m^2)} \\ &= \frac{\phi_m^4 - 2A\phi_m^2\psi_m^4 - 8B\phi_m\psi_m^6 + A^2\psi_m^8}{(4\psi_m^2)(\phi_m^3 + A\phi_m\psi_m^4 + B\psi_m^6)} = \frac{U}{V},\end{aligned}$$

where U and V are the numerator and denominator of the preceding expression. To show that U and V have no common roots, we need the following lemma.

Lemma 2.7.7. *Let* $\Delta = 4A^3 + 27B^2$ *and let*

$$\begin{aligned}F(x,z) &= x^4 - 2Ax^2z^2 - 8Bxz^3 + A^2z^4 \\ G(x,z) &= 4z(x^3 + Axz^2 + Bz^3) \\ f_1(x,z) &= 12x^2z + 16Az^3 \\ g_1(x,z) &= 3x^3 - 5Axz^2 - 27Bz^3 \\ f_2(x,z) &= 4\Delta x^3 - 4A^2Bx^2z + 4A(3A^3 + 22B^2)xz^2 + 12B(A^3 + 8B^2)z^3 \\ g_2(x,z) &= A^2Bx^3 + A(5A^3 + 32B^2)x^2z + 2B(13A^3 + 96B^2)xz^2 - 3A^2(A^3 + 8B^2)z^3.\end{aligned}$$

Then

$$Ff_1 - Gg_1 = 4\Delta z^7 \quad \textit{and} \quad Ff_2 + Gg_2 = 4\Delta x^7.$$

The proof of the lemma is a straightforward computation. Where do these polynomials come from? The polynomial $F(x,1) = \phi_2(x)$ and $G(x,1) = 4(x^3 + Ax + B)$ have no common roots, so using the Euclidean algorithm applied to polynomials, we find polynomials $f_1(x), g_1(x)$ such that $F(x,1)f_1(x) - G(x,1)g_1(x) = 1$. Substituting x with x/z and multiplying all by z^7 (to make things homogeneous), then multiplying also by 4Δ yields the first identity. The second identity is obtained similarly reversing the roles of x and z.

Applying the lemma we have that

$$\begin{aligned}U \cdot f_1(\phi_m, \psi_m^2) - V \cdot g_1(\phi_m, \psi_m^2) &= 4\psi_m^{14}\Delta \\ U \cdot f_2(\phi_m, \psi_m^2) + V \cdot g_2(\phi_m, \psi_m^2) &= 4\phi_m^7\Delta.\end{aligned}$$

If U and V have a common root, then so do ϕ_m and ψ_m^2; since $n = 2m$ is the smallest index such that there is a common root, this is a contradiction. It remains now to show that $U = \phi_{2m}$ and $V = \psi_{2m}^2$; since $U/V = \phi_{2m}/\psi_{2m}^2$, then $U\psi_{2m}^2 = V\phi_{2m}$, and since U and V have no common roots, it follows that ϕ_{2m} is a multiple of U and ψ_{2m}^2 is a multiple of V. Using that $U = \phi_m^4 - 2A\phi_m^2\psi_m^4 - 8B\phi_m\psi_m^6 + A^2\psi_m^8$ and applying Lemma 2.7.4, we have that $U = x^{4m^2} +$ lower degree terms. Lemma 2.7.4 and the fact that ϕ_{2m} is a multiple of U implies that $U = \phi_{2m}$, and hence $V = \psi_{2m}^2$. This shows that ϕ_{2m} and ψ_{2m}^2 have no common roots as wanted.

- Assume that the smallest index n such that there is a common root between ϕ_n and ψ_n^2 is odd, i.e. $n = 2m+1$. Let r be a common root of ϕ_n and ψ_n^2; since

$$\phi_n = x\psi_n^2 - \psi_{n-1}\psi_{n+1}$$

and since $\psi_{n-1}\psi_{n+1}$ is a polynomial in x, we have that $\psi_{n-1}\psi_{n+1}(r) = 0$. But ψ_{n-1}^2 and ψ_{n+1}^2 are both polynomials in x and their product vanishes at r; therefore $\psi_{n+\delta}^2(r) = 0$, where δ is either -1 or 1. Since n is odd, both ψ_n and $\psi_{n+2\delta}$ are polynomials in x. Moreover,

$$(\psi_n\psi_{n+2\delta})^2 = \psi_n^2\psi_{n+2\delta}^2$$

vanishes at r, and so $\psi_n\psi_{n+2\delta}$ vanishes at r. Since

$$\phi_{n+\delta} = x\psi_{n+\delta}^2 - \psi_n\psi_{n+2\delta},$$

we find that $\phi_{n+\delta}(r) = 0$. Therefore we find that $\phi_{n+\delta}$ and $\psi_{n+\delta}^2$ have a common root. Note that $n+\delta$ is even. Recall that, in the even case, we showed that if ϕ_{2m} and ψ_{2m}^2 have a common root, the same happens for ϕ_m and ψ_m. Let us apply this to $n+\delta$; since we assumed n to be the smallest index such that there is a common root, then we have $\frac{n+\delta}{2} \geq n$, which implies that $n = 1$. But clearly $\phi_1 = x$ and $\psi_1^2 = 1$ have no common roots, so we have a contradiction.

This proves that ϕ_n and ψ_n^2 have no common roots in all cases, implying that the multiplication-by-n map has degree n^2, as wanted. □

Notice that, if $n \in \mathbb{Z}$ is negative, then $\deg[n] = \deg([-|n|]) = \deg([-1] \circ [|n|]) = \deg([-1])\deg([|n|]) = |n|^2$ since $\deg([-1]) = 1$; hence the corollary holds also if n is negative.

We are now ready to prove Theorem 2.7.1. Recall that, if $\alpha(x,y) = (r(x), s(x)y)$ is an isogeny, of elliptic curves, then α is separable if and only if r' is not identically zero. We need first a suitable criterion for separability.

Corollary 2.7.8. *Let E be an elliptic curve over a field K. Then the multiplication by n endomorphism $[n] : E \to E$ is separable if and only if n it is not divisible by the characteristic of K.*

Proof. Using Theorem 2.7.5, we know that the multiplication by n map is given by the formula

$$nP = \left(\frac{\phi_n(x)}{\psi_n^2(x)}, \frac{\omega_n(x,y)}{\psi_n(x,y)^3}\right).$$

If n does not divide the characteristic of K then the leading term of $\phi_n(x)'$, i.e. $n^2x^{n^2-1}$, is not zero and therefore

$$\left(\frac{\phi_n(x)}{\psi_n^2(x)}\right)' \neq 0,$$

which implies that $[n]$ is a separable endomorphism. On the other hand if $n \mid \operatorname{char} K$ then the leading coefficient of $\psi_n^2(x)$ vanishes and $\deg \psi_n^2$ is less than $n^2 - 1$. This implies

that the kernel of $[n]$, which consists of the point $\mathcal{O}$ and the points (x_0, y_0) for which $\psi_n(x) = 0$ is strictly smaller than its degree, which is n^2. This shows that in this case $[n]$ is therefore inseparable as wanted. □

We are now ready to prove Theorem 2.7.1.

Proof of Theorem 2.7.1. Assume now that n is not a multiple of the characteristic p of the field; then, from Corollary 2.7.8, we have that the multiplycation-by-n map is separable, hence $E[n]$ has order the degree of the endomophism, which is n^2. Now, applying the structure theorem of finite abelian groups we have that

$$\mathbb{Z}/n_1\mathbb{Z} \times \cdots \times \mathbb{Z}/n_k\mathbb{Z}$$

for some integers $n_1, \ldots, n_k$ with $n_i \mid n_{i+1}$ for all i. Let ℓ be a prime dividing n_1; then, $\ell \mid n_i$ for all i; this means that $E[\ell] \subseteq E[n]$ has order ℓ^k; but we have just proved that the order of $E[\ell]$ is ℓ^2, hence $k = 2$. Multiplication by n annihilates $E[n] \cong \mathbb{Z}/n_1\mathbb{Z} \times \mathbb{Z}/n_2\mathbb{Z}$, so we must have $n_2 \mid n$. Using that $n^2 = \#E[n] = n_1 n_2$; therefore, if $p \nmid n$,

$$E[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}.$$

It remains to consider the case $p \mid n$. We first determine the p-power torsion on E. By Corollary 2.7.8, the multiplication-by-p is not separable, hence $E[p]$ has order strictly less than the degree of the endomorphism, which is p^2. Since every element of $E[p]$ has order either 1 of p, then the order of $E[p]$ is a power of p, hence it must be 1 or p. If $E[p]$ is trivial, then $E[p^k]$ must be trivial for all k. Assume now that $\#E[p] = p$; then we show that $E[p^k] \cong \mathbb{Z}/p^k\mathbb{Z}$. Notice that under this hypothesis we have that $\deg_s[p] = p$; now, the multiplication by p^k map is nothing else than the composition of the multiplication by p k times, and we showed that the separability degree is multiplicative with respect to the composition of isogenies; hence this implies that $\deg_s[p^k] = \deg_s[p]^k = p^k$, which implies that $\#E[p^k] = p^k$ for all $k \geq 1$. Let us now show that there exists an element of order exactly p^k. Assume that there exists an element P of order p^j; since the multiplication-by-p is surjective, there exists a point Q with $pQ = P$. Since $p^j Q = p^{j-1} P \neq \mathcal{O}$ and $p^{j+1} Q = p^j P = \mathcal{O}$, we have that Q has order exactly p^{j+1}. By induction this implies that there are points of order p^k for every k, and so $E[p^k] \cong \mathbb{Z}/p^k\mathbb{Z}$. We can now deal with the general case. Let us write $n = p^r n'$, with $p \nmid n'$; then

$$E[n] \cong E[p^r] \times E[n'].$$

Now from the previous part we have that $E[n'] \cong \mathbb{Z}/n'\mathbb{Z} \times \mathbb{Z}/n'\mathbb{Z}$, and $E[p^r]$ is either 0 or $\mathbb{Z}/p^r\mathbb{Z}$. Using that $\mathbb{Z}/n'\mathbb{Z} \times \mathbb{Z}/p^r\mathbb{Z} \cong \mathbb{Z}/n\mathbb{Z}$, we have

$$E[n] \cong \mathbb{Z}/n'\mathbb{Z} \times \mathbb{Z}/n'\mathbb{Z} \quad \text{or} \quad \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n'\mathbb{Z}.$$

□

Definition 2.7.9. Let E be an elliptic curve defined over a field of characteristic $p > 0$; if $E[p] \cong \mathbb{Z}/p\mathbb{Z}$, then we say that the elliptic curve is **ordinary**; otherwise, we say that the elliptic curve is **supersingular**.

Theorem 2.7.1 allows us to study the set of points of an elliptic curve defined over a finite field. Indeed, we have the following result.

Corollary 2.7.10. *Let E/K be an elliptic curve. Every finte subgroup G of $E(\overline{K})$ is the direct sum of at most two cyclic groups, at most one of which has order divisible by the characteristic p of K. In particular, when $K = \mathbb{F}_q$ is a finite field of characteristic p, we have*

$$E(\mathbb{F}_q) \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$$

for some positive integers m, n with $m \mid n$ and $p \nmid n$.

Proof. Assume that $\#G = n$; then, by Lagrange's theorem, for every $P \in G$ we have $nP = \mathcal{O}$. This implies that $G \subseteq E[n]$, and the thesis follows by applying Theorem 2.7.1. □

Exercise 2.7.11. Let E be an elliptic curve defined by the equation $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$, and let $P = (x_1, y_1) \in E(\mathbb{Q})$ a point of finite order $m > 0$. We want to prove that $x_1, y_1 \in \mathbb{Z}$. As shown in the last section, for any integer n not divisible by m, the x-coordinate x_n of the point $nP = (x_n, y_n)$ is given by $x_n = \phi_n(x_1)/\psi_n^2(x_1)$, where ϕ_n and ψ_n^2 are the division polynomials.

(a) Prove that, for any positive integer $n < m$, if x_n is an integer, then x_1 must be an integer. Use this to reduce to the case that m is prime.
(b) Prove that, if $m = 2$, then P has integer coordinates.
(c) If m is an odd prime, then x_1 is a root of $\psi_m(x) = mx^{(m^2-1)/2} + \cdots \in \mathbb{Z}[x]$. Using this, prove that x_1 is an integer, and deduce that y_1 is an integer as well.
(d) Prove that either $[2]P = \mathcal{O}$ or $y_1^2 \mid 4A^3 + 27B^2$. (**Hint:** Use Lemma 2.7.7 for suitable F and G)

2.8 Endomorphism rings

For any pairs of elliptic curves $E_1, E_2/K$, we have seen the definition of an isogeny $\phi : E_1 \to E_2$ as a non-constant mophism of curves such that $\phi(\mathcal{O}_1) = \mathcal{O}_2$. Recall that these have the property that these maps are also group homomoprhisms with respect to the group law operation over the points of the elliptic curves.

$$\mathrm{Hom}_K(E_1, E_2) = \{\text{isogenies } E_1 \to E_2 \text{ defined over } K\} \cup \{[0]\},$$

where $[0] : E_1 \to E_2$ is the morphism such that $[0]P = \mathcal{O}_2$ for all $P \in E_1(\overline{K})$.[12] We can define the sum on $\mathrm{Hom}_K(E_1, E_2)$ as

$$(\phi + \psi)(P) := \phi(P) + \psi(P) \quad \text{for all } P \in E_1(\overline{K}); \tag{2.1}$$

notice that $\phi + \psi$ is still a morphism from E_1 to E_2 sending $\mathcal{O}_1$ in $\mathcal{O}_2$, hence $\phi + \psi \in \mathrm{Hom}_K(E_1, E_2)$. This implies that $\mathrm{Hom}_K(E_1, E_2)$ is a group.

[12] This is not an isogeny in our definition because it is a constant map, but it is a morphism of curves.

If $E_1 = E_2$, we can also compose morphisms; thus, if we let E be an elliptic curve, we call

$$\mathrm{End}_K(E) := \mathrm{Hom}_K(E, E);$$

then $\mathrm{End}_K(E)$ is a ring with respect to the addition (2.1) and the composition $(\phi\psi)(P) = \phi(\psi(P))$. The ring $\mathrm{End}_K(E)$ is called the **endomorphism ring of** E. The invertible elements of $\mathrm{End}_K(E)$ form the **automorphism group of** E, usually denoted by $\mathrm{Aut}_K(E)$. We will denote by $\mathrm{Hom}(E_1, E_2)$, $\mathrm{End}(E)$ and $\mathrm{Aut}(E_1, E_2)$ the analogous groups and rings where we consider the morphisms defined over $\overline{K}$.

Recall that for an isogeny $\phi : E_1 \to E_2$ we defined its degree as the degree of the field extension $K(E_1)/\phi^*(K(E_2))$, where $\phi^* : K(E_2) \to K(E_1)$ is the map over the function field induced by the composition $f \mapsto f \circ \phi$. We will set by convention $\deg([0]) = 0$, so that it still holds that for every $\phi, \psi \in \mathrm{End}(E)$ we have $\deg(\phi \circ \psi) = \deg(\phi)\deg(\psi)$.

Example 2.8.1. Let $m \in \mathbb{Z}$; then the multiplication-by-m map $[m] : E \to E$ lies in $\mathrm{End}_K(E)$ for every m. Moreover, $[m]$ is nonconstant for all $m \neq 0$.

We have the following resut.

Proposition 2.8.2.

(a) Let E_1, E_2 be elliptic curves; then, the group $\mathrm{Hom}(E_1, E_2)$ is a torsion free group (i.e. without elements of finite order).

(b) Let E be an elliptic curve; then the endomorphism ring $\mathrm{End}(E)$ is a (not necessarily commutative) ring of characteristic zero with no zero divisors.

Proof. (a) Assume that $\phi \in \mathrm{Hom}(E_1, E_2)$ and $m \in \mathbb{Z}$ satisfying

$$[m] \circ \phi = [0].$$

Taking the degrees we have that $(\deg([m] \circ \phi)) = \deg([m])\deg([0]) = 0$, so either $m = 0$ or the fact that $[m] : E \to E$ is nonconstant implies that $\deg([m]) \geq 1$, in which case we must have $\phi = [0]$.

(b) From (a) we have that $\mathrm{End}(E)$ is torsion free. Suppose that there exist $\phi, \psi \in \mathrm{End}(E)$ such that $\phi \circ \psi = [0]$. Again taking the degrees we have that $(\deg \phi)(\deg \psi) = \deg(\phi \circ \psi) = [0]$, which implies that either $\phi = [0]$ or $\psi = [0]$. Therefore $\mathrm{End}(E)$ has no zero divisors.

□

Remark 2.8.3. Assume that $\mathrm{char}(K) = 0$; then, the map

$$[\,] : \mathbb{Z} \to \mathrm{End}(E)$$

sending $m \mapsto [m]$ is usually surjective, i.e. $\mathrm{End}(E) \cong \mathbb{Z}$. If E is strictly larger that $\mathbb{Z}$, then we say that E has *complex multiplication* (or CM). Elliptic curves with complex multiplication have may special properties. On the other hand, if K is a finite field, then $\mathbb{Z} \subsetneq \mathrm{End}(E)$.

Example 2.8.4. Assume that $\mathrm{char}(K) \neq 2$ and let $i \in K$ be a primitive fourth root of unity, i.e. $i^2 = -1$. Then, the elliptic curve E/K given by the equation $y^2 = x^3 - x$ has endomorphism ring $\mathrm{End}(E)$ strictly larger than $\mathbb{Z}$ since it contains a map, which we denote by $[i]$, given by

$$[i] : (x, y) \mapsto (-x, iy).$$

(Notice that, if $(x, y) \in E(\overline{K})$), then $(iy)^2 = -y^2 = -x^3 + x = (-x)^3 - (-x)$). Observe moreover that

$$[i] \circ [i](x, y) = [i](-x, iy) = (x, -y) = [-1](x, y),$$

so $[i] \circ [i] = [-1]$; if we look at the degrees we have that $(\deg([i]))^2 = \deg([-1]) = 1$, hence $\deg([i]) = 1$; but the only multiplication-by-n maps of degree 1 are $[1]$ and $[-1]$, and none of this is equal to $[i]$. This shows that $\mathbb{Z} \subsetneq \mathrm{End}(E)$. There is thus a ring homomorphism

$$\mathbb{Z}[i] \to \mathrm{End}(E), \quad m + ni \mapsto [m] + [n] \circ [i].$$

If $\mathrm{char}(K) = 0$, this map is an isomorphism, i.e. $\mathrm{End}(E) \cong \mathbb{Z}[i]$.

Example 2.8.5. If E is an elliptic curve defined over a finite field $K = \mathbb{F}_q$, then we have the q-power Frobenius endomorphism

$$\pi_q : E \to E \quad (x, y) \to (x^q, y^q)$$

is an endomorphism of E. Moreover, π_q commutes with all the elements of $\mathrm{End}(E)$; this comes from the fact that, over $\mathbb{F}_q$, for every rational function $r \in \mathbb{F}_q(x_1, \ldots, x_n)$ we have $r(x_1, \ldots, x_n)^q = r(x_1^q, \ldots, x_n^q)$, and we can apply this to the rational maps defining any $\alpha \in \mathrm{End}(E)$. This implies that the subring $\mathbb{Z}[\pi_q]$ generated by π_q lies in the center of E.

We notice that it can happen that $\mathbb{Z}[\pi_q] = \mathbb{Z}$. For example, when $E[p] = \{0\}$ and $q = p^2$, then the multiplication-by-p map $[p]$ is purely inseparable, then $[p]$ is necessarily the composition of $\pi^2 = \pi_q$ with an isomophism. If this isomorphism is $[\pm 1]$, then we have that $\pi_q \in \mathbb{Z}$.

Example 2.8.6. Assume again that $\mathrm{char}(K) \neq 2$ and let $a, b \in K$ such that $b \neq 0$ and $r = a^2 - 4b \neq 0$. Consider the two elliptic curves defined by

$$E_1 : y^2 = x^3 + ax^2 + bx$$
$$E_2 : Y^2 = X^3 - 2aX^2 + rX.$$

There are two isogenies of degree 2 connecting these two curves, namely

$$\begin{array}{cc} \phi : E_1 \longrightarrow E_2, & \hat{\phi} : E_2 \longrightarrow E_1 \\ (x, y) \mapsto \left(\frac{x^2+ax+b}{x}, \frac{b-x^2}{x^2} y \right) & (X, Y) \mapsto \left(\frac{X^2-2aX+r}{4X}, \frac{r-X^2}{8X^2} Y \right). \end{array}$$

A direct computation shows that $\hat{\phi} \circ \phi = [2]$ on E_1 and $\hat{\phi} \circ \phi = [2]$ on E_2. The maps ϕ and $\hat{\phi}$ are example of *dual isogenies*, which we will study in the next section.

2.9 Dual isogenies

To further develop our understanding of isogenies and of endomorphism rings, we will now show that, given an isogeny $\phi : E_! \to E_2$, then there exists a sort of *inverse* $\hat{\phi} : E_2 \to E_1$, called *dual isogeny*. This will have many important consequences, for instance the fact that being isogenous is an equivalence relation.

In order to show the existence of the dual isogeny, we will use two important properties of isogenies that we recall here:

- If $\alpha : E_1 \to E_2$ and $\alpha' : E_1 \to E_3$ are separable isogenies with $\ker \alpha = \ker \alpha'$, there exists an isomorphism $\iota : E_3 \to E_2$ such that $\alpha' = \iota \circ \alpha$.
- If $\alpha : E_1 \to E_2$ is an isogeny of degree n, then $\ker \alpha$ is a subgroup of $E_1[n]$; this is because $\# \ker \alpha = \deg_s \alpha$ is a divisor of $n = \deg \alpha$, so every point $P \in \ker(\alpha)$ has order dividing n, i.e. $P \in E_1[n]$.

Theorem 2.9.1. *For any isogeny $\alpha : E_1 \to E_2$ there exists a unique isogeny $\hat{\alpha} : E_2 \to E_1$ for which $\hat{\alpha} \circ \alpha = [n]$, where $n = \deg \alpha$.*

Proof. Let us first prove uniqueness. If $\alpha_1 \circ \alpha = \alpha_2 \circ \alpha$, then $\alpha_1(P) = \alpha_2(P)$ for all $P \in E_2(\overline{K})$ since α is surjective. This implies that the rational maps of α_1 and α_2 must be the same, since they agree on infinitely many points $P \in E_2(\overline{K})$.

Let us now prove existence. First, notice that, if $n = 1$, then α is an isomorphism, hence if we take $\hat{\alpha} = \alpha^{-1}$. Let us now assume $n \geq 2$; we prove the claim by induction on the number of prime factors of n (counted with multiplicity). Let us denote by p the characteristic of the field K.

First, assume that α has prime degree $\ell \neq p$, then α is separable and $\alpha(E_1[\ell])$ is a subgroup of cardinality $\#E_1[\ell]/\deg(\alpha) = \deg([\ell])/\deg(\alpha) = \ell^2/\ell = \ell$ by applying the first homomorphism theorem to $\alpha_{|E_1[\ell]}$ (recall that $\ker \alpha \subseteq E_1[\ell]$). Let $\alpha' : E_2 \to E_3$ be the separable isogeny with $\ker \alpha' = \alpha(E_1[\ell])$; then, $\ker(\alpha' \circ \alpha) = \{P \in E_1 \mid \alpha(P) \in \ker(\alpha')\} = E_1[\ell]$, and since $[\ell] : E_1 \to E_1$ is another isogeny with the same kernel, there exists an isomoprhism $\iota : E_3 \to E_1$ such that $\iota \circ \alpha' \circ \alpha = [\ell]$; if we put $\hat{\alpha} := \iota \circ \alpha'$ we have $\hat{\alpha} \circ \alpha = [\ell]$ as wanted. Let us now assume that α has prime degree equal to the characteristic p; hence there are two cases.

Case 1: If α is separable, then $\ker \alpha$ is a subgroup of $E_1[p]$ of order p, which is the largest possible size of $E_1[p]$ in characteristic p, so $\ker \alpha = E_1[p] \cong \mathbb{Z}/p\mathbb{Z}$ and $\deg_s[p] = p$ (in particular notice that this implies that the curve E_1 in this case is ordinary). Now $\deg[p] = p^2$, hence by Corollary 2.5.5 we have that $[p] = \alpha' \circ \pi$ with $\pi : E_1 \to E_1^{(p)}$ the p-power Frobenius morphism and $\alpha' : E_1^{(p)} \to E_1$ a separable isogeny. If we call $\pi_2 : E_2 \to E_2^{(p)}$, then we have that $\pi_2 \circ \alpha = \alpha^{(p)} \circ \pi_1$ where $\alpha^{(p)} : E_1^{(p)} \to E_1^{(p)}$ is obtained by replacing each coefficient of α by its pth power; indeed, if α is given by $(\frac{u(x)}{v(x)}, \frac{s(x)}{t(x)}y)$, then

$$\pi_2(\alpha(x,y)) = (\frac{(u(x))^p}{(v(x))^p}, \frac{(s(x))^p}{(t(x))^p}y^p) = (\frac{(u^{(p)}(x^p))}{(v^{(p)}(x^p))}, \frac{(s^{(p)}(x^p))}{(t^{(p)}(x^p))}y^p) = (\alpha^{(p)})(\pi_1(x,y)).$$

We have then

$$\ker(\alpha^{(p)} \circ \pi_1) = \ker(\pi_2 \circ \alpha) = \ker \alpha = \ker[p] = \ker(\alpha' \circ \pi_1),$$

since the Frobenius morphisms π_1 and π_2 have trivial kernel, and it follows that $\alpha^{(p)}$ and α' are separable isogenies with the same kernel; there is thus an isomorphism $\iota : E_2^{(p)} \to E_1$ such that $\alpha' = \iota \circ \alpha^{(p)}$. If we now put $\hat{\alpha} = \iota \circ \pi_2$ then

$$\hat{\alpha} \circ \alpha = \iota \circ \pi_2 \circ \alpha = \iota \circ \alpha^{(p)} \circ \pi_1 = \alpha' \circ \pi_1 = [p],$$

as wanted.
Case 2: If α is inseparable then we must have $\alpha = \iota \circ \pi$ for some isomorphism ι. If $E[p] = \{\mathcal{O}\}$, then $[p]$ is purely inseparable of degree p^2, so $[p] = \iota' \circ \pi^2$ for some isomorphism ι', and we may take $\hat{\alpha} = \iota' \circ \pi \circ \iota^{-1}$. If $E[p] \cong \mathbb{Z}/p\mathbb{Z}$, then $[p] = \alpha' \circ \pi$ for some separable isogeny α' of degree p and we may take $\hat{\alpha} = \alpha' \circ \iota^{-1}$.
Let us now treat the case n composite; if so, thanks to Exercise 2.6.5, we can decompose α into a sequence of isogenies of prime degree; in particular we can write $\alpha = \alpha_1 \circ \alpha_2$, where $\deg \alpha_i = n_i < n$ and $n_1 n_2 = n$. Let now $\hat{\alpha} := \hat{\alpha}_2 \circ \hat{\alpha}_1$; where the existence of the $\hat{\alpha}_i$ is given by the inductive hypothesis. Then

$$\hat{\alpha} \circ \alpha = (\hat{\alpha}_2 \circ \hat{\alpha}_1) \circ \alpha = \hat{\alpha}_2 \circ \hat{\alpha}_1 \circ \alpha_1 \circ \hat{\alpha}_2 = \hat{\alpha}_2 \circ [n_1] \circ \alpha_2 = \hat{\alpha}_2 \circ \alpha_2 \circ [n_1] = [n_2] \circ [n_1] = [n],$$

where we used the propery that $[n_1] \circ \alpha_2 = \alpha_2 \circ [n_1]$. This concludes the proof. □

The previous theorem allows us to give the following definition.

Definition 2.9.2. The isogeny $\hat{\alpha}$ given by Theorem 2.9.1 is the *dual isogeny* of α.

As a matter of convinience, we extend the notion of dual isogeny to the map $[0]$ by imposing that $\hat{[0]} = [0]$; this is consistent with the fact that the degrees are multiplicative with respect to compositions and that $\hat{[0]} \circ [0] = [0]$.
We are ready to prove now some properties of the isogeny duals.

Proposition 2.9.3. *Let E_1 and E_2 be two elliptic curves; then,*
(i) if α be an isogeny of degree n, then $\deg \hat{\alpha} = \deg \alpha = n$ and

$$\hat{\alpha} \circ \alpha = \alpha \circ \hat{\alpha} = [n],$$

hence $\hat{\hat{\alpha}} = \alpha$;
(ii) for any integer m, the multiplication-by-m map is self-dual, i.e. $\widehat{[n]} = [n]$;
(iii) for any $\alpha, \beta \in \mathrm{Hom}(E_1, E_2)$, we have that $\widehat{\alpha + \beta} = \hat{\alpha} + \hat{\beta}$;
(iv) for any $\alpha \in \mathrm{End}(E_1)$, we have that $\alpha + \hat{\alpha} = [1 + \deg \alpha - \deg(1 - \alpha)]$.

Notice that in this statement we denote by $1 - \alpha$ the endomorphism $[1] - \alpha$. From now on, when writing $m + \alpha$ for $m \in \mathbb{Z}$ we will implicitly view m as an element of $\mathrm{End}(E)$ via the embedding $\mathbb{Z} \hookrightarrow \mathrm{End}(E)$.

Proof.

(i) The first statement follows from the multiplicativity of the degrees in the composition; indeed, $\deg(\hat{\alpha} \circ \alpha) = \deg(\hat{\alpha}) \deg(\alpha) = \deg[n] = n^2$, hence $\deg(\hat{\alpha}) = n$. Moreover, we note that

$$(\alpha \circ \hat{\alpha}) \circ \alpha = \alpha \circ (\hat{\alpha} \circ \alpha) = \alpha \circ [n] = [n] \circ \alpha,$$

implies that $\hat{\alpha} \circ \alpha = [n]$. This follows from the fact that the isogenies involved are all surjective, hence, since the maps $\hat{\alpha} \circ \alpha$ and $[n]$ coincide on the image of α, they coincide everywhere. This implies that $\hat{\hat{\alpha}} = \alpha$.

(ii) This follows readily from the fact that $[m] \circ [m] = [m^2] = [\deg m]$.

(iii) The proof of this property makes use of the so called *Weil pairing*, that we will see later in the course. For this reason, we postpone the proof of this point.

(iv) For any $\alpha \in \mathrm{End}(E)$ (including $[0]$), we have that

$$[\deg(1-\alpha)] = \widehat{(1-\alpha)} \circ (1-\alpha) = (1-\hat{\alpha}) \circ (1-\alpha) = [1] - (\hat{\alpha} + \alpha) + [\deg(\alpha)],$$

and therefore $\alpha + \hat{\alpha} = [1 + \deg \alpha - \deg(1-\alpha)]$ as wanted.

□

A key consequence of the last statement of the proposition is that $\alpha + \hat{\alpha}$ is always a multiplication-by-t map for some $t \in \mathbb{Z}$. Then we can give the following definition.

Definition 2.9.4. The *trace* of an endomorphism α is the integer $\mathrm{tr}(\alpha) := \alpha + \hat{\alpha}$.

Notice now that for any $\alpha \in \mathrm{End}(E)$ we have $\mathrm{tr}(\hat{\alpha})$=$\mathrm{tr}(\alpha)$ and $\deg(\hat{\alpha}) = \deg(\alpha)$, hence α and $\hat{\alpha}$ have the same characteristic polynomial; more specifically, we have the following.

Proposition 2.9.5. *Let α be an endomorphism of an elliptic curve; both α and its dual $\hat{\alpha}$ are roots of the polynomial*

$$\lambda^2 - \mathrm{tr}(\alpha)\lambda + \deg \alpha = 0.$$

Proof. We have that $\alpha^2 - \mathrm{tr}(\alpha) + \deg(\alpha) = \alpha^2 - (\hat{\alpha} + \alpha)\alpha + \hat{\alpha}\alpha = 0$ and similarly for $\hat{\alpha}$. □

3 Elliptic curves over finite fields

3.1 Hasse bound

Especially for applications to elliptic curve cryptography, we need some properties of elliptic curves over finite fields. Let p be a prime, let $q = p^f$ and let E be an elliptic curve defined over $\mathbb{F}_q$. We want to understand what is the cardinality of $E(\mathbb{F}_q)$. If we want to give an estimate of $\#E(\mathbb{F}_q)$, we have of course that $\#E(\mathbb{F}_q) \leq q^2 + 1$. One can of course do better; indeed, if E is defined by a Weierstrass equation of the form

$$f(x, y) = y^2 + a_1xy + a_3y - x^3 + a_2x^2 - a_4x - a_6 = 0,$$

then for every $x_0 \in \mathbb{F}_q$ the equation $f(x_0, y) = 0$ has at most two solutions; this implies that, for every $x_0 \in \mathbb{F}_q$ there are at most 2 rational points of E such that $P = (x_0, y) \in E(\mathbb{F}_q)$, hence $\#E(\mathbb{F}_q) \leq 2q + 1$.

Remark 3.1.1. If we assume that $f(x_0, y)$ *behaves casually* for varying $x_0 \in \mathbb{F}_q$, then $f(x_0, y)$ should be irreducible the 50% of the times; hence morally one should have that $\#E(\mathbb{F}_q) \sim q$.

We will now prove Hasse's theorem, that gives the exact cardinality of this set.

Theorem 3.1.2. *Let $E/\mathbb{F}_q$ be an elliptic curve over a finite field; then,*

$$E(\mathbb{F}_q) = q + 1 - t,$$

where t is the trace of the Frobenius endomorphism π_E and $|t| \leq 2\sqrt{q}$.

In order to prove this theorem, we need to uderstand the separability and inseparability of the sum of two isogenies. This is given by the following lemma.

Lemma 3.1.3. *Let α and β be two isogenies, with β inseparable. Then, $\alpha + \beta$ is inseparable if and only if α is inseparable.*

Notice that this lemma implies that the sum of two inseparable isogenies is inseparable, while the sum of an inseparable isogeny with a separable one is always separable. On the other hand the sum of two separable isogenies can be either separable or inseparable (why?).

Proof. If both α and β are inseparable, we can write them as $\alpha = \alpha_{\text{sep}} \circ \pi^m$ and $\beta = \beta_{\text{sep}} \circ \pi^n$ with $m, n \geq 1$; this implies that

$$\alpha + \beta = (\alpha_{\text{sep}} \circ \pi^{m-1} + \beta_{\text{sep}}\pi^{n-1}) \circ \pi,$$

which is inseparable. On the other hand, if β is inseparable, then also $-(\beta)$ is inseparable, hence $\alpha+\beta-\beta=\alpha$ is the sum of two inseparable isogenies which we just showed to be inseparable. □

From this, we have in particular the following corollary which will be used later. Recall that, as said at the end of the previous chapter, we will denote the *multiplication-by-n map* $[n]:E\to E$ without the brackets, i.e. just $n:E\to E$. This is not a big deal since $[n]=[m]$ if and only if $n=m$.

Corollary 3.1.4. *Let $\pi_E:E\to E$ given by $(x,y)\mapsto(x^q,y^q)$ the Frobenius endomorphism; then, the map π_E-1 is separable[1].*

Moreover, we will need the following general properties of a *quadratic form* over an abelian group.

Definition 3.1.5. Let A be an abelian group; then, $d:A\to R$ is said to be a *quadratic form* if it satisfies the following properties:

(i) $d(a)=d(-a)$ for every $a\in A$;

(ii) the function $A\times A\to\mathbb{R}$ such that $(a,b)\to d(a+b)-d(a)-d(b)$ is bilinear.

Moreover, we say that d is *positive definite* if

(iii) $d(a=0)$ if and only if $a=0$;

(iv) $d(a)\geq 0$ for every $a\in A$.

We leave the following exercise.

Exercise 3.1.6. If $d:A\to\mathbb{R}$ is a quadratic form, then $d(m\cdot a)=m^2d(a)$ for every $m\in\mathbb{Z}$ and for every $a\in A$.

Recall that in general given two elliptic curves E_1 and E_2, the set $\mathrm{Hom}(E_1,E_2)$ is a group. A very important positive definite quadratic form on this group is given by the degree, as the following lemma shows.

Lemma 3.1.7. *The map* $\deg:\mathrm{Hom}(E_1,E_2)\to\mathbb{Z}\subseteq\mathbb{R}$ *is a positive definite quadratic form.*

Proof. Notice that the only property which is not clear to hold is (ii). To prove that the map $(\phi,\psi)\mapsto\deg(\phi+\psi)-\deg(\phi)-\deg(\psi)$ is bilinear, it is enough to prove it on one side. We have hence to prove that, given $\phi_1,\phi_2,\psi\in\mathrm{Hom}(E_1,E_2)$, we have that

$$\deg(\phi_1+\phi_2+\psi)-\deg(\phi_1+\phi_2)-\deg(\psi)=\\ \deg(\phi_1+\psi)-\deg(\phi_1)-\deg(\psi)+\deg(\phi_2+\psi)-\deg(\phi_2)-\deg(\psi).$$

Notice now that, if we take $\phi,\psi\in\mathrm{Hom}(E_1,E_2)$, we have that

$$\deg(\phi+\psi)-\deg(\phi)-\deg(\psi)=\widehat{(\phi+\psi)}\circ(\phi+\psi)-\hat{\phi}\circ\phi-\hat{\psi}\circ\psi=\\ \hat{\phi}\circ\psi+\hat{\psi}\circ\phi.$$

[1] We mean as explained before $\pi_E-[1]$

If we take now $\phi = \phi_1 + \phi_2$ we get

$$\begin{aligned}\deg(\phi_1 + \phi_2 + \psi) - \deg(\phi_1 + \phi_2) - \deg(\psi) = \\ \widehat{\phi_1} \circ \psi + \widehat{\phi_2} \circ \psi + \hat{\psi} \circ \phi_1 + \hat{\psi} \circ \phi_2 = \\ \deg(\phi_1 + \psi) - \deg(\phi_1) - \deg(\psi) + \deg(\phi_2 + \psi) - \deg(\phi_2) - \deg(\psi),\end{aligned}$$

as wanted. □

We will prove now the following general inequality for positive definite quadratic forms $d : A \to \mathbb{Z}$.

Lemma 3.1.8. *Let $d : A \to \mathbb{Z}$ be a positive definite quadratic form; then, for every $\phi, \psi \in A$*

$$|d(\phi + \psi) - d(\phi) - d(\psi)| \leq 2\sqrt{d(\phi)d(\psi)}.$$

Proof. Notice first that we can assume $\phi, \psi \neq 0$, otherwise there is nothing to prove.

For any $\phi, \psi \in A$, let

$$L(\phi, \psi) := d(\phi + \psi) - d(\phi) - d(\psi)$$

be the bilinear form associated to d; then, since d is positive definite, for every $m, n \in \mathbb{Z}$

$$0 \leq d(m\phi + n\psi) = L(m\phi, n\psi) + d(m\phi) + d(n\psi) = m^2 d(\phi) + n^2 d(\psi) + mnL(\phi, \psi).$$

In particular, taking $m = -L(\phi, \psi)$ and $n = 2d(\phi)$ we have that

$$0 \leq d(\phi)(4d(\phi)d(\psi) - L(\phi, \psi)^2).$$

But d is positive definite by assumption, hence if $\phi \neq 0$, then $|L(\psi, \phi)| \leq 2\sqrt{d(\psi)d(\phi)}$. □

We can now prove Hasse bound.

Proof of the Theorem 3.1.2. Recall that $\mathbb{F}_q$ is defined as the splitting field of the polynomial $x^q - x$ over $\mathbb{F}_p$, where $p = \mathrm{char}(\mathbb{F}_q)$; thus $\mathbb{F}_q$ is precisely the subfield of the q-power Frobenius automorphism $x \mapsto x^q$. The Frobenius endomorphism $\pi_E : E \to E$ is defined by $\pi_E(x : y : z) = (x^q : y^q : z^q)$, therefore

$$E(\mathbb{F}_q) = \{P \in E(\overline{\mathbb{F}}_q) : \pi_E(P) = P\} = \ker(\pi_E - 1),$$

where 1 denotes the multiplication-by-1 map $[1] \in \mathrm{End}(E)$ (i.e. the identity). The morphism $\pi_E - 1$ is separable by Lemma 3.1.4, thus the cardinality of its kernel is equal to its degree. But now

$$\deg(\pi_E - 1) = \widehat{(\pi_E - 1)} \circ (\pi_E - 1) = \hat{\pi}_E \circ \pi_E + 1 - (\hat{\pi}_E + \pi_E) = q + 1 - t.$$

It remains only to show that $|t| \leq 2\sqrt{q}$. To prove this, we use the fact that the degree map on $\mathrm{End}(E)$ is a positive definite quadratic form, hence by Lemma 3.1.8 $|t| = |\#E(\mathbb{F}_q) - q - 1| = |\deg(\pi_E - 1) - \deg(\pi_E) - \deg([1])| \leq 2\sqrt{\deg(\pi_E)\deg([1])} = 2\sqrt{q}$, which concludes the proof. □

Remark 3.1.9. Notice that Hasse's theorem gives a bound for the number of points of $E(\mathbb{F}_q)$, but it does not provide a practical algorithm to compute them if q is large.

Remark 3.1.10. Let $E/\mathbb{F}_q$ be an elliptic curve and let $P, Q \in E(\mathbb{F}_q)$ be points such that Q lies in the subgroup generated by P. The *elliptic curve discrete logarithm problem* (ECDLP) asks for an integer m satisfying $Q = [m]P$. If q is small, we can compute $P, [2]P, [3]P, ...$ until we find Q, but for large values of q it is quite difficult to find m. This led some mathematicians to create a public-key cryptosystem based on the difficulty to solve ECDLP. We will see this later in the course.

3.2 Ordinary and supersingular elliptic curves over finite fields

Next, we want to deepen our study of supersingular elliptic curves, which are commonly used in the applications. Let E/K be an elliptic curve over a field of characteristic $p > 0$; previously we proved that, for any non-zero integer n, the multiplication-by-n map $[n]$ is separable if and only if $p \nmid n$. This implies that the separable degree of the map $[p]$ cannot be $p^2 = \deg[p]$, hence it is either 1 or p, meaning that its kernel $E[p]$ is either cyclic of order p or it is trivial. We said that an elliptic curve E is *ordinary* if $E[p] \equiv \mathbb{Z}/p\mathbb{Z}$ and it is *supersingular* if $E[p] \equiv \{0\}$. We now explore this distinction further in the case when $K = \mathbb{F}_q$ with $q = p^f$. For an elliptic curve defined over $\mathbb{F}_q$, let us denote bt π_E the q-power Frobenius endomorphism and by $\pi : E \to E^{(p)}$ the p-power Frobenius map $(x, y) \mapsto (x^p, y^p)$, where, if E is defined by a Weierstrass equation $y^2 = x^3 + Ax + B$, the curve $E^{(p)}$ is defined by $y^2 = x^3 + A^p x + B^p$. While π_E is an endomorphism of the curve E, the map π is an isogeny but not necessarily an endomorphism. Let us recall some useful facts we proved earlier.

- An isogeny is separable if and only if the cardinality of its kernel is equal to its degree;
- any isogeny α can be decomposed as $\alpha = \alpha_{\text{sep}} \circ \pi$, where α_{sep} is separable;
- if $\alpha = \alpha_{\text{sep}} \circ \pi$, then $\deg(\alpha) = \deg_s(\alpha) \deg_i(\alpha)$, where $\deg_s(\alpha) := \deg(\alpha_{\text{sep}})$ and $\deg_i(\alpha) := p^n$;
- the separable and inseparable degrees are multiplicative with respect to the composition, i.e. $\deg_s(\alpha \circ \beta) = \deg_s(\alpha) \deg_s(\beta)$, and similarly for $\deg_i$;
- a composition of two separable isogenies is separable, and the composition of two inseparable isogenies is inseparable;
- the sum of an separable isogeny with an iseparable one is separable; the sum of two inseparable isogenies is inseparable.

Before analysing the situation over finite fields, let us prove that the property of being ordinary or supersingular is isogeny invariant.

Lemma 3.2.1. *Let $\phi : E_1 \to E_2$ be an isogeny. Then, E_1 is supersingular if and only if E_2 is supersingular.*

Proof. Let $p_1 \in \mathrm{End}(E_1)$ and $p_2 \in \mathrm{End}(E_2)$ denote the multiplication-by-p map on E_1 and E_2 respectively. Since the isogeny is in particular a group homomorphism, we have

$$\begin{aligned} \phi \circ p_1 &= p_2 \circ \phi \\ \deg_s(p_1 \circ \phi) &= \deg_s(p_2 \circ \phi) \\ \deg_s(p_1)\deg_s(\phi) &= \deg_s(p_2)\deg_s(\phi) \\ \deg_s(p_1) &= \deg_s(p_2). \end{aligned}$$

Thus E_1 is supersingular if and only if $\deg_s(p_1) = 1$, and the same holds for E_2 proving the claim. □

We are now ready to derive a criterion to detect whether an elliptic curve defined over a finite field is supersingular.

Theorem 3.2.2. *An elliptic curve $E/\mathbb{F}_q$ is supersingular if and only if $\mathrm{tr}(\pi_E) \equiv 0 \pmod p$, and this holds if and only if the map $[p]$ is purely inseparable.*

Proof. First let us assume that E is supersingular, and assume that $q = p^n$, so that $\pi_E = \pi^n$. Then, $\ker[p] = \ker(\pi \circ \hat{\pi})$ is trivial, hence also $\ker(\hat{\pi})$ is trivial. Thus $\hat{\pi}$ is inseparable since it has degree $p > 1$. Now, since the isogeny $\hat{\pi}^n = \widehat{\pi^n} = \hat{\pi}_E$ is also inseparable, we have that $\mathrm{tr}\,\pi_E = \hat{\pi}_E + \pi_E$ is the sum of two inseparable isomorphisms, hence it is inseparable. But we know that the multiplication-by-n map is inseparable if and only if $p \mid n$, hence $p \mid \mathrm{tr}\,\pi_E$ as wanted. Conversely, if $p \mid \mathrm{tr}\,\pi_E$, then the endomorphism $\mathrm{tr}\ \pi_E$ is inseparable, and also $\hat{\pi}_E = \mathrm{tr}\,\pi_E - \pi_E$ is inseparable. This implies also that $\hat{\pi}^n$, and therefore $\hat{\pi}$ is inseparable. This implies that $\ker\pi$ and $\ker\hat{\pi}$ are inseparable, since they have prime degree, and so $\ker[p] = \ker(\pi\circ\hat{\pi})$ is trivial, proving the claim. □

The previous theorem implies the following important corollart for curves defined over $\mathbb{F}_p$.

Corollary 3.2.3. *Let $E/\mathbb{F}_p$ be an elliptic curve over a field of prime order $p > 3$. Then, E is supersingular if and only if $\#E(\mathbb{F}_p) = p + 1$.*

Proof. By Hasse's Theorem, $|\mathrm{tr}\pi_E| \leq 2\sqrt{p} < p$ for $p > 3$. □

This should convince you that supersingular elliptic curves are rare: indeed, there are $\sim 4\sqrt{p}$ possible values for $\mathrm{tr}\pi_E$, and all but one correspond to ordinary curves. We will now show that, up to isomorphism, every supersingular elliptic curve over a field of characteristic p (not necessarily finite) can be defined over $\mathbb{F}_{p^2}$. More specifically, the following holds.

Theorem 3.2.4. *Let E be a supersingular elliptic curve defined over a field k of characteristic p. Then, $j(E)$ lies in $\mathbb{F}_{p^2}$ (and possibly in $\mathbb{F}_p$).*[2]

[2]Recall that the j-invariant always lies in the minimal field containing A, B.

Proof. Let us assume that E is defined by $y^2 = x^3 + Ax + B$ and, for any prime power q of p, let us denote by $E^{(q)}$ the elliptic curve defined by $y^2 = x^3 + A^q x + B^q$. Let π be the p-power Frobenius isogeny $\pi : E \to E^{(p)}$. Since E is supersingular, the endomoprhism $[p] = \hat{\pi}\pi$ has trivial kernel, so also the isogeny $\hat{\pi} : E^{(p)} \to E$ has trivial kernel, and is therefore purely inseparable of degree p. We can then decompose $\hat{\pi}$ as $\hat{\pi} = \hat{\pi}_{\text{sep}} \circ \pi$, where the separable isogeny $\hat{\pi}_{\text{sep}}$ has degree 1, hence it is an isomorphism. Therefore we have

$$[p] = \hat{\pi}\pi = \hat{\pi}_{\text{sep}}\pi^2,$$

hence it follows that $\hat{\pi}_{\text{sep}}$ is an isomorphism from $E^{(p^2)}$ to E. Now, since two isomorphic elliptic curves have the same j-invariant, we have that

$$j(E) = j(E^{(p^2)}) = (j(E))^{p^2},$$

so $j(E)$ is fixed by the field automorphism $\sigma : x \mapsto x^{p^2}$ of k. It follows that $j(E)$ lies in the subfield of k fixed by σ, which is either $\mathbb{F}_{p^2}$ or $\mathbb{F}_p$ depending on whether k contains a quadratic extension of its prime field or not; in any case $j(E) \in \mathbb{F}_{p^2}$. □

Definition 3.2.5. If E is supersingular, we say that E has *Hasse invariant* 0. Otherwise we say that E has *Hasse invariant* 1.

We have just seen that, up to isomorphism, there are only finitely many elliptic curves with Hasse invariant 0, since each such curve has j-invariant in $\mathbb{F}_{p^2}$. For $p = 2$, one can prove that the only supersingular elliptic curve (over $\overline{\mathbb{F}}_2$) is $E : y^2 + y = x^3$. For $p > 2$, the following theorem gives a criterion to detect whether the elliptic curve is supersingular.

Theorem 3.2.6. *Let $\mathbb{F}_q$ be a finite field of characteristic $p \geq 3$.*

(a) Let $E/\mathbb{F}_q$ be an elliptic curve defined by a Weierstrass equation

$$E : \ y^2 = f(x),$$

where $f(x) \in \mathbb{F}_q[x]$ is a cubic polynomial with distinct roots in $\overline{\mathbb{F}}_q$. Then, E is supersingular if and only if the coefficient of x^{p-1} in $(f(x))^{(p-1)/2}$ is zero;

(b) Let $m = (p-1)/2$ and define a polynomial

$$H_p(t) = \sum_{i=0}^{m} \binom{m}{i}^2 t^i.$$

Let $\lambda \in \overline{\mathbb{F}}_q$ with $\lambda \neq 0, 1$; then, the elliptic curve

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

is supersingular if and only if $H_p(\lambda) = 0$.

(c) The polynomial $H_p(t)$ has distinct roots in $\overline{\mathbb{F}}_q$. There is one supersingular elliptic curve in characteristic 3, and, for $p \geq 5$, the number of supersingular elliptic curves (up to $\overline{\mathbb{F}}_q$-isomorphism) is

$$\left\lfloor \frac{p}{12} \right\rfloor + \begin{cases} 0 & \text{if } p \equiv 1 \pmod{12}; \\ 1 & \text{if } p \equiv 5, 7 \pmod{12}; \\ 2 & \text{if } p \equiv 11 \pmod{12}. \end{cases}$$

For a complete proof of the theorem see [Sil09, Theorem V.4.1]. We give some examples.

Example 3.2.7. Let $p = 11$; then,

$$H_{11}(t) = t^5 + 3t^4 + t^2 + 3t + 1 \equiv (t^2 - t + 1)(t + 1)(t - 2)(t + 5) \pmod{11}.$$

From part c) of the theorem, we have that the curve $E_\lambda : y^2 = x(x-1)(x-\lambda)$ is supersingular if and only if $H_{11}(\lambda) = 0$, hence the possible λ are either one of the of the solutions of $t^2 - t + 1$ or $\lambda \in -1, 2, -5$. Notice moreover that $2 = 1 - (-1)$ and that $2 = -1/5$, hence the curves E_{-1}, E_2 and E_{-5} are isomorphic. If one compute the j-invariants for such curves, we have that the j-invariants of supersingular elliptic curves are either $j = 0$ or $j = 1728 = 1$.

Example 3.2.8. We compute for which primes $p \geq 5$ the elliptic curve

$$E : y^2 = x^3 + 1$$

with $j = 0$ is singular. Part a) of the previous theorem says that we need to compute the coefficient of x^{p-1} of the polynomial $(x^3 + 1)^{(p-1)/2}$. Now notice that, for the term x^{p-1} to appear, we should have that $p - 1 = 3a$ for some a non zero, hence i.e. $p \equiv 1 \pmod 3$. Hence we have that the elliptic curve is supersingular if $p \equiv 2 \pmod 3$, and is ordinary otherwise.

Example 3.2.9. Let us compute also for which primes the curve

$$E : y^2 = x^3 + x$$

with j-invariant 1728 is supersingular. Again, this is equivalent to compute the coefficient of the term x^{p-1} in the polynomial $(x^3 + x)^{(p-1)/2}$. As before, this is 0 if $p \equiv 3 \pmod 4$, hence the curve is supersingular, and it is not zero otherwise.

These example might suggest that, for a given Weierstrass equation with coefficients in $\mathbb{Z}$, the resulting elliptic curve modulo p is supersingular for half of the primes. This is in fact true, provided that the elliptic curve has complex multiplication[3]. On this there is a more precise result of Deuring which we do not state here. The situation is completely different for curves with $\mathrm{End}_{\mathbb{Q}}(E) = \mathbb{Z}$, where *supersingular primes* seems to be very rare.

[3] i.e. $\mathrm{End}_{\mathbb{Q}}(E) \supsetneq \mathbb{Z}$.

Example 3.2.10. Let E be the elliptic curve defined by

$$E: \ y^2 + y = x^3 - x^2 + 10x - 20;$$

this has j-invariant $j(E) = -2^{12}31^3/11^5$. Then, using the criterion a), one can prove that the only primes $p < 100$ such that E is supersingular are $p \in \{2, 19, 29\}$. More generally, Lehmer proved that there are exactly 27 primes $p < 31500$ such that E modulo p is supersingular.

One can prove that, given a curve $E/\mathbb{Q}$, there exists infinitely many primes p such that E modulo p is ordinary; we leave instead some theorems and open conjectures on this topic.

Definition 3.2.11. If $A \subseteq \mathbb{N}$ is a subset of the natural numbers, its density is defined as

$$d(A) = \lim_{n\to\infty} \frac{1}{n}\#(A \cap \{1, \dots n\}).$$

. In particular, if we take two subsets of the natural numbers $A \subseteq B$, we define the relative density of A in B as

$$d_B(A) = \lim_{x\to infinity} \frac{A \cap \{1, \dots, x\}}{\#\{p \in B \mid p \leq x\}}.$$

Theorem 3.2.12 (Serre, Elkies)**.** *Let $E/\mathbb{Q}$ be an elliptic curve such that $End(E) = \mathbb{Z}$; then the set of supersingular primes has relative density 0 (with respect to the prime numbers). More specifically, for every $\varepsilon > 0$ we have*

$$\#\{p < x \mid E/\mathbb{F}_p \text{ is supersingular } \} \ll x^{3/4+\varepsilon}.^4$$

Conjecture (Lang-Trotter)**.** *Let $E/\mathbb{Q}$ be an elliptic curve such that $End(E) = \mathbb{Z}$; then*

$$\#\{p < x \mid E/\mathbb{F}_p \text{ is supersingular } \} \sim \frac{c\sqrt{x}}{\log x}$$

for $x \to +\infty$, where $c > 0$ is a constant depending on E.

Although this conjecture is open in full generality, Elkies proved that there are infinitely many supersingular primes.

Theorem 3.2.13 (Elkies)**.** *Let $E/\mathbb{Q}$ be an elliptic curve such that $End(E) = \mathbb{Z}$; then, there exists infinitely many primes p such that E modulo p is supersingular.*

[4] If we write $A \ll B$ this means that there exists a constant $c > 0$ such that $A \leq cB$; this is usually called the Vinogradov simbol.

4 Algorithmic aspects of elliptic curves and application to cryptography

In this chapter we are going to see various applications of elliptic curves, with a special eye towards the use of elliptic curves in cryptography. Since we are going to analyse various algorithms, we are going to begin with a section where we are going to give a rough definition of algorithmic complexity, which will be useful later in the course.

4.1 Algorithmic complexity

In mathematics and computer science, an *algorithm* is a finite sequence of well-defined instructions, tipically used to solve a class of of specific problems or to perform a computation. When performing an algorithm, it is very important to know its complexity, i.e. a *measure* of how long an algorithm would take to end given an input of size n. Usually, we expect that an algorithm should compute the result within a finite and practical time even for large values of n; for this reason, complexity is usually calculated asymptotically as n goes to infinity. While complexity is usually expressed in terms of time, sometimes complexity is also analysed in terms of space, which translates to the algorithm's memory requirements. Analysis of an algorithm's complexity is helpful especially when comparing algorithms or seeking for improvements.

Big-O notation is the tipical notation used to represent algorithmic complexity; it gives morally an upper bound on complexity and hence it signifies the worse case of the algorithm. With such a notation, it is easy to compare different algorithms because the notation clearly tells how the algorithm scales when input size increases. This is often called *the order of growth*. The following is a list of the main types of complexity we are going to use:

- constant runtime is represented by $O(1)$;
- linear growth is $O(n)$;
- logarithmic growth is $O(\log n)$;
- log-linear growth is $O(n \log n)$;
- quadratic growth is $O(n^2)$
- polynomial growth is $O(p(n))$, where $p(n)$ is a polynomial in n;
- exponential growth is $O(2^n)$;
- factorial growth is $O(n!)$.

The order of growth can also be compared from the best to the worst:

$$O(1) < O(\log n) < O(\sqrt{n}) < O(n) < O(n \log n) < O(n^2) < O(2^n) < O(n!).$$

In complexity analysis, only the dominant term is retained; for example, if an algorithm requires $2n^3 + \log n + 4$ steps, its complexity is $O(n^3)$, since the dominant term is $2n^3$.

4.2 Double-and-add Algorithm

Let E/K be an elliptic curve and let $P \in E(K)$ be a point on E. In most of the applications of elliptic curves, we will need first to have a good algorithm to compute the multiple of a point $[n]P$ for large value of n. The naive way of doing it is to compute successively

$$P, \quad [2]P = P + P, \quad [3]P = [2]P + P, \ \ldots, \quad [n]P = [n-1]P + P.$$

This algorithm takes $n - 1$ steps, where a step consists of adding two points; but if n is large, it is completely useless. All the practical applications of elliptic curves over large finite fields rely on the following exponential improvement.

Theorem 4.2.1 (Double-and-add Algorithm)**.** *Let E/K be an elliptic curve, let $P \in E(K)$ and let $n \geq 2$ be an integer. The following algorithm computes $[n]P$ using no more than $\log_2(n)$ point doublings and no more than $\log_2(n)$ point additions.*

Algorithm:

1. Write the binary expansion of n as
$$n = a_0 + a_1 \cdot 2 + \cdots a_t \cdots 2^t,$$
with $a_i \in \{0, 1\}$ and $a_t = 1$.
2. Set $Q := P$ and $R = \begin{cases} \mathcal{O} & \text{if } a_0 = 0; \\ P & \text{if } a_0 = 1. \end{cases}$
3. for $i = 1, \ldots, t$
   - set $Q = [2]Q$;
   - if $a_i = 1$ set $R = R + Q$;

   end for;
4. return R, which is equal to $[n]P$.

Proof. During the i^{th} iteration of the loop, the value of Q is $[2^i]P$. Since R is incremented by Q if and only if $a_i = 1$, the final value of R is

$$\sum_{i \text{ with } a_i=1} [2^i]P = \sum_{i=0}^{t} [a_i 2^i]P = \left[\sum_{i=0}^{i} a_i 2^i\right] P = [n]P,$$

as wanted.

Notice that each iteration of the loop requires one point duplication and at most one point addition, since $t \leq \log_2 n$, the running time of the algorithm is as stated. □

Remark 4.2.2. Notice that the average running time of the *double-and-add* algorithm to compute $[n]P$ is $\log_2(n)$ doublings and $\frac{1}{2}\log_2(n)$ additions, since the binary expansion of a random integer n has an equal number of 1's and 0's.

Remark 4.2.3. Koblitz has suggested the use of the Frobenius map to further speed the computation of $[n]P$. THe idea is to use an elliptic curve $E/\mathbb{F}_p$ with p small and to take a point $P \in \mathbb{F}_{p^r}$. Then, we replace the doubling map with the easier-to-compute Frobenius map. For more about this, see [Sil09][Remark XI.1.5].

4.3 Counting the number of points in $E(\mathbb{F}_q)$: Schoof's algorithm

Let $E/\mathbb{F}_q$ be an elliptic curve defined over a finite field. Hasse's theorem says that

$$\#E(\mathbb{F}_q) = q + 1 - t \quad \text{with } |t| \leq 2\sqrt{q},$$

where t is the trace of the q-power of the Frobenius endomorphism $\pi_E : E \to E$ sending $(x, y) \mapsto (x^q, y^q)$.

For many applications, especially in cryptography, we need an efficient way to compute the number of points in $E(\mathbb{F}_q)$. For simplicity, let us assume q odd and E given by a Weierstrass equation of the form

$$E : y^2 = f(x),$$

where $f(x)$ is a cubic polynomial without multiple roots. With minor modifications, everything that we do also works in characteristic 2.

The more naive algorithm to compute $\#E(\mathbb{F}_q)$ takes $O(\sqrt{q})$ steps. In this section, we describe Schoof's algorithm, which computes $\#E(\mathbb{F}_q)$ in $O\left((\log q)^8\right)$ steps. The idea is to compute the value of t for a lot of small primes ℓ and then to use the Chinese remainder theorem to reconstruct t.

We first treat separately the case $\ell = 2$. In this case, we know that, if $f(x)$ has a root $e \in \mathbb{F}_q$, then $(e, 0) \in E[2]$ and $(e, 0) \in E(\mathbb{F}_q)$, hence $\#E(\mathbb{F}_q)$ is even (since as a group it contains a point of order 2). In this case, this implies that $q + 1 - a \equiv 0 \pmod 2$, so $a \equiv 0 \pmod 2$. On the other hand, if $f(x)$ has no roots in $\mathbb{F}_q$, then $E(\mathbb{F}_q)$ has no points of order 2, hence $\#E(\mathbb{F}_q)$ must be odd and a must be odd too.

To determine whether the polynomial $f(x)$ has or not a root in $\mathbb{F}_q$, one could in principle try all the elements of $\mathbb{F}_q$, but there is a faster way when q is big. Indeed, recall that $\mathbb{F}_q$ is defined as the field of all the roots of $x^q - x$; therefore, $f(x)$ has a root in $\mathbb{F}_q$ if and only if the greatest common divisor of $f(x)$ and $x^q - x$ is not 1, and this can be efficiently computed by applying the Euclidean algortihm to polynomials.

From now on, we assume $\ell \neq 2$; moreover, for simplicity we will also assume that $\ell \neq p$ where p is the characteristic of the field.

Recall that the q-power Frobenius map has degree q, hence it satisfies

$$\pi_E^2 - t\pi_E + q = 0 \quad \text{in } \mathrm{End}(E).$$

In particular, if $P \in E(\overline{\mathbb{F}}_q)[\ell]$, then

$$\pi_E^2(P) - [t]\pi_E(P) + [q]P = \mathcal{O},$$

so if we assume $P \neq \mathcal{O}$ and we write $P = (x, y)$, then we have

$$(x^{q^2}, y^{q^2}) - [t](x^q, y^q) + [q](x, y) = \mathcal{O}.$$

A key observation is that, since the point $P = (x, y)$ has order ℓ, then we have that

$$[t](x^q, y^q) = [n_\ell](x^q, y^q), \quad \text{where } n_\ell \equiv t \pmod{\ell} \text{ with } 0 \leq n_\ell < \ell.$$

Similarly, we can compute $[q](x, y)$ by first reducing q modulo ℓ. Of course, we don't know the value of n_ℓ, so we need to compute $[n](x^q, y^q)$ for a point $(x, y) \in E[\ell] \setminus \{\mathcal{O}\}$ and check to see whether it satisfies

$$[n](x, y) = (x^{q^2}, y^{q^2}) + [q](x, y).$$

However, the individual points of $E[\ell]$ tends to be defined over fairly large extensions of $\mathbb{F}_q$, so instead of work with one point at the time we work with all the points simultaneously by making use of the division polynomial ψ_ℓ (recall that since ℓ is odd, then $\psi_\ell \in \mathbb{Z}[x]$ and $P = (x_P, y_P) \in E[\ell]$ if and only if $\psi_\ell(x_P) = 0$). For simplicity, we will assume $\ell \neq 2$. The division polynomial $\psi_\ell(x)$ has degree $\frac{\ell^2-1}{2}$ if $p \nmid \ell$ and can be easily computed using the recurrence formulae seen in Section 2.7. We will perform the computations in the quotient ring

$$R_\ell = \frac{\mathbb{F}_q[x, y]}{(\psi_\ell(x), y^2 - f(x))};$$

this implies that anytime we have a power of y we replace y^2 with $f(x)$, and anytime we have a power x^d with $d \geq \frac{\ell^2-1}{2}$ we divide by $\psi_\ell(x)$ and take the reminder. In this way we will work with polynomials of degree at most $\frac{\ell^2-1}{2} - 1$.

Our goal will be to compute the value of the trace of the Frobenius endomoprhism modulo ℓ for enough primes ℓ to determine t. Recall that Hasse's theorem tells that $|t| \leq 2\sqrt{q}$, so it will be sufficient to use all primes $\ell \leq \ell_{\max}$ such that

$$\prod_{\ell \leq \ell_{\max}} \ell > 4\sqrt{q}.$$

The following theorem shows that the following algorithm allows us to compute $\#E(\mathbb{F}_q)$ in $O((\log q)^8)$ steps.

Theorem 4.3.1. *Let $E/\mathbb{F}_q$ be an elliptic curve; the following algorithm computes the cardinality $\#E(\mathbb{F}_q)$ in $O((\log q)^8)$ steps.*

Schoof's Algorithm:

1. Set $A = 1$ and $\ell = 3$;

2. For $A \leq \lfloor 4\sqrt{q} \rfloor$
   3. for $n = 0, \ldots, \ell - 1$
   4. if $(x^{q^2}, y^{q^2}) + [q](x, y) = [n](x^q, y^q)$ in the ring R_ℓ, break;
   5. end for
6. set $A = \ell \cdot A$;
7. set $n_\ell = n$;
8. replace ℓ with the next largest prime;
9. end for;
10. Use the Chinese reminder theorem to find an integer t such that $t \equiv n_\ell \pmod{\ell}$ for all the stored values of n_ℓ;
11. Return the value $\#E(\mathbb{F}_q) = q + 1 - t$.

Sketch of proof. To show that the running time of Schoof's algorithm is $O((\log q)^8)$ we firstly verify three claims.

(a) *The largest prime ℓ used in the algorithm satisfies $O(\log q)$.*
For the prime number theorem, we have that

$$\lim_{X \to +\infty} \frac{1}{X} \sum_{\ell \text{ prime } \leq X} \log \ell = 1.$$

This implies that, asymptotically, $\prod_{\ell \leq X} \ell \sim e^X$, so in order to make the product $\prod_{\ell \leq X} >\geq 4\sqrt{q}$, we have to take $X \sim \frac{1}{2} \log 16q$.

(b) *Multiplication in the ring R_ℓ can be done in $O(\ell^4 (\log q)^2)$ bit operations.*[1]
Notice that elements of the ring R_ℓ are polynomials of degree $O(\ell^2)$. Multiplication of two such polynomials and reduction modulo ψ_ℓ takes $O(\ell^4)$ elementary operations (additions and multiplications) in the field $\mathbb{F}_q$. Similarly, fast multiplication algorithms in $\mathbb{F}_q$ takes $O((\log q)^2)$ bit operations. So, basic operations in R_ℓ take $O(\ell^4 (\log q)^2)$ bit operations.

(c) *It takes $O(\log q)$ ring operations in R_ℓ to reduce $x^q, y^q, x^{q^2}, y^{q^2}$ in the ring R_ℓ.*
Notice that the square-and-multiply algorithm allows us to compute powers x^n and y^n in $O(\log n)$ multiplications in R_ℓ. This computation is done only once, and then the points

$$(x^{q^2}, y^{q^2}) + [q \pmod{\ell}](x, y) \quad \text{and} \quad (x^q, y^q)$$

are computed and stored using step 4 of the algorithm.

Using $(a), (b)$ and (c) we can now estimate the running time of Schoof's algorithm. From (a), we need to use only primes ℓ that are less than $O(\log q)$. Now, there are $O(\log q / \log \log q)$ such primes, so this is how many times the A-loop (2.-9.) is executed.

[1] A bit operation is a basic computer operation on one or two bits. Example of bit operations include addition, moltiplication, and, or, xor, complement. Fancier multiplication methods based for example on fast Fourier transform can be used to reduce multiplication in R_ℓ to $= ((\ell^2 \log q)^{1+\epsilon})$, at the cost of a bigger constant.

Then, each time the A-loop is executed, the n-loop (3.-5.) is executed $\ell = O(\log q)$ times. Furthermore, since $\ell = O(\log q)$, claim (b) says that the basic operations in R_ℓ take $O((\log q)^6)$ bit operations. Moreover, the value $[n](x^q, y^q)$ in step 4. can be computed in $O(1)$ operations in R_ℓ from the previous value $[n-1](x^q, y^q)$ (or else in $O(\log n) = O(\log \log q)$ R_ℓ-operations using the double-and-add algorithm). In total we have that the number of operations in Schoof algorithm is

$$O(\log q) \cdot O(\log q) \cdot O((\log q)^6) = O((\log q)^8) \quad \text{bit operations.}$$

□

Remark 4.3.2. There are optimized variants of Schoof's algorithm, like the SEA algorithm developed by Schoof, Elkies and Atkin, where they work with factors of ψ_ℓ and modular polynomials. For a description of the algorithm, see [BSS00, Chapter 7]

Remark 4.3.3. Let $q \sim 2^{256}$, which is a typical size used in cryptographic applications. We have

$$\prod_{\ell \leq 103} \ell \sim 2^{133.14} > 4\sqrt{q} = 2^{130},$$

so the largest prime ℓ required by Schoff's algorithm in this case is $\ell = 103$. An element of $\mathbb{F}_q[x]/\psi_\ell(x)$ is represented by a $\mathbb{F}_q$-vector of dimension $103^2 \sim 2^{13.4}$, and each element of $\mathbb{F}_q$ is a 256-bit number, so elements of $\mathbb{F}_q[x]/\psi_\ell(x)$ are approximately 2^{22} bits, which is more than 16 KB.

Example 4.3.4. Let $E/\mathbb{F}_{19}$ be the elliptic curve defined by $y^2 = x^3 + 2x + 1$; then $\#E(\mathbb{F}_{19}) = 19 + 1 - t$. We want to determine t by applying Schoof's algorithm. We will show that

$$t \equiv \begin{cases} 1 \pmod 2 \\ 2 \pmod 3 \\ 3 \pmod 5 \end{cases}.$$

Putting this all together, we have that $t \equiv 23 \pmod{30}$; since $|t| < 2\sqrt{19} < 9$, we must have $t = -7$.

Let us start with $\ell = 2$. The polynomial $f(x) = x^3 + 2x + 1$ has no roots in $\mathbb{F}_{19}$; hence, there is no 2-torsion, and $a \equiv 1 \pmod 2$. Let us compute the other two values applying Schoof's algorithm.

We have $q^2 = 361$ and $q \equiv 1 \pmod 3$, and we have to check whether

$$(x^{361}, y^{361}) + (x, y) = [\pm 1](x^{19}, y^{19})$$

for some $(x, y) \in E[3]$ (if not, then $t \equiv 0 \pmod 3$). The third division polynomial is

$$\psi_3(x) = 3x^4 + 12x^2 + 12x - 4.$$

We compute the x-coordinate of $(x^{361}, y^{361}) + (x, y)$ using the addition-law formula:

$$\left(\frac{y^{361} - y}{x^{361} - x}\right)^2 - x^{361} - x = (x^3 + 2x + 1)\left(\frac{(x^3 + 2x + 1)^{180} - 1}{x^{361} - x}\right)^2 - x^{361} - x,$$

where we used the relation $y^2 = x^3 + 2x + 1$. We want now to reduce this quantity modulo ψ_3; to do this, we first need to find (if it exists) the inverse of $x^{361} - x$ modulo ψ_3. On the other hand, a computation using Euclidean algorithm shows that

$$\gcd(x^{361} - x, \psi_3(x)) = x - 8 \neq 1,$$

hence 8 is a root of the polynomial $\psi_3(x)$, and the point of coordinates $(8, 4) \in E[\mathbb{F}_{19}]$ is a point of order 3. This implies that $\#E(\mathbb{F}_{19}) = 19 + 1 - t \equiv 0 \pmod 3$, and so $t \equiv 2 \pmod 3$.

Finally, let us compute $t \pmod 5$. Notice that $19 \equiv -1 \pmod 5$, and so $[19](x, y) = [-1](x, y) = (x, -y)$ for all $(x, y) \in E[5]$. We need to check whether

$$(x', y') := (x^{361,y^{361}}) + (x, -y) = [\pm 2](x^{19}, y^{19}) = [\pm 1](x'', y'') = (x'', -y'')$$

for all $(x, y) \in E[5]$ (if not, then $t \equiv 0 \pmod 5$). Computing the fifth division polynomial using the recurrence formulae gives

$$\psi_5(x) = 5x^{12} + 10x^{10} + 17x^8 + 5x^7 + x^6 + 9x^5 + 12x^4 + 2x^3 + 5x^2 + 8x + 8.$$

The equation for the x-coordinates yields

$$x' = \left(\frac{y^{361} + y}{x^{361} - x}\right)^2 - x^{361} - x \pmod{\psi_5}$$

and

$$x'' = \left(\frac{3x^{88} + 2}{2y^{19}}\right)^2 - 2x^{19} \pmod{\psi_5}.$$

When y^2 is changed to $x^3 + 2x + 1$, this reduces to a polynomial relation in x, which is verified. Therefore

$$t \equiv \pm 2 \pmod 5.$$

To decide the sign, one can look at the y-coordinates, and find that $a \equiv -2 \pmod 5$. As we showed before, this is sufficient to find that $t = -7$, and so $\#E(F_{19}) = 27$.

4.3.1 Arithmetic in R_ℓ

The main idea behind Algorithm 4.3.1 is to run the computation in the subgroup $E[\ell]$ of $E(\mathbb{F}_q)$. However we first need to show that the characteristic polynomial of the Frobenius restricted to $E[\ell]$ gives the reduction of its trace $\bmod \ell$. More precisely we let π_ℓ denote the restriction of the Frobenius morphism to the ℓ-torsion. Similarly we denote by q_ℓ, resp. t_ℓ, the restriction of the multiplications $[q]$, resp. $[t]$. Note that these maps can also be thought as $q_\ell \cdot [1]_\ell$ where q_ℓ is now the element in $\mathbb{Z}/\ell\mathbb{Z}$ corresponding to $q \bmod \ell$ and $[1]_\ell$ is the identity element in $\mathrm{End}(E[\ell])$ (and similarly for t_ℓ).

Lemma 4.3.5. *Let E be an elliptic curve defined over $\mathbb{F}_q$ with Frobenius endomorphism π_E. Let ℓ be a prime non dividing q and $P \in E[\ell]$ nonzero. If*

$$\pi_\ell^2(P) - c\pi_\ell(P) + q_\ell(P) = 0, \tag{4.1}$$

for some integer c, then $c \equiv \mathrm{tr}\,\pi \bmod \ell$.

Proof. Since π_E has degree q, Proposition 2.9.5 implies that

$$\pi_\ell^2(P) - t_\ell \pi_\ell(P) + q_\ell(P) = 0,$$

Subtracting this from equation (4.1) yields $(c - t_\ell)\pi_\ell(P) - 0$. Since $pi_\ell(P)$ is a nonzero element of $E[\ell]$ and ℓ is prime, the point $\pi_\ell(P)$ has order ℓ and hence $\ell \mid c - t_\ell$ as wanted. □

To run Algorithm 4.3.1 we make use of the characterization of $E[\ell]$ via division polynomials. Recall that if $h = \psi_\ell(x, y)$ is the ℓ-th division polynomial of E, by Lemma 2.7.2, since ℓ is odd, $h \in \mathbb{F}_q[x]$. Moreover a point $P \in E(\overline{\mathbb{F}}_q)$ lies in $E[\ell]$ if and only if $h(x_P) = 0$, where $P = (x_P, y_P)$. In particular to represent elements of $\mathrm{End}(E[\ell])$ as rational amps, we can this treat the polynomials appearing in these maps as elements of the ring

$$R_ell = \frac{\mathbb{F}_q[x, y]}{(h(x), y^2 - f(x))},$$

where the Weierstrass equation of E is $y^2 = f(x)$. For example the element π_ℓ can be written as

$$\begin{aligned} \pi_ell = \big(&x^q \bmod h(x), \quad y^q \bmod (h(x), y^2 - f(x))\big) \\ \Big(&x^q \bmod h(x), \quad f(x)^{\frac{q-1}{2}} \bmod h(x)\Big). \end{aligned}$$

Similarly

$$[1]_\ell = (x \bmod h(x), \quad (1 \bmod h(x))y).$$

In particular every nonzero endomorphism appearing in equation (4.1) can be represented in the form $(a(x), b(x)y)$ with a, b elements of $\mathbb{F}_q[x]/(h(x))$, which may be represented as polynomials in $\mathbb{F}_q[x]$ of degree less than $\deg h = (\ell^2 - 1)/2$.

Multiplication in $\mathrm{End}(E[\ell])$ If $\alpha_1 = (a_1(x), b_1(x)y)$ and $\alpha_2 = (a_2(x), b_2(x)y)$ are two elements in $\mathrm{End}(E[\ell])$ their product $\alpha_1\alpha_2$ is defined as the composition

$$\alpha_1 \circ \alpha_2 = (a_1(a_2(x)), \quad b_1(a_2(x))b_2(x)y),$$

where one can reduce $a_3(x) = a_1(a_2(x))$ and $b_3(x) = b_1(a_2(x))b_2(x)$ modulo $h(x)$.

Addition in $\mathrm{End}(E[\ell])$ The sum in the ring $\mathrm{End}(E[\ell])$ is defined using the addition on the elliptic curve E. Thus if $\alpha_1 = (a_1(x), b_1(x)y)$ and $\alpha_2 = (a_2(x), b_2(x)y)$ their sum $\alpha_3 = \alpha_1 + \alpha_2$ can be computed simply applying the formulas for point addition to the coordinates functions of α_1 and α_2.

4.4 Application of elliptic curves to cryptography

In the next lectures we are going to discuss several cryptosystems based on elliptic curves over finite fields, especially on the discrete logarithmic problem for elliptic curves (which we are going to speak about later). The first question that arises is why elliptic curves appear to be useful in cryptographic situations; the main reason is that elliptic curves provide security equivalent to classical system using fewer bits. For example, it has been estimated that a key size of size 4096 bits for RSA gives the same level of security as 313 bits in an elliptic curve system. This means that implementations of elliptic curve cryptosystems require smaller chip size, less power consumption etc. This means that elliptic curves cryptosystems are especially useful for small devices.

We will first give a small introduction to symmetric and asymmetric encryption, and then we will see generalizations of usual cryptographic systems such as Diffie-Hellman key exchange on elliptic curves, as well as cryptographic systems based on elliptic pairings. Finally, we will give a general introduction to isogeny-based cryptography, which is very important nowdays due to the advent of the quantum computers.

4.4.1 A small introduction to public-key cryptography

Assume that Alice wants to send a message, often called the *plaintext*, to Bob, without allowing Eve to read it. In order to do so, she encrypts it to obtain a *ciphertext*. In order to ecrypt the message, Alice uses an *encryption key*; then, Bob uses a *decryption key* to decrypt the ciphertext, which must of course be kept secret from Eve. There are two basis types of encryption; in **symmetric encryption**, the encryption and decryption key are the same, or one can easily be deduced from the other. Popular symmetric encryption methods include the Data Encryption Standard (DES) and the Advanced Encryption Standard (AES). In this case, Alice and Bob need a way to estabilish a common key, and this usually means that they need to have a secret channel in order to send it. This is usually not practical in many situations.

The second type of encryption is called **public key encryption**. This relies on what are known as *one-way trapdoor fucntions*. These are usually easy to compute injective functions $f : A \to B$ with the property that f^{-1} is usually very difficult to compute in general, but becomes quite easy to compute if someone is in possess of some extra piece of information. Thus, if Bob knows the value of k, then Alice can send him a message $a \in A$ by sending $b = f(a)$; then, Bob easily recovers $a = f^{-1}(b)$ since he knows k, but this is impossible for Eve who doesn't know k.

It is not clear whether that such one-way trapdoor functions exist, and indeed, it is still an open problem to prove their existence; however, a number of hard mathematical problems have been posed as the bases for one-way trapdoor functions, including in particular the *discrete logarithm problem*.

Definition 4.4.1. Let G be a group, and let $x, y \in G$ be elements such that y belongs to the subgroup generated by x. The *Discrete logarithm problem* is the problem of

determining an integer $m \geq 1$ such that

$$x^m = y.$$

Obviously, each group has its own discrete logarithm problem, and one can ask how difficult is to solve the problem in any group G. The so called *Pollard ρ algorithm* (see [Sil09, Section IX.5]) is a collision algorithm that takes $O(\sqrt{\#G})$ steps to solve the DLP virtually in any group G. However, this square root estimate is only an upper bound for the computational complexity of the DLP; obviously, the difficulty of the DLP depends also on the nature of the group; here we list some examples.

Example 4.4.2 (The additive group $\mathbb{F}_q^+$)**.** The DLP for the additive group of a finite field $\mathbb{F}_q$ asks for a solution m to the linear equation $mx = y$ for given $x, y \in \mathbb{F}_q$. To solve this equation, we only need to find the multiplicative inverse of x in $\mathbb{F}_q$, which takes $O(\log q)$ steps using the Euclidean algorithm. Thus the DLP in $\mathbb{F}_q^+$ is a very easy problem.

Example 4.4.3 (The multiplicative group $\mathbb{F}_q^*$)**.** The DLP for the multiplicative group of a finite field $\mathbb{F}_q$ asks for a solution m to the exponential equation $x^m = y$ for a given $x, y \in \mathbb{F}_q^*$. In this case there are faster ways to solve the DLP, taking fewer than $O(q^\epsilon)$ steps for every $\epsilon > 0$. These go under by the general name of *index calculus* methods, and they solve the DLP in $\mathbb{F}_q^*$ in

$$\exp\left(c\sqrt[3]{(\log q)(\log\log q)^2}\right) \text{ steps,}$$

where c is a small absolute constant.

Example 4.4.4 (The group $E(\mathbb{F}_q)$)**.** The elliptic curve discrete logarithm problem, usually abbreviated by ECDLP, asks for a solution m to the equation $[m]P = Q$ for given points $P, Q \in E(\mathbb{F}_q)$. Despite extensive research to solve the ECDLP on general curve, the fastest known algorithm is still the general ρ-Pollard algorithm which takes $O(\sqrt{q})$ steps. This fact is one of the primary attractions for using elliptic curves in cryptography.

We will see that there is a key exchange system based on the DLP that is due to Diffie and Hellman, and a public key cryptosystem based on DLP which is called ElGamal. These systems works for any group and are tipically applied either to $\mathbb{F}_q^*$ or $E(\mathbb{F}_q)$. As already noted, at present to solve the ECDLP is much difficult than to solve the DLP in $\mathbb{F}_q^*$; this means that elliptic curve cryptography has key and message sizes that are 5 to 10 times smaller than those for other systems, including RSA and $\mathbb{F}_q^*$-based DLP systems.

Diffie-Hellman Key Exchange: The following procedure allows Alice and Bob to securely exchange the value of a point on an elliptic curve, although neither of them initially knowm the value of the point.

1. Alice and Bob agree on a finite field $\mathbb{F}_q$, an elliptic curve $E/\mathbb{F}_q$, and a point $P \in E(\mathbb{F}_q)$;
2. Alice selects a secret integer a and compute the point $A = [a]P \in E(\mathbb{F}_q)$;
3. Bob selects a secret integer b and computes the point $B = [b]P \in E(\mathbb{F}_q)$;
4. Alice and Bob exchange the values A and B on a (possible insecure) communication line;
5. Alice computes $[a]B = [ab]P$ and $[b]A = [ab]P$; this is now the shared secret between Alice and Bob.

Notice that Diffie-Hellman key exchange allows Alice and Bob to exchange a piece of data they didn't know in advance. This in principle doesn't seem to be so useful; on the other hand, they could in principle use this datum to perform a private key cryptosystem such as the advanced encryption standard (AES).

Diffie-Hellman key exchange security is based on the difficulty to solve the ECDLP. Indeed, Alice and Bob's adversary Eve knows the values of P, $A = [a]P$ and $B = [b]P$, so if Eve can solve the ECDLP, then she can find a (or b) and recover Alice and Bob's secret value. However, in principle, Eve does not need to find a or b; what she needs to do is to compute the point $[ab]P$. However, at present, the only way known to solve the elliptic curve Diffie-Hellman problem is to solve the associated elliptic curve discrete logarithm problem.

Usually the choice of the finite field $\mathbb{F}_q$, of the elliptic curve $E/\mathbb{F}_q$ and of the starting point $P \in E(\mathbb{F}_q)$ is standard. Indeed, one should pay attention to some issues in order to avoid some easy attacks as the following remark shows.

Remark 4.4.5. It is essential than the order of P be divisible by a large prime, since otherwise a Chinese reminder algorithm due to Pohlig and Hellman shows that the solution time of the ECDLP depends only on the largest prime dividing the order of P. For this reason, it is always advisable to use a point P of prime order.

Diffie-Hellman key exchange allow Alice and Bob to exchange a random bit string, but a true public key cryptosystem such as RSA allows Bob to send a specific message to Alice. A public key cryptosystem based on DLP in $\mathbb{F}_q^*$ was proposed in 1985 by ElGamal We will see an elliptic curve version. Before doing this, we first have to show how to represent a message as a point of an elliptic curve. We use a method proposed by Koblitz.

Koblitz' method: Suppose that E is an elliptic curve given by $y^2 = x^3 + Ax + B$ over $\mathbb{F}_p$. The case of an arbitrary field $\mathbb{F}_q$ is similar. Let m be a message, expressed as a number $0 \leq m < p/100$. Let $x_j = 100m + j$ for $0 \leq j < 100$. For every $j = 0, \ldots, 99$, compute $s_j = x_j^3 + Ax_j + B$. If $s_j^{(p-1)/2} \equiv 1 \pmod{p}$, then s_j is a square modulo p, in which case we do not need to try any more values of j. When $p \equiv 3 \pmod{4}$, then a squareroot of s_j is given by $y_j \equiv s_j^{(p+1)/4} \pmod{p}$ (we leave it as an exercise). When $p \equiv 1 \pmod{4}$, a square root of s_j can also be computed (but it is more complicated). In this way, we obtain a point (x_j, y_j) on E. To recover the message m from the point, we have simply to compute $m = \lfloor \frac{x_j}{100} \rfloor$. Notice that, since s_j is essentially a random element of $\mathbb{F}_p^*$, which is cyclic of even order, the probability that s_j is a square is approximately

1/2. Consequently, the probability of not being able to find a point for m after trying 100 values is around 2^{-100}.

We are now ready to describe an elliptic curve version of ElGamal cryptosystem.

ElGamal Public Key Cryptosystem. The following procedure allows Bob to send securely a message to Alice without any previous communication.

1. Alice and Bob agree on a finite field $\mathbb{F}_q$, an elliptic curve $E/\mathbb{F}_q$, and a point $P \in E(\mathbb{F}_q)$.
2. Alice selects a secret integer a and compute the point $A = [a]P \in E(\mathbb{F}_q)$.
3. **Public key**: $(\mathbb{F}_q, E, P, A)$; **private key**: a;
4. Bob chooses a *plaintext* (i.e. a message) $M \in E(\mathbb{F}_q)$ and a random integer k. He computes the two points
$$B_1 = [k]P \in E(\mathbb{F}_q) \quad \text{and} \quad B_2 = M + [k]A \in E(\mathbb{F}_q).$$
5. Bob sends the *ciphertext* (B_1, B_2) to Alice;
6. Alice uses her secret key to compute
$$B_2 - [a]B_1 = (M + [k]A) - [a][k]P = M + [a][k]P - [a][k]P = M.$$

Remark 4.4.6.

- Just as in Diffie-Hellman key exchange, the field $\mathbb{F}_q$, the curve $E/\mathbb{F}_q$ and the point $P \in E(\mathbb{F}_q)$ are chosen from a list published by some trusted authority;
- the ElGamal plaintext is a point $M \in E(\mathbb{F}_q)$, while the ciphertext is a pair of two points (B_1, B_2), so Bob has to send two bits of information for one message. This is less efficient than RSA cryptosystem, where the exchange is 1-1.
- In practise, there is no natural way to assign a message written in, for example English, to a point $M \in E(\mathbb{F}_q)$. A variant of ElGamal system due to Menezes and Vanstone uses the coordinates of a point in E as a mask for the actual message. Moreover, we point out that the ElGamal cryptosystem as described before is in a very raw state and it is subject to various sorts of attacks. All practical secure implementations of modern public key cryptosystems include some sort of internal message structure that allows Alice to verify that Bob's message was properly encrypted. An example of such method is the Integrated Encryption Scheme (IES) due to Abdalla, Bellare, and Rogaway (for more see [MEÁ10]).

Remark 4.4.7. As with Diffie-Hellman key exchange, the ElGamal cryptosystem can be broken by breaking the Diffie-Hellman key exchange. Indeed, Eve knows $A = [a]P$ and $B_1 = [k]P$, so if she can break the Diffie-Hellman problem, then she can compute $[ak]P = [k]A$. Since she also knows B_2, she is then able to compute $B_2 - [k]A = M$ and reach the message.

A public key cryptosystem allows Alice and Bob to exchange information. A *digital signature scheme* has a different purpose, namely it allows Alice to use a private key to sign a digital document, e.g., a computer file, in such a way that Bob can use Alice's

public key to verify the validity of the signature. There are a number of practical digital signature algorithms, we describe here one of them which uses elliptic curves.

Elliptic Curve Digital Signature Algorithm (ECDSA): The following procedure allows Alice to sign a digital document and Bob to verify that the signature is valid.

1. Alice and Bob agree on a finite field $\mathbb{F}_p$, an elliptic curve $E/\mathbb{F}_p$, and a point $P \in E(\mathbb{F}_p)$ of (prime) order N;
2. Alice selects a secret integer a and computes the point $A = [a]P \in E(\mathbb{F}_p)$;
3. Alice publishes the point A, which is the **public verification key**; the secret multiplier a is her **private signing key**.
4. Alice chooses a digital document $d \bmod N$ to sign. To do this, she also chooses a random integer k modulo N. Alice computes $[k]P$ and sets
$$s_1 \equiv x([k]P) \pmod{N} \quad \text{and} \quad s_2 \equiv (d + as_1)k^{-1} \pmod{N}.$$
(We choose a integer representative of $x([k]P)$ between 0 and $p-1$). Alice publishes the signature (s_1, s_2) for the document d.[2]
5. To verify the signature, Bob computes
$$v_1 \equiv ds_2^{-1} \pmod{N} \quad \text{and} \quad v_2 \equiv s_1 s_2^{-1} \pmod{N}.$$
He then computes $[v_1]P + [v_2]A \in E(\mathbb{F}_p)$ and verifies that
$$x([v_1]P + [v_2]A) \equiv s_1 \pmod{N}.$$

Proof. We need to check that, if Alice follows the procedure described in step (4), then the procedure that Bob has to follow to verify the signature works. The point computed by Bob in step (5) is

$$\begin{aligned}[v_1]P + [v_2]A &= [ds_2^{-1}]P + [s_1 s_2^{-1}][a]P \text{ using the values of } s_1, s_2, and A, \\ &= [s_2^{-1}(d + as_1)]P = [k]P.\end{aligned}$$

Hence
$$x([v_1]P + [v_2]A) = x([k]P) \equiv s_1 \pmod{N}$$
by definition of s_1. □

Remark 4.4.8. Before using elliptic curves to exchange keys or messages or to sign documents, Alice and Bob need to chose a finite field $\mathbb{F}_q$, an elliptic curve $E/\mathbb{F}_q$ and a point $P \in E(\mathbb{F}_q)$ having large prime order, and in this choice the have to match certain requests in order to have a secure procedure. In order to make people's life easier, the United States National Institute of Standards and Technology (NIST) publish a list of fifteen fields, curves, and points for Alice and Bob to use. For each of five different security levels, NIST gives one curve $E/\mathbb{F}_p$ with a large prime, one curve $E/\mathbb{F}_{2^k}$ and one so called *Koblitz curve* $E/\mathbb{F}_2$ with a point $P \in E(\mathbb{F}_{2^k})$ for some $k \geq 1$ of large prime order.

[2] We need $s_2 \neq 0 \pmod{N}$ because N is prime and we want s_2 to be invertible. If this is not the case we change the integer k.

4.5 Pairings in cryptography

Another important aspect of the use of elliptic curves in cryptography is the existence of pairing functions that map pairs of points on an elliptic curve into a finite field. The unique properties of these pairing functions have enabled many new cryptographic protocols that had not been previously feasible.

In general, a pairing is a suitable map from the product of two additive groups G_1 and G_2 to a multiplicative group G_T satisfying the following properties:

$$e : G_1 \times G_2 \longrightarrow G_T$$

1. e is **bilinear**, i.e.
   - for every $P_1, P_2 \in G_1$ and $Q \in G_2$ we have
   $$e(P_1 + P_2, Q) = e(P_1, Q) + e(P_2, Q);$$
   - for every $P \in G_1$ and $Q_1, Q_2 \in G_2$ we have
   $$e(P, Q_1 + Q_2) = e(P, Q_1) + e(P, Q_2).$$
2. e is **non-degenerate**, i.e. if $e(P, Q) = 1$ for all P then $Q = 0$, and viceversa if $e(P, Q) = 1$ for all Q then $P = 0$.
3. for practical purposes, e has to be efficiently computable.

Pairings are useful in cryptography because, if constructed properly, they can produce finite fields that are large enough to make the discrete logarithm problem hard to compute, but small enough to make computations efficient.

Pairing-based cryptography has been used to construct identity-based encryption (aka IBE), which allows a sender to excrypt a message without needing a receiver's public key to have been certified and distributed in advance. IBE uses some form of a person (or entity's) identification to generate a public key. This could be an email address, for instance. Besides IBE, there are a number of other applications of pairing-based cryptography, including other identity-based cryptosystems and signature schemes, key establishment schemes, but also attacks to the ECDLP (such as MOV attack, which trasforms the ECDLP into the DLP of the multiplicative group of a finite field, where a sub-exponential index calculus attack can be used to attack the problem). On the other hand, if the target group (i.e. the finite field) to which points are mapped is made sufficiently large, then the discrete logarithmic problem becomes hard again.

Before giving the definition of pairings and of the Weil pairing on elliptic curves, we need to recall some properties of degree zero divisors on Elliptic curves. For more details on this and on divisors on curve we refer to [Was08, Chapter 11] or [Sil09, Chapter II and III].

4.5.1 Degree zero divisors on elliptic curves

Let E be an elliptic defined over a field K; Recall that we defined in Definitions 1.2.10 and 1.2.20 the group of divisors $\mathrm{Div}(E)$ and the subgroup of divisors of degree 0, $\mathrm{Div}_0(E)$.

Recall that a divisor $D \in \mathrm{Div}(E)$ is a formal finite linear combination of symbols with integer coefficients:

$$D = \sum_{P \in E(\overline{K})} a_P P,$$

where $P \in E(\overline{K})$ and $a_P \in \mathbb{Z}$ are non zero only for finitely many P.

There is a **degree** map of divisors

$$\deg \left(\sum_{P \in E(\overline{K})} a_P P \right) = \sum_{P \in E(\overline{K})} a_P \in \mathbb{Z}.$$

Notice that the sum of two divisors of degree 0 is again a divisor of degree 0, hence the subset $\mathrm{Div}_0(E)$, of the divisors of degree 0 forms a subgroup of $\mathrm{Div}(E)$. We can also define the map

$$\begin{aligned} \mathrm{sum} : \mathrm{Div}(E) &\to E(\overline{K}) \\ \sum_{P \in E(\overline{K})} a_P P &\mapsto \sum_{P \in E(\overline{K})} [a_P] P, \end{aligned}$$

where the second addition is the group law on the elliptic curve, and $[a_P]$ denotes the multiplication by a_P isogeny. If we restrict the map to the divisors of degree 0, we obtain a surjective map, since, for every $P \in E(\overline{K})$, then

$$\mathrm{sum}(P - \mathcal{O}) = P.$$

Recall that given a rational function $f \in K(E)$ its divisor, $\mathrm{div}\, f$, defined in 1.2.11, is called a **principal divisor**, and by definition $\mathrm{div}\, f \in \mathrm{Div}_0(E)$. We want to show that in fact the kernel of the map sum restriced to the degree-0 divisors coincide with the set of principal divisors.

Notice that the sum of principal divisors is again a principal divisor, i.e. the set of principal divisors, usually denoted by $\mathrm{Prin}(E)$, forms a subgroup of $\mathrm{Div}^0(E)$.

The following result characterize divisors of degree 0 that are principal divisors.

Theorem 4.5.1. *Let E be an elliptic curve; let D be a divisor on E with $\deg D = 0$. Then, there is a function f on E with $div(f) = D$ if and only if $sum(D) = \mathcal{O}$.*

Proof. To prove this we use the isomorphism $\Phi : E \cong \mathrm{Pic}^0(E)$ of Proposition 1.2.21: the map $\Phi(P) \mapsto P - \mathcal{O}$ is an isomorphism of groups, where Pic^0 is the quotient of $\mathrm{Div}_0(E)$ by the subgorup $\mathrm{Princ}(E)$ of principal divisors. Then, given $D = \sum_Q n_Q Q$ we have that

$$\sum [n_Q] Q = \mathcal{O} \Leftrightarrow \sum [n_Q](Q + \mathcal{O}) = \mathcal{O} \Leftrightarrow \Phi \left(\sum_Q [n_Q](Q + \mathcal{O}) \right) = \sum_Q n_Q Q \in \mathrm{Princ}(E)$$

□

4.5.2 The Weil pairing

We want now to construct a suitable pairing using the elliptic curves. Let E/K be an elliptic curve, and let us fix an integer $m \geq 2$ which we assume to be coprime with the characteristic p of K if $p \neq 0$. We know that, if $p \nmid m$, then $E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}$. Without loss of generality (eventually replacing K with a suitable finite extension), we can assume that K is big enough so that $E[m] \subset K$. We want to construct a pairing

$$e_m : E[m] \times E[m] \to \mu_m,$$

where $\mu_m = \{\alpha \in \overline{K} \mid \alpha^m = 1\}$ is the set of the m-th root of unity in $\overline{K}$.

Let $T \in E[m]$; by Theorem 4.5.1, there exists a function f such that

$$\operatorname{div}(f) = m(T) - m(\mathcal{O}).$$

Choose $T' \in E[m^2]$ such that $mT' = T$; we will show that there exists a function g such that

$$\operatorname{div}(g) = \sum_{R \in E[m]} \left((T' + R) - (R)\right).$$

Indeed, such a function exists if and only if $\sum_{R \in E[m]} (T' + R) - (R) = \mathcal{O}$. Notice that the points R in the sum cancel, so the sum on the left-hand-side is equal to $[\#E[m]]T'$; since $\#E[m] = m^2$, this sum is equal to $\mathcal{O}$. Notice moreover that the choice of g does not depend on the choice of T' since given another point $T'' \in E[m^2]$ such that $mT'' = T$ we have

$$m(T' - T'') = mT' - mT'' = \mathcal{O},$$

hence the formula for $\operatorname{div} g$ using T' or T'' returns the same divisor.

Let us consider the composition $f \circ [m]$; then,

$$\operatorname{div}(f \circ [m]) = m \left(\sum_{mT'' = T} (T'') \right) - m \left(\sum_{R \in E[m]} (R) \right);$$

but we know that two elements A and B such that $mA = mB = T$ differ by an element of $E[m]$, hence we can write $\sum_{mT''=T}(T'') = \sum_{R \in E[m]}(T' + R)$, and so

$$\operatorname{div}(f \circ [m]) = m \left(\sum_{R \in E[m]} (T' + R) \right) - m \left(\sum_{R \in E[m]} (R) \right) = \operatorname{div}(g^m).$$

Therefore, $f \circ [m]$ is a constant multiple of g^m, hence by multiplying f by a suitable constant, we may assume that

$$f \circ [m] = g^m.$$

Let now $S \in E[m]$ and let $P \in E(\overline{K})$; then,

$$g(P + S)^m = f([m](P + S)) = f([m]P) = g(P)^m;$$

consequently,

$$\left(\frac{g(P+S)}{g(P)}\right)^m = 1,$$

i.e. $\frac{g(P+S)}{g(P)} \in \mu_m$. In fact, the quantity is also independent of P; indeed, in the Zariski topology, $g(P+S)/g(P)$ is a continuous function in P, and E is connected, hence it is constant.

We define the **Weil pairing** as

$$e_m : E[m] \times E[m] \to \mu_m \quad e_m(S,T) = \frac{g(P+S)}{g(P)}.$$

Since g is determined up to a scalar multiple, this definition does not depend on the choice of g and, as explained before, does not depend on the choice of the auxiliary point P.

The following theorem states the main properties of the pairing.

Theorem 4.5.2. *Let E be an elliptic curve defined over a field K and let m be a positive integer. Assume that the characteristic of K does not divide m. Then, the Weil pairing:*

$$e_m : E[m] \times E[m] \to \mu_m$$

satisfies the following properties:

1. *e_m is bilinear in each variable; this means that for all $S, S_1, S_2, T, T_1, T_2 \in E[m]$,*

$$e_m(S_1 + S_2, T) = e_m(S_1, T) + e_m(S_2, T)$$

and

$$e_m(S, T_1 + T_2) = e_m(S, T_1) + e_m(S, T_2).$$

2. *e_m is non degenerate in each variable; this means that, if $e_m(S,T) = 1$ for all $T \in E[m]$, then $S = \mathcal{O}$, and if $e_m(S,T) = 1$ for all $S \in E[m]$, then $T = \mathcal{O}$.*
3. *$e_m(T,T) = 1$ for all $T \in E[m]$;*
4. *$e_m(S,T) = e_m(T,S)^{-1}$ for all $S, T \in E[m]$;*
5. *$e_m(\sigma(S), \sigma(T)) = \sigma(e_m(S,T))$ for all automorphisms of $\overline{K}$ that fixes the coefficients of E.*
6. *$e_m(\alpha(S), \alpha(T)) = e_n(S,T)^{\deg \alpha}$ for every endomorphism of E.*

For a proof of this result, see [Was08, Theorem 11.7]. We derive some important consequences given by the properties of the Weil pairing.

Corollary 4.5.3. *Let $\{T_1, T_2\}$ be a basis of $E[m]$; then, $e_m(T_1, T_2)$ is a primitive m^{th}-root of unity.*

Proof. Suppose $e_m(T_1, T_2) = \zeta$ with $\zeta^d = 1$; then, $e_m(T_1, [d]T_2) = 1$ and $e_m([d]T_2, T_2) = 1$. Let $S \in E[m]$ and write $S = [a]T_1 + [b]T_2$; then,

$$e_m(S, [d]T_2) = e_m(T_1, [d]T_2)^a e_m(T_2, [d]T_2)^b = 1.$$

Since this holds for all S, this implies that $[d]T_2 = \mathcal{O}$. Since $[d]T_2 = \mathcal{O}$ if and only if $m \mid d$, it follows that ζ is a primitive m^{th}-root of unity as wanted. □

Corollary 4.5.4. *If $E[m] \subset E(K)$, then $\mu_m \subset K$.*

Proof. Let σ be an automorphisms of $\overline{K}$ such that $\sigma_{|_K}$ =id and let T_1, T_2 be a basis of $E[m]$. Since by assumption $T_1, T_2 \in E(K)$, we have that

$$\zeta = e_m(T_1, T_2) = e_m(\sigma(T_1), \sigma(T_2)) = \sigma(e_m(T_1), e_m(T_2)) = \sigma(\zeta);$$

by the fundamental theorem of Galois theory, since ζ is fixed by every element in $\mathrm{Aut}_K(\overline{K})$, then $\zeta \in K$ as wanted.[3] □

Corollary 4.5.5. *Let E be an elliptic curve defined over $\mathbb{Q}$; then $E[m] \not\subseteq \mathbb{Q}$ for every $n \geq 3$.*

Proof. If $E[m] \subseteq \mathbb{Q}$, then $\mu_m \subset \mathbb{Q}$ and this is not the case if $n \geq 3$. □

Another very important property of the Weil pairing is that it allows to detect whether two points are linearly related. Let $P, Q \in E(\mathbb{F}_q)$ and let N be the order of P. Assume that $\mathrm{GCD}(N, q) = 1$. the following proposition gives a necessary and sufficient condition for the existence of $k \in \mathbb{Z}$ such that $Q = kP$.

Proposition 4.5.6. *There exists k such that $Q = [k]P$ if and only if $[N]Q = \mathcal{O}$ and the Weil pairing $e_N(P, Q) = 1$.*

Proof. If $Q = [k]P$, then $[N]Q = [Nk]P = \mathcal{O}$. Also,

$$e_N(P, Q) = e_N(kQ, Q) = e_N(Q, Q)^k = 1.$$

Conversely, if $NQ = \mathcal{O}$, then $Q \in E[N]$; since $\mathrm{GCD}(N, q) = 1$, then $E[N] \cong \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$. Choose now a point R such that (P, R) is a basis of $E[N]$. Then,

$$Q = [a]P + [b]R,$$

for some integers $a, b \in \mathbb{Z}$. By Corollary 4.5.3, then $e_N(P, R) = \zeta$ is a primitive N^{th}-root of unity. Therefore, if $e_N(P, Q) = 1$, then

$$1 = e_N(P, Q) = e_N(P, P)^a e_N(P, R)^b = \zeta^b,$$

which implies that $N \mid b$ and $[b]R = \mathcal{O}$. Therefore $Q = [a]P$ as wanted. □

The Weil pairing has many applications in cryptography; indeed, pairings can be both used to build cryptographic protocols, but also to to make some attacks for example to the elliptic logarithm problems. One of the attacks for example, called the MOV attack, allows to translate the elliptic discrete logarithm problem in $E(\mathbb{F}_q)$ into one in $\mathbb{F}_{q^m}$. For more about this, see [Was08, Section 5.3]. Other applications to other problems, like the Decision Diffie-Hellman problem or the tripartite Diffie-Hellman can be found at [Was08, Section 6.2].

For the applications, we will need to understand a bit more on the field where the N torsion is defined. We give the following definition.

[3] Notice that the fundamental theorem of Galois theory gives that ζ lies in a purely inseparable extension of K, but since $p \nmid K$ then $K(\zeta)/K$ is separable, hence $\zeta \in K$.

Definition 4.5.7. Let $\mathbb{F}_q$ be a finite field and let $N \geq 1$ be an integer; the *embedding degree* of N in $\mathbb{F}_q$ is the smallest integer $d \geq 1$ such that $\mu_N \subset \mathbb{F}_{q^d}$.

Since $\mathbb{F}_{q^d}^*$ is cyclic of order $q^d - 1$, this is equivalent to d being the smallest integer satisfying $q^d \equiv 1 \pmod{N}$. We have then the following property.

Exercise 4.5.8. Let $E(\mathbb{F}_q)$ be an elliptic curve, let $N \geq 1$ be an integer satisfying $GCD(q-1, N) = 1$, let d be the embedding degree of N in $\mathbb{F}_q$, and suppose that $E(\mathbb{F}_q)$ contains a point of order exact N. Then, $E[N] \subset E(\mathbb{F}_{q^d})$.

Let $P \in E(\mathbb{F}_q)$ be the given point of exact order N defined over $\mathbb{F}_q$ and choose a point $T \in E[N]$ such that $\{P; T\}$ is a basis for $E[N]$. To prove the theorem it is enough to show that $T \in E(\mathbb{F}_{q^d})$. Let $\phi \in Gal_{\overline{F}_q/\mathbb{F}_q}$ be the q-power Frobenius map. Since $P \in E(\mathbb{F}_q)$ we have

$$P^\phi = P \quad \text{and} \quad T^\phi = [a]P + [b]T \ \text{ for some } a, b \in \mathbb{Z}.$$

Using the basic properties of the Weil pairing, we have that

$$\begin{aligned} e_N(P,T)^q &= e_N(P^\phi, T^\phi) \\ &= e_N(P, [a]P + [b]T) = e_N(P,P)^a e_N(P,T)^b = e_N(P,T)^b. \end{aligned}$$

Since $e_N(P,T)$ is a primitive N-th root of unity, we have that $b \equiv q \pmod{N}$. This implies that $T^\varphi = [a]P + [q]T$. Applying repeatedly ϕ to T and using the fact that ϕ fixes P, we have that

$$T^{\phi^d} = [a(1 + q + q^2 \cdots + q^{d-1})]P + [q^d]T.$$

By definition of embedding degree, we have that $q^d \equiv 1 \pmod{N}$, so $[q^d]T = T$.Furthermore, since $N \mid q^d - 1$ but $N \nmid q - 1$, we have that

$$1 + q + \cdots + q^{d-1} \equiv 0 \pmod{N},$$

so $[1 + q + \cdots + q^{d-1}]P = \mathcal{O}$; therefore, $T^{\phi^d} = T$, which proves that $T \in E(\mathbb{F}_{q^d})$ as wanted.

4.5.3 A cryptosystem based on the Weil pairing

We will now analyze a cryptosystem based on the Weil pairing introduced by Boneh and Franklin which falls in the **identity-based** cryptography.

In the cryptosystem we are going to describe each user ahs a public key based on her/his identity, such as an email address. A central trusted authority assigns a corresponding private key to each user. In most of public-key systems, when Alice wants to send a message to Bob, she looks at Bob's public key; however, she needs some way of being sure that this key actually belongs to Bob, rather than someone such as Eve who is acting like Bob. In the present system, the authentication happens in the initial

communication between Bob and the trusted authority. After that, Bob is the only one who has the information necessary to decrypt messages that are encrypted using his public identity.

A simple question is why a system like RSA cannot be used to create such a system. For example, the users could share the same common modulus n, whose factorization is known only to the trusted authority. Bob's identity, call it *Bobid*, could be his encryption exponent. The TA would then compute his decryption exponent and communicate it to him.

When Alice sends a message m to Bob, then she encrypts it as $m^{Bobid} \pmod{n}$. Bob decrypts using the secret exponent provided by the TA. However, anyone who knows the encryption and decryption exponent can find the factorization of n, and then read all the messages in the system. Therefore, the system would not protect secrets. If, instead, a different n is used for each user, then some type of authentication is needed in order to make sure that n is the correct one, and we are back to the original problem.

We will show how to construct such a cryptosystem using pairings; in particular, we will need a pairing $\hat{e} : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$, where $\mathbb{G}_1$ is a group of prime order q. We will then show how to derive such a pairing using the Weil paring defined in the previous section.

The algorithm proposed by Boneh and Franklin consists of four steps:

1. Setup;
2. Authentication;
3. Encryption;
4. Decryption.

Let us analyze the various steps more in detail. Before doing this, we will need the following definition, which we will use later.

Definition 4.5.9. A **Hash function** is any function that can be used to map data of arbitrary size to fixed-size values.

A good hash function satisfies two important properties:

- it should be very fast to compute;
- it should minimize duplication of output values (collisions).

Example 4.5.10. One of the easiest has functions id the Rabin function. Let us choose an integer N which is the product of two primes p and q. Then, the Rabin hash functions takes an integer X, squares it and then take the reminder modulo N.

Let us see how the protocol works.

Setup: In order to set up the system, the trusted authority does the following:

1. Generates a prime q and a group $\mathbb{G}_1$ of order q and put $\mathbb{G}_2 = \mathbb{F}_q$;
2. Chooses a random generator P of $\mathbb{G}_1$;
3. Chooses a space of messages $\mathcal{M} = \{0,1\}^n$;

4. Chooses two hash functions

$$H_1 : \{0,1\}^n \to \mathbb{G}_1 \quad \text{and} \quad H_2 : \mathbb{G}_2 \to \{0,1\}^n,$$

which do the following:
   - the function H_1 takes a string of length n and outputs a point of $\mathbb{G}_1$;
   - the function H_2 inputs an element of $\mathbb{G}_2$ and outputs a binary string of length n, where n is the length of the messages.

5. Chooses a random secret $s \in \mathbb{F}_q^*$ and computes $P_{pub} = sP$.

We will have then

- **Space of messages:** $\mathcal{M} = \{0,1\}^n$;
- **Private key:** $\mathbf{s}$;
- **Public key:** $\mathbf{q}, \mathbf{H_1}, \mathbf{H_2}, \mathbf{n}, \mathbf{P}, \mathbf{P_{pub}}$.

Authentication: If a user (say Bob) with identity ID wants a private key, the TA does the following:

1. Computes $Q_{ID} = H_1(ID)$, giving a point on $\mathbb{G}_1$.
2. Lets $D_{ID} = sQ_{ID}$.
3. After verifying that ID is the identification for the user with whom he is communicating, sends D_{ID} to the user.

Encryption: If Alice wants to send a message M to Bob, she does the following:

1. Looks up Bob's identity, for example $ID = bob@computer.com$ (written as a binary string) and computes $Q_{ID} = H_1(ID)$.
2. Chooses a random $r \in \mathbb{F}_q^*$.
3. Computes $g_{ID} = \hat{e}(Q_{ID}, P_{pub})$.
4. Lets the ciphertext be the pair

$$c = (rP, M \oplus H_2(g_{ID}^r)),$$

where $\oplus$ denotes the XOR (i.e. the bitwise addition modulo 2).

Decryption: Bob decrypts the ciphertext (u, v) as follows:

1. Uses his private key D_{ID} to compute $h_{ID} = \hat{e}(D_{ID}, u)$.
2. Computes $m = v \oplus H_2(h_{ID})$.

The decryption works because

$$h_{ID} = \hat{e}(D_{ID}, rP) = \hat{e}(sQ_{ID}, rP) = \hat{e}(Q_{ID}, P)^{rs} = \hat{e}(Q_{ID}, P_{pub})^r = g_{ID}^r.$$

Therefore,

$$m = (M \oplus H_2(g_{ID}^r)) \oplus H_2(h_{ID}) = (M \oplus H_2(g_{ID}^r)) \oplus H_2(g_{ID}^r) = M,$$

as wanted.

Notice that, given $P, Q_{ID} = aP$, $P_{pub} = sP$ and rP, the goal of an attacker is to find $\mathbf{g_{ID}^r} = \mathbf{\hat{e}(Q_{ID}, P_{pub})^r}$. The security of the system is based on a variant of the Bilinear

Diffie-Hellman problem:

Bilinear Diffie-Hellman problem: Given $\{P, aP, bP, cP\}$ for some $a, b, c \in \mathbb{F}_q^*$, compute

$$W = \hat{e}(P, P)^{abc}.$$

Let us now see how to construct such a pairing to implement the IBE using the Weil pairing. To do this, we will focus on a specific example. Let us give a prime $p \equiv 2 \pmod 3$ and let us consider the curve defined $E/\mathbb{F}_p$ defined by $y^2 = x^3 + 1$. One can prove that this curve is supersingular, hence $\#E(\mathbb{F}_p) = p + 1$. Let us take a prime $q > 3$ such that $q \mid p + 1$ and let $P \in E(\mathbb{F}_p)$ be a point of order q. We know that $E[q] \cong \mathbb{Z}/q\mathbb{Z} \times \mathbb{Z}/q\mathbb{Z}$; moreover, since $q \mid p^2 - 1$, the embedding degree of q is 2, and by Exercise 4.5.8 we have that $E[q] \subset \mathbb{F}_{q^2}$. Let us now take $\mathbb{G}_1 =< P >$, which is a cyclic group of order q, and let $\mathbb{G}_2$ be the subgroup of $\mathbb{F}_{p^2}$ of order q.

Notice that, by Proposition 4.5.6, we have that for every $Q, R \in \mathbb{G}_1$, then $e_q(Q, R) = 1$. We need hence to modify the Weil pairing to avoid this. Let us consider the map

$$\phi : E \to E \quad (x, y) \mapsto (\omega x, y),$$

where ω is a primitive third root of unity. This is an isomorphism, so if P has order q, also $\phi(P)$ has order q; but now $3 \nmid p - 1$, hence $\omega \notin \mathbb{F}_p$, and so $\phi(P) \notin E(\mathbb{F}_p)$. This implies that, if $Q \in< P >$, then $\phi(Q) \notin< P >$, hence $e_q(P, \phi(Q)) \neq 1$.

We can define the *modified Weil pairing* as

$$\hat{e}_q : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2 \quad (P, Q) \mapsto e_q(P, \phi(Q)).$$

Remark 4.5.11. There are also other types of pairings that one can define on elliptic curves; one is the Tate-Lichtenbaum pairing. For more about this, see [Was08, Section 11.3]. To apply the pairings, one should have an efficient way to compute the Weil pairing. This can be done by applying Miller's algorithm (for more details about this, see [Was08, Section 11.4]).

Bibliography

[BM02] Ezra Brown and Bruce T. Myers. "Elliptic Curves from Mordell to Diophantus and Back". In: **The American Mathematical Monthly** 109.7 (2002), pp. 639–649.

[BSS00] I. F. Blake, G. Seroussi, and N. P. Smart. **Elliptic curves in cryptography**. Vol. 265. London Mathematical Society Lecture Note Series. Reprint of the 1999 original. Cambridge University Press, Cambridge, 2000, pp. xvi+204. ISBN: 0-521-65374-6.

[Fal83] G. Faltings. "Endlichkeitssätze für abelsche Varietäten über Zahlkörpern". In: **Invent. Math.** 73.3 (1983), pp. 349–366.

[Kob87] Neal Koblitz. "Elliptic curve cryptosystems". In: **Math. Comp.** 48.177 (1987), pp. 203–209.

[Mat70] Ju. V. Matijasevič. "The Diophantineness of enumerable sets". In: **Dokl. Akad. Nauk SSSR** 191 (1970), pp. 279–282. ISSN: 0002-3264.

[Maz77] B. Mazur. "Modular curves and the Eisenstein ideal". In: **Inst. Hautes Études Sci. Publ. Math.** 47 (1977). With an appendix by Mazur and M. Rapoport, 33–186 (1978).

[MEÁ10] V Gayoso Martínez, L Hernández Encinas, and C Sánchez Ávila. "A survey of the elliptic curve integrated encryption scheme". In: **ratio** 80.1024 (2010), pp. 160–223.

[Mil86] Victor S. Miller. "Use of elliptic curves in cryptography". In: **Advances in cryptology—CRYPTO '85 (Santa Barbara, Calif., 1985)**. Vol. 218. Lecture Notes in Comput. Sci. Springer, Berlin, 1986, pp. 417–426.

[Sil09] Joseph H. Silverman. **The arithmetic of elliptic curves**. Second. Vol. 106. Graduate Texts in Mathematics. Springer, Dordrecht, 2009, pp. xx+513.

[Sut23] Andrew Sutherland. "Lecture notes for the course *18.783 Elliptic Curves*". 2023. URL: https://math.mit.edu/classes/18.783/2023/lectures.html.

[TW95] Richard Taylor and Andrew Wiles. "Ring-theoretic properties of certain Hecke algebras". In: **Ann. of Math. (2)** 141.3 (1995), pp. 553–572.

[Vél71] Jacques Vélu. "Isogénies entre courbes elliptiques". In: **C. R. Acad. Sci. Paris Sér. A-B** 273 (1971), A238–A241. ISSN: 0151-0509.

[Was08] Lawrence C. Washington. **Elliptic curves**. Discrete Mathematics and its Applications. Second Ed. Chapman & Hall/CRC, Boca Raton, FL, 2008, pp. xviii+513.

[Wil95] Andrew Wiles. "Modular elliptic curves and Fermat's last theorem". In: **Ann. of Math. (2)** 141.3 (1995), pp. 443–551.

[Zan09] Umberto Zannier. **Lecture notes on Diophantine analysis**. Vol. 8. Appunti. Scuola Normale Superiore di Pisa (Nuova Serie) [Lecture Notes. Scuola Normale Superiore di Pisa (New Series)]. With an appendix by Francesco Amoroso. Edizioni della Normale, Pisa, 2009, pp. xvi+237.