Elliptic curves over finite fields and their pairings

An elementary and rigorous account

Matan Prasma

September 14, 2023

Contents

1 Introduction

Since the construction of Miller's algorithm [Mil], the Cryptography community started to use elliptic curves and their pairing extensively. By now, many publicly available code libraries allow one to efficiently compute elliptic curves over finite fields and evaluate their pairings. However, compared to Machine Learning, where the mathematical prerequisites consist of Linear Algebra, Calculus and basic Statistics, elliptic curves require more background and are usually taught at a master level in pure Mathematics. This state of affairs poses a challenge to engineers and others who wish to understand the mathematical building blocks.

These notes grew as part of a math seminar I gave in Aragon Association during 2022, to assist overcoming the challenge mentioned above. They aim to give a self-contained, rigorous and elementary account of most of the math required for pairing-based Cryptography. I collected material from several standard sources, and sometimes formulated elementary arguments to replace non-elementary explanations. In particular, I completely avoid relying on Galois Theory or Algebraic Geometry unlike most textbooks on the subject.

I'd like to thank Amir Taaki, Alex Kampa, Artem Grigor, Roger Baig and Arnaucube for many useful comments during the writing of this manuscript.

2 Naive Set Theory

As our logic syntax we use the symbols $\forall$, $\exists$, $!$, $\neg$, $\Rightarrow$ and $\iff$ to denote 'for all', 'exists', 'unique', 'not', 'implies' and 'if and only if' (or 'implies and implied') respectively.

We typically define a new notion by saying that something is called 'name' **if** it satisfies a certain condition. By convention, this 'if' is meant as an 'if and only if' in that we will call something 'name' only if it satisfies this condition.

2.1 Sets and functions

Slogan. *Sets are the machine code of modern Mathematics.*

On a fundamental level, modern Math is built on Set Theory. From that point of view, a **Set** S is a collection of elements such that for every object x in our 'universe' we can determine whether x is an element of S, denoted $x \in S$ or that x is not an element of S, denoted $x \notin S$.

When we want to specify the elements of a set S, we do so with curly brackets and commas separating between elements e.g. $S = \{a, b, c\}$. Repeated elements in a set are ignored so $\{1, 1, 2, 3\} = \{1, 2, 3\}$. Also the order of elements does not matter so $\{2, 3, 1\} = \{1, 2, 3\}$.

If S has finite number of elements (or just 'finite') we denote by $\#S$ (or $|S|$) the number of elements of S. Of course, S need not be finite, and in this case, we need a rule in order to specify the elements of S, e,g. $S = \{n | n \geq 2\}$ or if the rule is obvious, we can write $S = \{2, 3, 4, ...\}$. For sets A, B we write $A \subseteq B$ if $\forall a \in A, a \in B$ and say that A is included in B. The basic operations on sets include **union**

$$A \cup B = \{x | x \in A \vee x \in B\},$$

intersection

$$A \cap B = \{x | x \in A \wedge x \in B\}$$

and **complement** (or subtraction)

$$A \setminus B = \{x | x \in A \wedge x \notin B\}.$$

Remark 2.1. More generally, let I be a set that we refer to as an 'index set'. Suppose that for every $i \in I$ we are given a set U_i. Then we can form the union

$$\bigcup_{i \in I} U_i = \{x | \exists i \in I : x \in U_i\}$$

and the intersection

$$\bigcap_{i \in I} U_i = \{x | \forall i \in I : x \in U_i\}.$$

Our fundamental assumption is that there exist a special set, called the **empty set** and denoted $\varnothing$ that has no elements. More formally, we can write

$$\varnothing = \{x | x \neq x\}$$

and observe that for every set A we have $\varnothing \subseteq A$. Using the empty set, we can in fact define all natural numbers as follows:

$$0 := \varnothing,$$

$$1 := \{\varnothing\},$$

$$2 := \{\varnothing, \{\varnothing\}\} = \{0, 1\},$$

...,

$$n := (n-1) \cup \{n-1\} = \{0, 1, ..., n-1\}.$$

Let's define addition of natural numbers. For $n \in \mathbb{N}$ we take **successor** of n to be $S(n) = n \cup \{n\} = n+1$. Then define recursively the addition by setting first $\forall n \in \mathbb{N}$, $n + 0 := n$. For a fixed $m \in \mathbb{N}$, assume we defined $n + m \in \mathbb{N}$ for all $n \in \mathbb{N}$. Then for any $n \in \mathbb{N}$ we define $n + (m+1) = n + S(m) := S(n) + m$.

Exercise 2.2. Define multiplication of natural numbers in a similar fashion.

Example 2.3. Some special sets and their notation are given below.

1. $\mathbb{N} = \{1, 2, 3, ...\}$ the natural numbers.

2. $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ the integers.

3. $\mathbb{Q} = \{\frac{a}{b} | a, b \in \mathbb{Z} \wedge b \neq 0\}$ the rational numbers.

4. $\mathbb{R}$ the real numbers.

The elements of a set have no particular order and we remove duplicated elements so that for example $\{a, b, c\} = \{a, c, a, b\}$. In case we wish to talk about **ordered elements** we can do the following: Given to objects, a, b we can consider the set of two elements

$$O_{ab} = \{\{a\}, \{a, b\}\}.$$

If a', b' are any two other elements, we have that

Observation 2.4. $O_{ab} = O_{a'b'}$ if and only if $a = a'$ and $b = b'$.

Proof. If $a = a'$ and $b = b'$ then clearly $O_{ab} = O_{a'b'}$. Conversely, if $O_{ab} = O_{a'b'}$ then in particular $\{a\} \in \{\{a'\}, \{a', b'\}\}$. But $\{a\} \neq \{a', b'\}$ since the first set has one element and the second has two elements. Thus, $\{a\} = \{a'\}$ which means $a = a'$. Next, we have $\{a, b\} \in \{\{a'\}, \{a', b'\}\}$ and since $\{a, b\} \neq \{a'\}$ we get $\{a, b\} = \{a', b'\}$. Since we already know $a = a'$, it follows that $b = b'$. □

We refer to the set O_{ab} as the **ordered pair** of a, b and denote it $(a, b) := O_{ab}$. The definition of O_{ab} could have been different and we gave the one above as a convention. In fact all we need from it is the property stated in 2.4.

Using the notion of ordered pairs we can make the following

Definition 2.5. Given two sets A, B, the **Cartesian product** of A and B is the set
$A \times B := \{(a, b) | a \in A, b \in B\}$.

For example, if $A = \{0,1\}$ and $B = \{1,2\}$ then

$$A \times B = \{(0,1),(0,2),(1,1),(1,2)\}.$$

As another example, if $A = B = \mathbb{R}$, then

$$A \times B = \{(x,y)|x,y \in \mathbb{R}\} =: \mathbb{R}^2$$

i.e. the two dimensional plane aka the X-Y plane.

Example 2.6. Let

$$A = [2,4] = \{x \in \mathbb{R}|2 \leq x \leq 4\} \subseteq \mathbb{R}$$

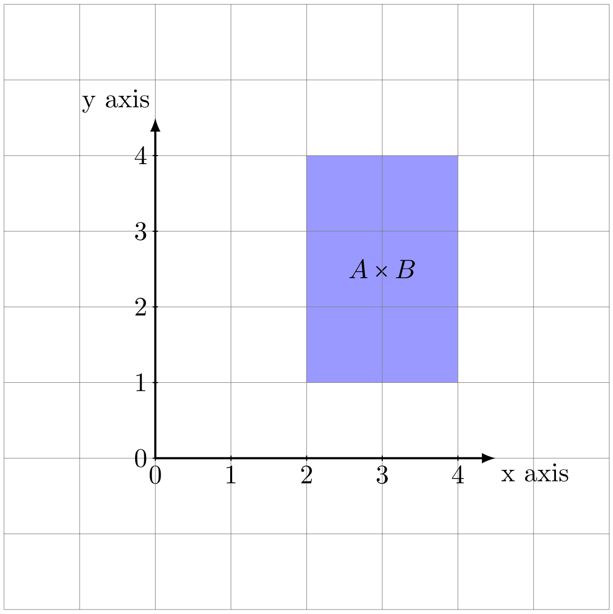
and

$$B = [1,4] = \{x \in \mathbb{R}|1 \leq x \leq 3\} \subseteq \mathbb{R}$$

be a pair of intervals.
Then

$$A \times B \subseteq \mathbb{R}^2$$

can be depicted as follows:



In order to 'move' between sets, we need functions. Given sets A, B a function f from A to B is a rule that assigns for each $a \in A$ a unique element in B, denoted $f(a)$. We denote such a function by $f : A \longrightarrow B$. Informally speaking, we can think of the function f as a machine with set of inputs A and a set of possible outputs B; the machine f assigns to each input $a \in \mathcal{A}$ a unique output $f(a) \in B$. But what is a rule? to give a more formal definition we go as follows:

Definition 2.7. A function $f : A \longrightarrow B$ is a subset $f \subseteq A \times B$ that satisfies two properties:

1. for all $a \in A$, there is $b \in B$, also denoted as $b := f(a)$ such that $(a,b) \in f$ (so that f is defined on all elements in A).
2. if $(a,b) \in f$ and $(a,b') \in f$ then $b = b'$ (so that f gives a unique element in B for every element in A).

The set A is called the **domain** of f and the set B is the **range** of f. We will use the terms 'function' and 'map' interchangeably.

Example 2.8. Let $A = \varnothing$ and B any set. Then $A \times B = \varnothing$ so that there can be at most one function $f : \varnothing \longrightarrow B$, namely the one corresponding to $\varnothing \subseteq A \times B$. One can see that the conditions of Definition 2.7 are vacantly satisfied. This function is called the empty function. On the other hand, if $A \neq \varnothing$ and $B = \varnothing$ we have again $A \times B = \varnothing$ but now $\varnothing \subseteq A \times B$ is not a function $A \longrightarrow B$ since condition 1 of Definition 2.7 is not satisfied (since there is at least one element a in A).

We say two functions $A \underset{f'}{\overset{f}{\rightrightarrows}} B$ (with the same domain and range) are equal, denoted $f = f'$, if for every $a \in A$, $f(a) = f(a')$ or in other words if $f = f' \subseteq A \times B$ as sets.

Definition 2.9. Suppose $f : A \longrightarrow B$ and $g : B \longrightarrow C$ are functions. We define their **composition** $g \circ f$ to be the function $g \circ f : A \longrightarrow C$ specified by the rule

$$(g \circ f)(a) := g(f(a))$$

for all $a \in A$.

Exercise 2.10. In light of the definition above...

1. Prove that $g \circ f$ is indeed a function.
2. Suppose in addition that $h : C \longrightarrow D$ is another function. Prove that there is an equality of functions $h \circ (g \circ f) = (h \circ g) \circ f$.

2.2 Isomorphisms

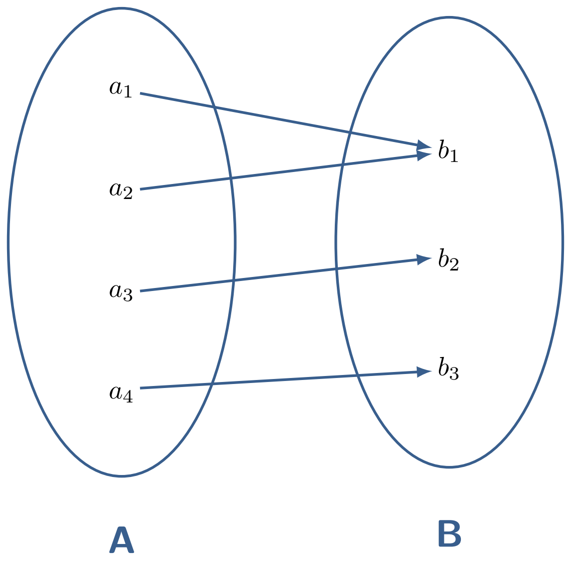
Definition 2.11. Let A, B be (possibly infinite) sets. A function $f : A \longrightarrow B$ is called:

1. **injective (monomorphism)** if for all $a \neq a' \in A$, we have $f(a) \neq f(a') \in B$.
2. **surjective (epimorphism)** if for every $b \in B$, there exists some $a \in A$ such that $f(a) = b$.
3. **bijective (isomorphism)** if it is injective and surjective.

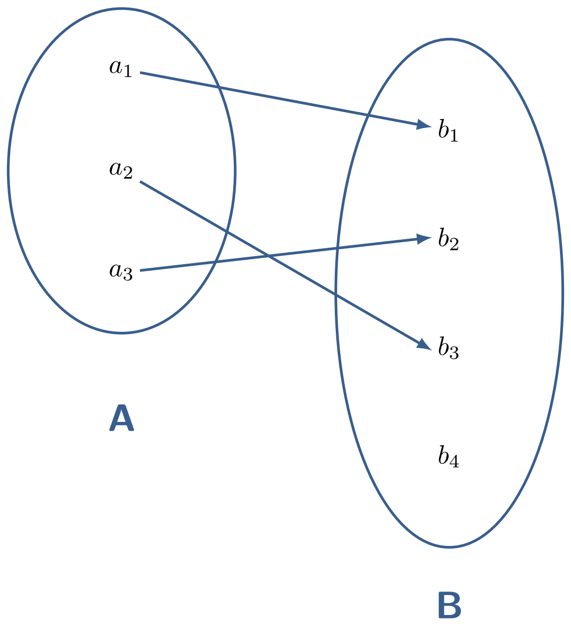
Remark 2.12. It also customary to call a monomorphism a 'one-to-one' function and an epimorphism an 'onto' function.

The intuition for Definition 2.11 often comes from the case of finite sets as the following examples demonstrate.

Example 2.13. The function below is surjective but not injective:



Example 2.14. The function below is injective but not surjective:



In light of the examples above, we can formulate the following

Proposition 2.15. *Let A, B be finite sets.*

1. *There exists an injective function $f : A \longrightarrow B$ iff $\#A \leq \#B$.*
2. *There exists a surjective function $f : A \longrightarrow B$ iff $\#A \geq \#B$.*

3. *There exists a bijective function $f : A \longrightarrow B$ iff $\#A = \#B$*

Proof.

1. Suppose $f : A \longrightarrow B$ is injective and enumerate A as $A = \{a_1, ..., a_n\}$. For $i = 1, ..., n$ denote $b_i = f(a_i)$. Since f is injective, the b_i's are all distinct so $\#A = n \leq \#B$. Conversely, if $\#A \leq \#B$ we can enumerate A and B as $A = \{a_1, ..., a_n\}$ and $B = \{b_1, ..., b_k\}$ where $n \leq k$. Define a function $f : A \longrightarrow B$ by $f(a_i) = b_i$. Then f is injective.
2. Suppose $f : A \longrightarrow B$ is surjective. Enumerate B as $B = \{b_1, ..., b_k\}$. Since f is surjective, for any $1 \leq i \leq k$ there is $a_i \in A$ such that $f(a_i) = b_i$. Thus $\#A \geq k = \#B$.
3. Immediate from (1) and (2).

□

Proposition 2.16. *Let A, B, C be sets and let $f : A \longrightarrow B$ and $g : B \longrightarrow C$ be two functions.*

1. *if f and g are injective, then so is $g \circ f$.*
2. *if f and g are surjective, then so is $g \circ f$.*
3. *if f and g are bijective, then so is $g \circ f$.*

Proof.

1. to show that $g \circ f : A \longrightarrow C$ is injective, let $a \neq a' \in A$. Since f is injective, $f(a) \neq f(a')$. Thus, since g is injective, $g(f(a)) \neq g(f(a'))$ and we are done.
2. to show that $g \circ f : A \longrightarrow C$ is surjective, let $c \in C$ be an arbitrary element. Since $g : B \longrightarrow C$ is surjective, there exists $b \in B$ such that $g(b) = c$. Since $f : A \longrightarrow B$ is surjective, there exists $a \in A$ such that $f(a) = b$. It follows that
$$(g \circ f)(a) = g(f(a)) = g(b) = c$$
and we are done.
3. this follows from the other two claims.

□

It is useful to notice that an epimorphism (a surjective functions) are "right cancallable", in the following sense:

Lemma 2.17. *Let A, B be (possibly infinite) sets and $f : A \longrightarrow B$ a surjective function. Then for any set C and functions $g, h\colon B \rightrightarrows C$ such that $g \circ f = h \circ f$, we have $g = h$.*

Proof. Let $b \in B$ be any element. Since f is surjective, there exists $a \in A$ such that $f(a) = b$. Therefore,
$$g(b) = g(f(a)) = h(f(a)) = h(b).$$
Since b was arbitrary, $g = h$. □

Exercise 2.18. Dually to Lemma 2.17, an injective function between two sets is "left cancallable". Formulate the precise statement that expresses this idea and prove it.

Let A be a set. Then the function $A \longrightarrow A$ given by $a \mapsto a$ is called the **identity function** of A and denoted as id_A. We have:

Theorem 2.19. *Let $f : A \longrightarrow B$ be a function. Then f bijective iff there is a (unique) function $f^{-1} : B \longrightarrow A$ such that $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$.*

Remark 2.20. The function f^{-1} in Theorem 2.19 is called the inverse of f.

Proof. Suppose f is bijective and let $b \in B$ be an arbitrary element. Since f is surjective, there exists $a \in A$ such that $f(a) = b$. We claim that there is a unique $a \in A$ with that property: otherwise, there would be $a' \neq a \in A$ such that $f(a) = b = f(a')$ which would contradict the assumption that f is injective. Since that a is unique, we define $f^{-1}(b) = a$ and we obtain a function $f^{-1} : B \longrightarrow A$. By the definition of f^{-1}, we see that for any $a \in A$,

$$(f^{-1} \circ f)(a) = f^{-1}(f(a)) = a$$

so that $f^{-1} \circ f = \mathrm{id}_A$. By that same definition, for any $b \in B$, $f^{-1}(b) = a$ where $a \in A$ is the unique element such that $f(a) = b$. Thus, for any $b \in B$,

$$(f \circ f^{-1})(b) = f(f^{-1}(b)) = f(a) = b$$

so that $f \circ f^{-1} = \mathrm{id}_B$. For uniqueness, use 2.17 and exercise 2.18.

Conversely, suppose there is $f^{-1} : B \longrightarrow A$ such that $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$. To show that f is injective, suppose that there are $a, a' \in A$ such that $f(a) = f(a')$. Then $f^{-1}(f(a)) = f^{-1}(f(a'))$ ie $(f^{-1} \circ f)(a) = (f^{-1} \circ f)(a')$ so that $a = a'$. To show that f is surjective, let $b \in B$ be an arbitrary element and denote $a := f^{-1}(b) \in A$. Then $f(a) = f(f^{-1}(b)) = (f \circ f^{-1})(b) = b$ so f is surjective.

□

Example 2.21.

1. Let $A = \{1, 2, 3\}$, $B = \{a, b, c\}$ and suppose $f : A \longrightarrow B$ is given by $f(1) = c$, $f(2) = b$, $f(3) = a$. Then f is an isomorphism and its inverse $f^{-1} : B \longrightarrow A$ is given by $f^{-1}(a) = 3$, $f^{-1}(b) = 2$, $f^{-1}(c) = 1$. It is easy to check that $f \circ f^{-1} = \mathrm{id}_B$ and $f^{-1} \circ f = \mathrm{id}_A$.

2. Let $A = [0, 1]$ and $B = [3, 5]$ and define $f : [0, 1] \longrightarrow [3, 5]$ by $f(x) = 2x + 3$. Then f is an isomorphism whose inverse $f^{-1} : [3, 5] \longrightarrow [0, 1]$ is given by $f^{-1}(y) = \frac{y-3}{2}$. Observe that for any $x \in [0, 1]$, $(f^{-1} \circ f)(x) = f^{-1}(f(x)) = f^{-1}(2x + 3) = \frac{(2x+3)-3}{2} = x$ so that $f^{-1} \circ f = \mathrm{id}_{[0,1]}$. Similarly $f \circ f^{-1} = \mathrm{id}_{[3,5]}$.

2.2.1 Aside: Cantor's theorem

We finish this section with an aside that will not be used later in the text but may be of independent interest.

Theorem 2.22 (Cantor)**.** *There is no surjective function $\mathbb{N} \longrightarrow \mathbb{R}$.*

Proof. Suppose by contradiction that there is a surjective map $f : \mathbb{N} \longrightarrow \mathbb{R}$. Let us view $\mathbb{R}$ as the set of all (infinite) decimal representations $x = x_1.x_2x_3...$ where $x_1 \in \mathbb{Z}$ and $\forall 1 < i$, $x_i \in \{0, ..., 9\}$. Set $f(n) = x^{(n)} \in \mathbb{R}$ and denote

$$x^{(n)} = x_1^{(n)} x_2^{(n)} x_3^{(n)} \dots$$

We organise the $x^{(n)}$ in a table as follows:

$$
\begin{aligned}
x^{(1)} &= x_1^{(1)}.x_2^{(1)}x_3^{(1)}\dots x_n^{(1)}\dots\\
x^{(2)} &= x_1^{(2)}.x_2^{(2)}x_3^{(2)}\dots x_n^{(2)}\dots\\
x^{(3)} &= x_1^{(3)}.x_2^{(3)}x_3^{(3)}\dots x_n^{(3)}\dots\\
&\vdots\\
x^{(n)} &= x_1^{(n)}.x_2^{(n)}x_3^{(n)}\dots x_n^{(n)}\dots\\
&\vdots
\end{aligned}
\tag{1}
$$

Define a new real number $y = y_1.y_2y_3\dots y_n\dots$ as follows. $y_1 = x_1^{(1)} + 1$ and for $1 < n$,

$$y_n := \begin{cases} x_n^{(n)} + 1 & \text{if } x_n^{(n)} \neq 9 \\ x_n^{(n)} - 1 & \text{if } x_n^{(n)} = 9 \end{cases}$$

Observe that $y \neq x^{(1)}$ since $y_1 \neq x_1^{(1)}$ and for all $1 < n$, $y \neq x^{(n)}$ because they differ in the nth digit. Thus, there is no $n \in \mathbb{N}$ such that $f(n) = y$ – contradiction. □

Example 2.23. Fix a programming language and consider the set $\overline{\mathcal{P}}$ of all computer programs in this language. Consider a subcollection $\mathcal{P} \subseteq \overline{\mathcal{P}}$ consisting of all programs that print some decimal representation x. We do not assume that a program in $\mathcal{P}$ necessarily stops so we can consider the "theoretical" output of a program $P \in \mathcal{P}$ to be the print output obtained by letting P run indefinitely. For example, the theoretical output of the program that prints "0." and then prints "3" in an indefinite loop is 0.333.. which is the decimal representation of 1/3. Similarly, one can construct a program whose theoretical output is the decimal representation of $\sqrt{2}$.

We claim that there is at least one number $x \in \mathbb{R}$ whose decimal representation can not be obtained as the theoretical output of any computer program. Suppose by contradiction that any $x \in \mathbb{R}$ can be obtained as the theoretical output of some program in $\mathcal{P}$. Then we would get a surjective function $f : \overline{\mathcal{P}} \longrightarrow \mathbb{R}$ by setting $f(P) := 0$ if $P \notin \mathcal{P}$ and for $P \in \mathcal{P}$ we set $f(P) := x$ where x is the theoretical output of P. On the other hand, we can treat all programs in $\overline{\mathcal{P}}$ as finite strings over a finite set of letters, and order them lexicographically: for example if the set of letters is the english alhpabet then the lexicographic order will be

$$"a", "b", ..., "z", "aa", "ab", ...$$

Using this lexicographic order we can define a surjective (in fact bijective) function $g : \mathbb{N} \longrightarrow \overline{\mathcal{P}}$ that assigns to $n \in \mathbb{N}$ the nth program in $\overline{\mathcal{P}}$ according to the lexicographic order defined above. Since the composition of surjective functions is surjective, we get a surjective map $f \circ g : \mathbb{N} \longrightarrow \mathbb{R}$, contradicting Cantor's theorem.

2.3 Equivalence Relations

We saw how one can construct the natural numbers $\mathbb{N}$ as sets. How does one go about defining, for example, the rational numbers as sets? One immediate problem is that a rational number doesn't have a unique representation. For example $\frac{1}{2} = \frac{2}{4} = \frac{3}{6} = ...$ Informally speaking, we would like to say that $\frac{1}{2}$ is 'equivalent' to $\frac{2}{4}$ and so on.

Definition 2.24. Let X be a set. A **relation** R on X is a subset $R \subseteq X \times X$. When $(x, y) \in R$ we denote $x \sim_R y$ or $x \sim y$ if R is understood from the context.

Example 2.25. A function $f : X \longrightarrow X$ is a relation on X: $x \sim y$ iff $y = f(x)$.

We are interested in a particular type of relations:

Definition 2.26. A relation R on X is called an **equivalence relation** if it satisfies the following properties:

1. Reflexive: for any $x \in X$, $x \sim x$.
2. Symmetric: for any $x, y \in X$, $x \sim y$ iff $y \sim x$.
3. Transitive: for any $x, y, z \in X$, if $x \sim y$ and $y \sim z$ then $x \sim z$.

Example 2.27. Let $X = \mathbb{Z}$ be the integers and let $n \in \mathbb{N}$ be a natural number. The relation $=\ (mod\, n)$ is an equivalence relation on $\mathbb{Z}$:

1. for any $x \in \mathbb{Z}$, $x = x\ (mod\, n)$.
2. for any $x, y \in \mathbb{Z}$, $x = y\ (mod\, n)$ iff $y = x\ (mod\, n)$.
3. for any $x, y, z \in \mathbb{Z}$, if $x = y\ (mod\, n)$ and $y = z\ (mod\, n)$ then $x = z\ (mod\, n)$

Construction 2.28. Let R be an equivalence relation on X. For $x \in X$ we denote $[x] := \{y \in X | x \sim y\}$ and call it the **equivalence class** of x. Note that if $a, b \in [x]$, then $a \sim x$ and $b \sim x$. By symmetry, $x \sim b$ and by transitivity $a \sim b$. All elements in $[x]$ are equivalent to each other. It follows that $[x] = [y]$ iff $x \sim y$. If $[x] \bigcap [y] \neq \varnothing$ and $z \in [x] \bigcap [y]$ then $z \sim x$ and $z \sim y$. By symmetry, $x \sim z$ and by transitivity $x \sim y$ so that $[x] = [y]$. In other words, two equivalence classes $[x], [y]$ are either equal or disjoint.

We can consider the collection of equivalence classes of elements of X, $Q = \{[x] | x \in X\}$. Note that this description of Q includes many repetitions since $[x] = [y]$ whenever $x \sim y$. Clearly, every x belongs to some set in Q, namely $[x]$. Thus, the collection Q forms a **partition** of X in that $\bigcup_{x \in X} [x] = X$ and each pair $[x] \neq [y]$ satisfies $[x] \bigcap [y] = \varnothing$.

We will denote

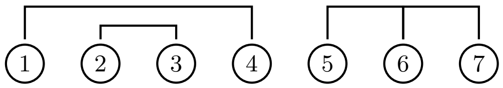
$$X/\sim := \{[x] | x \in X\}$$

and refer to it as the **quotient set** of X by R. Note that we always have a function $q : X \longrightarrow X/\sim$ given by $q(x) = [x]$. We call q the **quotient map**.

Conversely, suppose we have a partition of X, $\{U_i\}_{i \in I}$ (where I is an 'index' set), i.e. $X = \bigcup_{i \in I} U_i$ and for any $i, j \in I$ U_i, U_j are either equal or disjoint. Then we can define an equivalence relation on X by declaring $x \sim y$ if and only if $x, y \in U_i$ for some i. One readily verifies that this is indeed an equivalence relation on X.

Example 2.29. The picture below describes a partition of the set

$$X = \{1, 2, 3, 4, 5, 6, 7\} :$$



This partition corresponds to the equivalence relation given by

$$1 \sim 4,$$

$$2 \sim 3$$

and

$$5 \sim 6 \sim 7.$$

The quotient set is given by definition as

$$X/\backsim = \{[1],[2],[3],[4],[5],[6],[7]\}$$

and after omitting repetitions we get

$$X/\backsim = \{[1],[2],[5]\}.$$

note that we could also write, for example,

$$X/\backsim = \{[4],[3],[6]\}$$

since $[1] = [4],\ [2] = [3],\ [5] = [6]$

Example 2.30. Consider $X = \mathbb{Z}$ with the equivalence relation $= (mod\, n)$. The equivalence classes of the quotient set can be represented as

$$\mathbb{Z}/\backsim = \{[0],[1],...[n-1]\}$$

which is a set of size n. One typically denotes

$$\mathbb{Z}/\backsim := \mathbb{Z}_n = \{0,1,...,n-1\}.$$

Example 2.31. Earlier in this section we said that all mathematical objects can be described in terms of sets. Below Remark 2.1, we constructed the all natural numbers $\mathbb{N}$ from the empty set (and basic set operations) so one may wonder if the integers $\mathbb{Z}$ be constructed in a similar way. To this end, we define a relation on the Cartesian product $\mathbb{N} \times \mathbb{N} = \{(a,b)|a,b \in \mathbb{N}\}$ by setting $(a,b) \sim (a',b') \iff b + a' = b' + a$. The pair (a,b) will represent what we want to think of as "$b-a$" since $b+a' = b'+a \iff b-a = b'-a'$. The relation $\sim$ is an equivalence relation: clearly $(a,b) \sim (a,b)$ and $(a,b) \sim (b,a)$. For transitivity, suppose $(a,b) \sim (a',b')$ and $(a',b') \sim (a'',b'')$. Then $b + a'' = b'' + a$ (verify yourself). We may denote the quotient set as $\mathbb{Z} := \mathbb{N} \times \mathbb{N}/\sim$. This is the axiomatic definition of the integers. As an exercise, try to define + and × on $\mathbb{N}$ and then on $\mathbb{Z}$ in terms of set operations (union, intersection...).

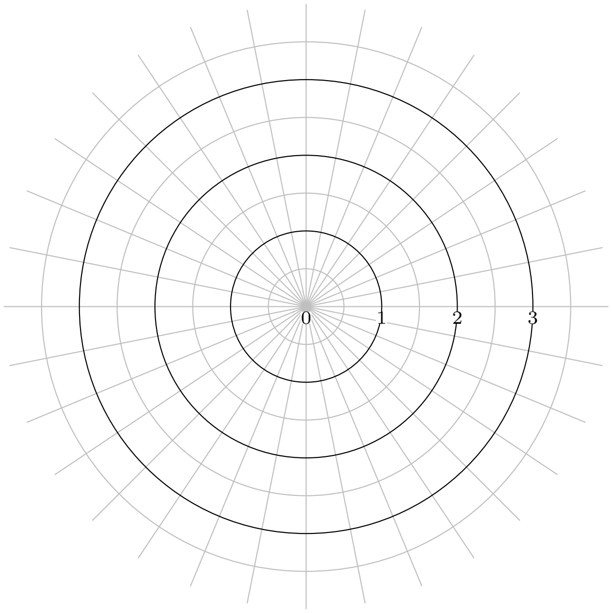
Exercise 2.32. Define an equivalence relation on $\mathbb{Z} \times \mathbb{Z}$ so that the quotient set can be identified with $\mathbb{Q}$.

Example 2.33. Let $I = [0,1]$ be the unit interval and define a relation $\sim$ by: $x \sim x$ for any $x \in I$ and $0 \sim 1$. One easily verifies that this is an equivalence relation with equivalence classes

$$I/\backsim = \{[x]|0 < x < 1\} \bigcup \{[0]\}.$$

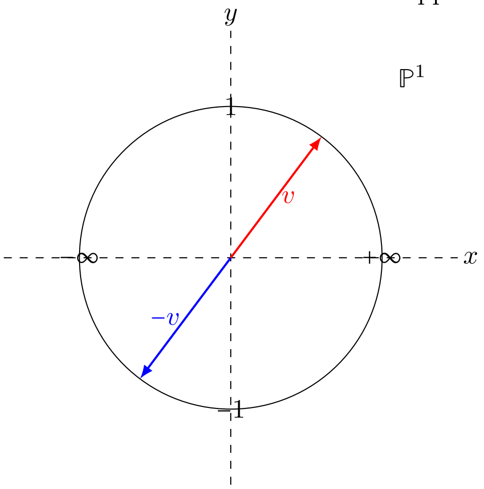
We can identify $I/\sim$ with the circle $S^1 \subseteq \mathbb{C}$ via the map $w : I/\backsim \longrightarrow S^1$ given by $w(x) = e^{2\pi i x}$ (note that $w(0) = w(1)$ so this map is well-defined).

Example 2.34. Define a relation on the plane $\mathbb{R}^2$ by setting $(x,y) \sim (z,w)$ iff $x^2 + y^2 = z^2 + w^2$. It is easy to see that this is an equivalence relation. For example, for any $(x,y) \in \mathbb{R}^2$ we have $(x,y) \sim (x,y)$ since $x^2 + y^2 = x^2 + y^2$ so the relation is reflexive. What are the equivalence classes of $\sim$? Let $P = (x_0, y_0)$ be an arbitrary fixed point and denote $r_0 = x_0^2 + y_0^2$. The equivalence class of P is the set of all points (x,y) such that $x^2 + y^2 = r_0$ namely the circle with centre at the origin $O = (0,0)$ and radius r_0. When $P = O = (0,0)$ we have $r_0 = 0$ and the equivalence class of P is the singeleton $\{(0,0)\}$ (that can be viewed as a circle with radius 0). Note that the collection of all these circles is a partition of the plane $\mathbb{R}^2$ as can be seen below:

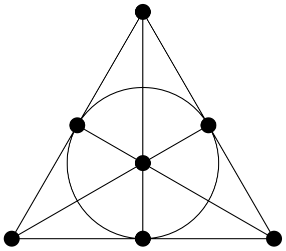


Exercise 2.35. Describe the quotient set of the equivalence relation in Example 2.34. Is it isomorphic to some known set?

Example 2.36. Define an equivalence relation on the punctured plane $\mathbb{R}^2 \smallsetminus \{(0,0)\}$ by setting $(x,y) \sim (x',y') \iff \exists \lambda \neq 0 : (x,y) = (\lambda x', \lambda y')$ We call the quotient set the **projective line** and denote it as $\mathbb{P}^1 \equiv \mathbb{P}^1(\mathbb{R})$ We can represent an equivalence class by a vector v of length 1 where the antipodal vector $-v$ is identified with it. Thus, $\mathbb{P}^1$ can be depicted as a unit circle in which the upper half is "glued" to the lower half (in the manner described above).



Example 2.37 (For recreational readers)**.** The construction of the projective space in the example above can make sense over any field, not just $\mathbb{R}$. When we take the field with 2 elements, the projective plane $\mathbb{P}^2(\mathbb{F}_2)$ is called the **Fanu plane** and can be modelled as drawn below. It is the minimal "Geometric model" with 7 points and 7 lines in which every line passes through exactly 3 points and any two distinct lines intersect exactly at one point.



3 Elementary number theory

Let $1 < n$ be an integer. If n is not prime, we can write $n = ab$ for positive integers a, b such that $a, b < n$. If a and/or b are not prime, we can further decompose them into such products and we end (after a finite number of steps) with a decomposition of the form $n = p_1^{e_1} \cdot ... \cdot p_2^{e_2} \cdot \ldots p_k^{e_k}$ where $p_1, ..., p_k$ are prime numbers and $e_1, ..., e_k \in \mathbb{N}$. This is called the **prime decomposition** of n and we note that it is unique up to a permutation of the p_i's (alternatively, it is unique if we assume in addition that $p_1 < ... < p_k$).

The following proposition is well-known but the proof we give here is useful since it can be adjusted to other cases.

Proposition 3.1. *The are infinitely many primes.*

Proof. Suppose that there are only finitely many primes and enumerate them as $p_1, ..., p_n$. We define $p_{n+1} = p_1 \cdot ... p_n + 1$. Then for every $1 \le i \le n$, $p_i \nmid p_{n+1}$ (as the remainder of dividing p_{n+1} by p_i is 1). Thus p_{n+1} is prime – contradiction. □

The **Euclidean algorithm** is a method for finding the greatest common divisor (GCD) of two possitive integers $b < a$. The algorithm proceeds in a series of steps, with the output of each step used as the input for the next. Track the steps using an integer counter i, so the initial step corresponds to $i = 0$, the next step to $i = 1$, and so on.

A step i begins with two nonnegative remainders $r_{i-2} > r_{i-1}$ and performs a division with remainder $r_{i-2} \% r_{i-1}$ to get non-negative integers q_i, r_i such that $r_i < r_{i-1}$ and

$$r_{i-2} = q_i r_{i-1} + r_i.$$

In the initial step $i = 0$, the remainders are set to $r_{-2} = a$ and $r_{-1} = b$, the numbers for which the GCD is sought. In the next step $i = 1$, the remainders are $r_{-1} = b$ and the remainder r_0 of the initial step, and so on. The algorithm proceeds in a sequence of equations

$$\begin{aligned} a &= q_0 b + r_0 \\ b &= q_1 r_0 + r_1 \\ r_0 &= q_2 r_1 + r_2 \\ r_1 &= q_3 r_2 + r_3 \\ &\vdots \\ r_{n-2} &= q_n r_{n-1} + r_n \\ &\vdots \end{aligned} \tag{2}$$

Since the remainders are non-negative integers that decrease with every step, the sequence

$$r_{-1} > r_0 > r_1 > r_2 > \cdots \ge 0$$

is finite, so the algorithm must eventually fail to produce the next step; but the division algorithm can always proceed to the $(N+1)$th step provided $r_N > 0$. Thus the algorithm must eventually produce a zero remainder $r_N = 0$. The final nonzero remainder is the greatest common divisor of a and b:

$$r_{N-1} = \gcd(a,b).$$

Proposition 3.2. *The Euclidean algorithm always computes $g = \gcd(a,b)$.*

Proof. In the notation of the discussion above, observe that

$$\gcd(a,b) = \gcd(b,r_0) = \gcd(r_0,r_1) = \cdots = \gcd(r_{N-2},r_{N-1}) = r_{N-1}$$

where the first equality follows from the first equation of 2 and so on. □

Construction 3.3. The **extended Euclidean algorithm** inputs two (possibly negative) integers a, b and outputs $x, y \in \mathbb{Z}$, called **Bezout coefficients**, such that $ax + by = \gcd(a,b)$. It proceeds similarly to the Euclidean algorithm, but adds two other sequences, as follows

$$\begin{aligned} r_0 &= a & r_1 &= b \\ s_0 &= 1 & s_1 &= 0 \\ t_0 &= 0 & t_1 &= 1 \\ &\vdots & &\vdots \\ r_{n+1} &= r_{n-1} - q_n r_n & &\text{and } 0 \le r_{n+1} < |r_n| \quad \text{(this defines } q_n\text{)} \\ s_{n+1} &= s_{n-1} - q_n s_n \\ t_{n+1} &= t_{n-1} - q_n t_n \\ &\vdots \end{aligned}$$

The computation also stops when $r_{N+1} = 0$ and gives r_N is the greatest common divisor of the input $a = r_0$ and $b = r_1$. The Bézout coefficients are s_N and t_N , that is $\gcd(a,b) = r_N = as_N + bt_N$ The quotients of a and b by their greatest common divisor are given by

$$s_{N+1} = \pm \frac{b}{\gcd(a,b)}$$

and

$$t_{N+1} = \pm \frac{a}{\gcd(a,b)}$$

Exercise 3.4. Adjust the argument of Proposition 3.2 to prove that the extended Euclidean algorithm indeed outputs the Bezout coefficient.

Example 3.5. Below is a Sage code for the (extended) euclidean algorithm and also the one-line commands

```
def euclid( a, b ):
    r = a%b
    while r != 0:
        a = b; b = r
        r = a%b
    g = b
    return g

g = gcd(a,b)


def extended_euclid( a , b ):
    if a == 0:
        return (b , 0 , 1) if b >= 0 else ( -b , 0 , -1)
    else:
        g, x, y = extended_euclid(b%a, a)
    return (g, y - (b//a) * x, x)


g, x, y = xgcd(a,b)
```

4 Groups

One of the most fundamental objects in Mathematics is a group. The notion of a group goes back to Galois, who studied solutions to polynomial equations and used groups in order to develop **Galois Theory**.

Definition 4.1. A **group** consists of the data of a set G with a chosen element $e \in G$, called the **neutral** (or unit) element together with a binary operation (i.e. a function) $\mu : G \times G \longrightarrow G$ satisfying the following conditions:

1. (unitary) For every $x \in G$, $\mu(x,e) = \mu(e,x) = x$.

2. (associativity) for every $x,y,z \in G$, $\mu(\mu(x,y),z) = \mu(x,\mu(y,z))$.

3. (invertability) for every $x \in G$, there exists an element $x^{-1} \in G$ (called the inverse of x) such that $\mu(x,x^{-1}) = \mu(x^{-1},x) = e$

Remark 4.2. As a meta principle, the definition of a notion in math can be divided to two parts:

1. Data – given by a collection of sets. E.g: G, $e \in G$, $\mu : G \times G \longrightarrow G$ (recall that a function is also a set).

2. Conditions – given by logical statements that describe the desired properties of the data. E.g $\forall x,y,z \in G$, $\mu(\mu(x,y),z) = \mu(x,\mu(y,z))$.

We will focus mostly on abelian groups, ie ones in which $\mu(x,y) = \mu(y,x)$.

Notation 4.3. The binary operation μ often comes from familiar multiplication or addition operations. In these cases we will adopt two notational conventions:

1. (multiplicative) $\mu(x,y)$ is replaced by $x \cdot y$ (or even xy), e is replaced by 1 and x^{-1} is left as it is. E.g $(\mathbb{R} \setminus \{0\}, \cdot, 1)$ the group of non-zero real numbers wrt multiplication. Here the inverse of an element $x \in \mathbb{R} \setminus \{0\}$ is denoted as x^{-1}.

2. (additive) $\mu(x,y)$ is replaced by $x + y$, e is replaced by 0 and x^{-1} is replaced by $-x$. E.g $(\mathbb{Z}, +, 0,)$ the group of integers wrt addition. Here the inverse of $n \in \mathbb{Z}$ is denoted as $-n$.

If there is no particular description of the operation in G, we will typically employ the multiplicative notation but keeping the neutral element as e, as it is a general convention in math for (not necessarily abelian) groups. Under this notation, the conditions of Definition 4.1 read:

1. For every $x \in G$, $x \cdot e = e \cdot x = x$.
2. For every $x, y, z \in G$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.
3. For every $x \in G$, there exists an element $x^{-1} \in G$ such that $x \cdot x^{-1} = e = x^{-1} \cdot x$

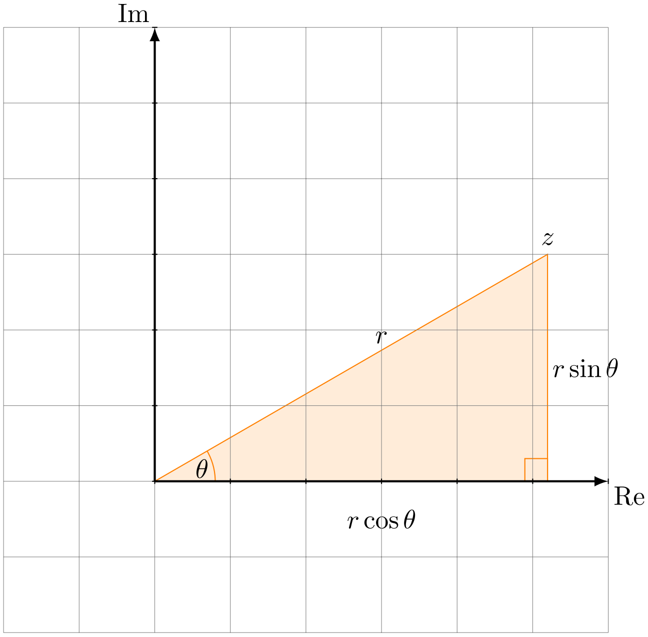
Observation 4.4. Let G be a group. The neutral element e is unique: if there exists an element $e' \in G$ such that for every $x \in G$, $x \cdot e' = e' \cdot x = x$ then $e' \cdot e = e$ since e is neutral but also $e' \cdot e = e'$ since e' is neutral. Thus $e = e'$.

Exercise 4.5. In a similar fashion, prove that

1. for any $x \in G$, the inverse x^{-1} is unique: if there exists $\tilde{x}^{-1} \in G$ such that $x \cdot \tilde{x}^{-1} = e = \tilde{x}^{-1} \cdot x$ then $x^{-1} = \tilde{x}^{-1}$.
2. For $x \in G$, if there is $y \in G$ such that $xy = e$ then $y = x^{-1}$. In other words, a one-sided inverse for x is automatically a two-sided inverse.

Examples 4.6. [Famous groups]

1. Choose an object e and let $G = \{e\}$. Then G has an obvious group structure and is called the **trivial group**.
2. The **integers** with addition, $(\mathbb{Z}, +, 0)$ is a group. The inverse of $x \in \mathbb{Z}$ is $-x$. Note that the natural numbers with addition $(\mathbb{N}, +, 0)$ is **not** a group since the inverse axiom is not satisfied.
3. The integers modulo n, $(\mathbb{Z}_n = \{0, ..., n-1\}, + \ (mod\ n), 0)$, is a group: the inverse of $a \in \mathbb{Z}_n$ is $n - a$ since $a + (n - a) = n = 0(mod\, n)$. The operation $+ \ (mod\, n)$ is associative: if $a, b, c \in \mathbb{Z}_n$ then
$$[(a+b)(mod\, n) + c](mod\, n) = [a + (b+c)(mod\, n)](mod\, n)$$
4. Let $\mathbb{C} = \{x + iy | x, y \in \mathbb{R}\}$ be the complex numbers. For $z = x + iy$ and $z' = x' + iy'$ in $\mathbb{C}$ the multiplication zz' is defined as
$$zz' := (x+iy)(x'+iy') = (xx' - yy') + i(xy' + x'y) \in \mathbb{C}.$$
Recall the Polar representation of $z \in \mathbb{C}$ is given by $z = r(\cos\theta + i\sin\theta)$ where r is the radius and $\theta \in [0, 2\pi]$. If $z' = r'(\cos\theta' + i\sin\theta')$ then
$$zz' = rr'(\cos\theta + i\sin\theta)(\cos\theta' + i\sin\theta') =$$
$$rr'(\cos\theta\cos\theta' - \sin\theta\sin\theta' + i[\sin\theta\cos\theta' + \sin\theta'\cos\theta])$$
$$= rr'(\cos(\theta+\theta') + i\sin(\theta+\theta'))$$
where the last equality is obtained from trigonometric identities:
$$\cos(\theta+\theta') = \cos\theta\cos\theta' - \sin\theta\sin\theta',$$
$$\sin(\theta+\theta') = \sin\theta\cos\theta' + \sin\theta'\cos\theta.$$



The exponential form is

$$z = re^{i\theta} := r(\cos\theta + i\sin\theta).$$

If $z' = r'e^{i\theta'}$ then from the discussion above, multiplication of complex numbers takes the form $z \cdot z' = r \cdot r'e^{i(\theta+\theta')}$. In light of this, we define the (complex) **unit circle** $S^1 \subseteq \mathbb{C}$ as

$$S^1 = \{z = e^{i\theta} | \theta \in \mathbb{R}\}.$$

Note that the radius of an element of S^1 is $r = 1$ as expected. Furthermore, multiplication of complex numbers that belong to S^1 yields a complex number in S^1:

$$e^{i\theta} \cdot e^{i\theta'} = e^{i(\theta+\theta')} \in S^1.$$

The element $1 = 1 + 0i = e^{i\cdot 0}$ satisfies $e^{i\theta} \cdot 1 = e^{i\theta}$ and for $z = e^{i\theta}$ we can take $z^{=1} = e^{-i\theta}$ to have $z \cdot z^{-1} = 1$. This means that $(S^1, \cdot, 1)$ is a group. Unlike the groups previously described, S^1 carries a **geometric structure** as well.

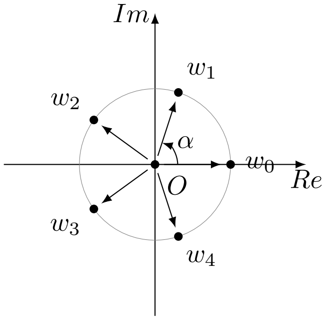
5. Let n be a natural number and consider the polynomial $p(x) = x^n - 1$. When we look for zeros of $p(x)$ over the complex numbers, i.e. $z \in \mathbb{C}$ such that $p(z) = 0$ we find that there are exactly n distinct such numbers, given by

$$w_k = e^{\frac{2k\pi i}{n}}$$

for $k = 0, 1, ..., n - 1$. The set

$$\mu_n(\mathbb{C}) = \{w_k\}_{k=0}^{n-1}$$

is called the n**th roots of unit**. When $n = 5$, for example, these roots can be depicted as follows:



If $w, w' \in \mu_n(\mathbb{C})$ then $(ww')^n = w^n w'^n = 1 \cdot 1 = 1$ so that ww' is a zero of $p(x) = x^n - 1$ hence belongs to $\mu_n(\mathbb{C})$. Furthermore, for $k \neq 0$ and $w_k = e^{\frac{2k\pi i}{n}} \in \mu_n(\mathbb{C})$, we have

$$w_{n-k} = e^{\frac{2(n-k)\pi i}{n}}$$

and thus

$$w_k w_{n-k} = e^{\frac{i(2k\pi + 2(n-k)\pi)}{n}} = e^{i2\pi} = 1.$$

It follows that $w_{n-k} = w_k^{-1}$ and so $\mu_n(\mathbb{C})$ is a group under complex multiplication.

The groups in the examples above satisfy the property that the operation $\cdot$ is commutative, i.e. that for any x, y, $x \cdot y = y \cdot x$. Such groups are called **abelian**.

Let us see an example of a non-abelian group:

Example 4.7. Let $[n] = \{1, 2, ..., n\}$ be a set of size n. Consider the collection of all functions $\sigma : [n] \longrightarrow [n]$ that are bijective (i.e. injective and surjective). We denote

$$\mathbb{S}_n = \{\sigma : [n] \longrightarrow [n] | \sigma \; is \text{ bijective}\}.$$

We define a binary operation on $\mathbb{S}_n$ by assigning for $\sigma, \tau \in \mathbb{S}_n$ their composition $\mu(\sigma, \tau) := \sigma \circ \tau$. The neutral element is the identity function $\mathrm{id}_{[n]} : [n] \longrightarrow [n]$. Associativity axiom follows from associativity of composition, and the inverse axiom holds by Theorem 2.19 that says that a bijective map admits an inverse map. The group $\mathbb{S}_n$ is called the **symmetric group** (or the **permutation group**) on n letters. Note that since composition of function is not commutative, the binary operation on $\mathbb{S}_n$ is not commutative as well. We say that $\mathbb{S}_n$ is a **non-abelian group**.

Let $n = 3$ and consider the symmetric group $\mathbb{S}_3$. We can represent an element $\sigma \in \mathbb{S}_n$ in the form

$$\sigma = \left(\begin{smallmatrix} 1 & 2 & 3 \\ \sigma(1) & \sigma(2) & \sigma(3) \end{smallmatrix}\right)$$

which means that 1 is mapped to $\sigma(1)$, 2 is mapped to $\sigma(2)$ and 3 is mapped to $\sigma(3)$.

Suppose $\sigma \in \mathbb{S}_3$ is given by $1 \mapsto 2$, $2 \mapsto 3$, $3 \mapsto 1$ and $\tau \in \mathbb{S}_3$ is given by $1 \mapsto 3$, $2 \mapsto 1$, $3 \mapsto 2$. Then we can depict $\sigma \cdot \tau$ as

$$\sigma \cdot \tau = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{smallmatrix}\right) = \mathrm{id}_{[3]} = e$$

Exercise 4.8. Find a pair of elements $\alpha, \beta \in \mathbb{S}_3$ such that $\alpha \cdot \beta \neq \beta \cdot \alpha$ and prove your claim in a similar way to the calculation in Example 4.7.

For the next example, we need to invoke a

Theorem 4.9 (Fermat's little theorem)**.** *If p is prime and $1 \leq a \leq p-1$ then $a^{p-1} = 1 \; (mod \, p)$.*

whose proof relies on a

Lemma 4.10. *For any $x, y \in \mathbb{F}_p$ and any n,*

$$(x+y)^{p^n} = x^{p^n} + y^{p^n} \ (mod\, p).$$

Proof. To see the claim for $n = 1$ we use the binom formula

$$(x+y)^p = \sum_{i=0}^{p} \binom{p}{i} x^i y^{p-i}.$$

Observe that for every $0 < i < p$, p divides $\binom{p}{i} = \frac{p!}{i!(p-i)!}$: on one hand $p \mid p!$ and on the other hand, since p is prime and $i!$ and $(p-i)!$ are both product of numbers smaller than p, we get that $p \nmid i!$ and $p \nmid (p-i)!$. Thus, for every $0 < i < p$,

$$\binom{p}{i} x^i y^{p-i} = 0 \ (mod\, p)$$

so we are left with $(x+y)^p = x^p + y^p \ (mod\, p)$. If $n = 2$ then $(x+y)^{p^2} = ((x+y)^p)^p = (x^p + y^p)^p = x^{p^2} + y^{p^2}$ by a repeated application of the $n = 1$ case. The proof continues by straightforward induction.

□

Proof of Theorem 4.9. We will prove an equivalent statement, that for any $a \in \mathbb{F}_p^{\times}$, $a^p = a \ (mod\ p)$ by induction on a. If $a = 1$ the statement is clear. Suppose the statement is true for a and consider the case of $a+1$. Then $(a+1)^p = a^p + 1^p = a + 1 \ (mod\, p)$ by Lemma 4.10 and the induction hypothesis so we are done. □

Example 4.11. Let p be a prime and consider

$$\mathbb{F}_p^{\times} = \{1, 2, ..., p-1\} = \mathbb{F}_p \setminus \{0\}$$

with the operation $\times \ (mod\ \ p)$ and unit element 1. Then $\mathbb{F}_p^{\times}$ is a group. Unitary and associativity axioms clearly hold. To prove the inverse axiom, we revoke Theorem 4.9 to get

$$a(a^{p-2}) = 1 \ (mod\ \ p)$$

so that

$$a^{-1} = a^{p-2}.$$

It will be useful for us to have a way of building new groups from old ones.

Construction 4.12. Let $G = (G, \cdot_G, e_G)$ and $H = (H, \cdot_H, e_H)$ be two groups and consider the Cartesian product

$$G \times H = \{(x, y) | x \in G \wedge y \in H\}.$$

Define a binary operation on $G \times H$ as follows: if $(x, y), (x', y') \in G \times H$,

$$(x, y) \cdot (x', y') := (x \cdot_G x', y \cdot_H y').$$

Proposition 4.13. *Under the binary operation defined above, $G \times H$ is a group with unit element (e_G, e_H) and the inverse for $(x, y) \in G \times H$ is given by (x^{-1}, y^{-1}).*

Proof. Left as an exercise. □

4.1 Subgroups and quotient groups

The last two examples in Examples 4.6 of the groups S^1 and μ_n suggest an interesting phenomena. We have $\mu_n \subseteq S^1$ and the group operation in both is given by multiplication of complex numbers.

Definition 4.14. Let $G = (G, \cdot, e)$ be a group and $H \subseteq G$ a subset. We say that H is a **subgroup** of G if $(H, \cdot, e)$ is a group under the restricted binary operation and the unit element of G. In that case, we denote $H \leq G$.

Equivalently, a subset $H \subseteq G$ of a group G is a subgroup if the following conditions hold:

1. $e \in H$.

2. for any $x, y \in H$, $xy \in H$.

3. for any $x \in H$, $x^{-1} \in H$.

Examples 4.15.

1. Let $G = (\mathbb{Z}, +, 0)$ and $H = n\mathbb{Z} = \{na | a \in \mathbb{Z}\}$ (eg $2\mathbb{Z}$ is the set of even numbers). Then $H \leq G$. Note that the set of odd numbers is not a subgroup of G since it does not contain 0 but also since it is not closed under addition.

2. Let $G = \mathbb{Z}_{10}$. The set $H = \{0, 5\}$ is a subgroup of G. Note that the set $H' = \{0, 4\}$ is not a subgroup of G since $4 + 4 = 8 \ (mod\ 10)$ and $8 \notin H'$.

3. As mentioned, for any natural number n, the group of nth roots of unity μ_n is a subgroup of the unit circle S^1.

4. Let $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$ be the multiplicative group of the field of complex numbers. Then S^1 is a subgroup of $\mathbb{C}^\times$. Since μ_n is a subgroup of S^1, it follows that $\mu_n \leq \mathbb{C}^\times$.

5. The group $\mathbb{Z}_n = \{0, 1, ..., n-1\}$ can be viewed as a subset of the group $\mathbb{Z}$ but $\mathbb{Z}_n \nleq \mathbb{Z}$. The reason is that the binary operation in $\mathbb{Z}$ is $+$ whereas in $\mathbb{Z}_n$ it is $+ \ (mod\ n)$. Thus, for example we have $n - 1 \in \mathbb{Z}_n \cap \mathbb{Z}$ but $(n-1) + (n-1) = 2n - 2 \notin \mathbb{Z}_n$.

6. Let $(G, e_G), (H, e_H)$ be groups. The Cartesian product $G \times H$ is a group by Proposition 4.13 where the multiplication is done coordinate-wise. The subsets $G \times \{e_H\} = \{(g, e_H) | g \in G\}$ and $\{e_G\} \times H = \{(e_G, h) | h \in H\}$ are both subgroups of $G \times H$. Note that $G \times \{e_H\}$ can be identified with G and $\{e_g\} \times H$ can be identified with H.

Let us perform a central

Construction 4.16. Let G be a group group and $H \leq G$ a subgroup. For an element $x \in G$, we define the (left) **H-coset**

$$xH := \{xh | h \in H\}.$$

Clearly, $eH = H$ but moreover, if $x \in H$ then $xH = H$: if $xh \in xH$ then since $x, h \in H$ and H is closed under multiplication, $xh \in H$ so that $xH \subseteq H$. Conversely, if $h \in H$ then since $x \in H$ and H is a subgroup, $x^{-1}h \in H$. Thus, $h = x(x^{-1}h) \in xH$ so that $H \subseteq xH$ and we have $H = xH$.

Example 4.17. Let $G = \mathbb{Z}$ and $H = n\mathbb{Z}$. Since we use additive notation in $\mathbb{Z}$, we write a coset as $x + n\mathbb{Z} = \{x + na | a \in \mathbb{Z}\}$. Note that $1 + n\mathbb{Z} = (n+1) + n\mathbb{Z} = (2n+1) + n\mathbb{Z} = ... = (-n+1) + n\mathbb{Z} = ...$ and similarly $2 + n\mathbb{Z} = (n+2) + n\mathbb{Z} = ...$ In other words, $x + n\mathbb{Z} = y + n\mathbb{Z}$ iff $x = y \ (mod\ n)$. Thus the collection of all cosets $\{x + n\mathbb{Z}\}_{x \in G}$ (after removing repetitions) can be written as

$$\{0 + n\mathbb{Z}, 1 + n\mathbb{Z}, ..., (n-1) + n\mathbb{Z}\}.$$

Proposition 4.18. *For a group G and a subgroup $H \leq G$, the set of cosets $\{xH\}_{x \in G}$ form a partition of G, i.e.,*

$$G = \bigcup_{x \in G} xH$$

and for any $x, y \in G$, the cosets xH, yH are either equal or disjoint.

Proof. First, for any $x \in G$ we have $x = x \cdot e \in xH$ since $e \in H$. Thus, $G = \bigcup_{x \in G} xH$. Second, let $x, y \in G$ and suppose that xH, yH are not disjoint. Then there is

$$\alpha \in xH \cap yH$$

ie there are $h, h' \in H$ such that $\alpha = xh = yh'$. Then

$$x = yh'h^{-1} \in yH$$

and

$$y = xh(h')^{-1} \in xH.$$

If $a = xh'' \in xH$ then

$$a = yh'h^{-1}h'' \in yH$$

and if $b = yh'''$ then

$$b = xh(h')^{-1}h'''$$

so $b \in xH$. It follows that $xH = yH$. □

Lemma 4.19. *For $H \leq G$ as above, and any $x, y \in G$, we have $xH = yH$ iff $x^{-1}y \in H$ (iff $y^-1x \in H$).*

Proof. Suppose $xH = yH$. Since $y = y \cdot e \in yH$ and $yH \subseteq xH$, there is $h \in H$ such that $y = xh$, so that $x^{-1}y = h \in H$. Conversely, if $x^{-1}y \in H$ then $y = y \cdot e \in yH$ but also $y = x(x^{-1}y) \in xH$ and since xH, yH are either equal or disjoint by Proposition 4.18 we deduce that $xH = yH$. □

Note that if $\alpha \in xH$ then $\alpha = xh$ for a unique $h \in H$: if also $\alpha = xh'$ then $xh = xh'$ and by multiplying both sides with x^{-1} on the left we get $h = h'$.

Suppose $H \leq G$ is a subgroup and let $x, y \in G$ be such that $xH \neq yH$. Define a function $f : xH \longrightarrow yH$ by $f(xh) = yh$. Clearly, f is surjective since every element $yh \in yH$ satisfies $f(xh) = yh$. The function f is also a injective: if $xh, xh' \in xH$ and $f(xh) = f(xh')$ then by definition $yh = yh'$ and multiplying the last equation by y^{-1} we get $h = h'$ so that $xh = xh'$.

Now suppose G is finite. Then for each $x \in G$, xH is a finite set and the discussion above means that for any $x, y \in G$, $\#xH = \#yH = \#H$. Thus, the set G can be partitioned to sets of equal size $G = \bigcup_{x \in G} xH$. We get the following

Theorem 4.20 (Lagrange)**.** *Let G be a finite group and $H \leq G$ a subgroup. Then $\#H \mid \#G$.*

Proof. We have a partition $G = \bigcup_{x \in G} xH$ where each of the partition sets xH has equal size $\#xH = \#H$ so $\#G = k(\#H)$ where k is the number of distinct cosets xH. □

Notation 4.21. For a group G and $H \leq G$ we denote the collection of all H-cosets by $G/H := \{xH | x \in G\}$ and refer to it as the **quotient** of G by H. Note that this description may include repetitions: if $x, y \in G$ are such that $x^{-1}y \in H$ then $xH = yH$.

Corollary 4.22. *For a group G and a subgroup $H \leq G$, we have an equivalence relation on G defined by $x \sim y$ iff $x^{-1}y \in H$ or equivalently $x \sim y$ iff $\exists z \in G$ such that $x, y \in zH$. The quotient set is G/H.*

We saw in Corollary 4.22 that for any group G and a subgroup $H \leq G$, we have a quotient set G/H. It is natural to ask: when is G/H a group?

We can ask when does the quotient G/H is again (naturally) a group. A sufficient condition is that G is abelian.

Proposition 4.23. *Suppose G is an **abelian** group and $H \leq G$ a subgroup. The quotient $G/H = \{xH | x \in G\}$ is a group under the operation $(xH) \cdot (yH) := (xy)H$ with unit $e_{G/H} := e \cdot H = H$ and the inverse of xH is given by $x^{-1}H$.*

Proof. Note that it could be that $xH = x'H$ with $x \neq x' \in G$ so we need to check that multiplication in G/H is well-defined (i.e. does not depend on representatives). So suppose $xH = x'H$ and $yH = y'H$. By Lemma 4.19 we have $xx'^{-1}, yy'^{-1} \in H$. In order to show that multiplication is well-defined, we need to show that $(xy)H = (x'y')H$ which by Lemma 4.19 is equivalent to show that $(xy)(x'y')^{-1} \in H$. But we have

$$(xy)(x'y')^{-1} = xy(x')^{-1}(y')^{-1} = xx'^{-1} \cdot yy'^{-1} \in H$$

since H is abelian. Associativity, unitary and existence of inverse in G/H follow from the corresponding axioms in G. □

Definition 4.24. For an abelian group G and a subgroup $H \leq G$ we call G/H with the multiplication defined above the **quotient group**. The map $q : G \longrightarrow G/H$ given by $x \mapsto xH$ is (a group homomorphism and) called the **quotient map**.

Exercise 4.25. Let G be a finite abelian group and $H \leq G$ a subgroup. Then $\#(G/H) = \#G/\#H$.

Examples 4.26.

1. Suppose G is a group and take $H = G$ (a legitimate subgroup). Then $G/G = \{eG\}$ ie the trivial group.

2. Suppose $G = \mathbb{Z}$ and $H = n\mathbb{Z}$. Then we saw that

   $$G/H = \{0 + n\mathbb{Z}, 1 + n\mathbb{Z}, ..., (n-1) + n\mathbb{Z}\}.$$

   For the group structure (which we will write in additive notation), we have $(x + n\mathbb{Z}) + (y + n\mathbb{Z}) = (x + y) + n\mathbb{Z}$. However, we saw that $z + n\mathbb{Z} = [z \pmod n] + n\mathbb{Z}$ so in fact we get

   $$(x + n\mathbb{Z}) + (y + n\mathbb{Z}) = [(x + y) \pmod n] + n\mathbb{Z}.$$

   We can thus "identify" the quotient group G/H with $\mathbb{Z}_n$. We will later define precisely what it means to "identify" two groups.

3. Let $(\mathbb{R}, +, 0)$ be the additive group of the real numbers and $\mathbb{Z} \leq \mathbb{R}$ the subgroup of the integers. Since $\mathbb{R}$ is abelian, the quotient $\mathbb{R}/\mathbb{Z}$ is a group. We may choose a set of representatives to be

   $$\mathbb{R}/\mathbb{Z} = \{x + \mathbb{Z} | x \in [0,1)\}.$$

   If $x, y \in [0,1)$ then by definition $(x + \mathbb{Z}) + (y + \mathbb{Z}) = (x + y) + \mathbb{Z}$ and if $x + y \geq 1$ we can represent $(x + y) + \mathbb{Z}$ as $(x + y) - 1 + \mathbb{Z}$, ie as the coset corresponding to the fractional part of $x + y$. One can depict it geometrically as the unit interval $[0,1]$ with its edges "glued" together, ie as the circle S^1.

4.2 Homomorphisms

As will be the case with many of the future mathematical objects we encounter, a group is a set together with an additional structure, namely a binary operation which in turn is subject to some conditions (associativity, unitary, existence of inverses). In order to 'move' between groups we need a function between their underlying sets that respects that structure.

Definition 4.27. Let G, H be groups. A function $f : G \longrightarrow H$ is called a (group) **homomorphism** if for any $x, y \in G$, $f(xy) = f(x)f(y)$ (note that the multiplication of the left-hand side is that of G and the one on the right-hand side is that of H).

Remark 4.28. It follows from the definition that a homomorphism $f : G \longrightarrow H$ must satisfy $f(e) = e$ as we have for any $x \in G$: $f(x) = f(e \cdot x) = f(e)f(x)$ and multiplying this equation by $f(x)^{-1}$ we get the desired. Similarly, a homomorphism f automatically satisfies $f(x^{-1}) = f(x)^{-1}$ for any $x \in G$ since

$$e_H = f(e_G) = f(xx^{-1}) = f(x)f(x^{-1})$$

and multiplying the equation by $f(x)^{-1}$ on the right gives the desired result.

Example 4.29. Let $n \in \mathbb{N}$ and $f : \mathbb{Z} \longrightarrow \mathbb{Z}_n$ be defined by $f(x) = x \ (mod\, n)$. Then f is a homomorphism since for any $x, y \in \mathbb{Z}$ we have $f(x+y) = (x+y) \ (mod\, n) = f(x) + f(y) \ (mod \ n)$.

Example 4.30. Let n, m be positive integers and set $G = \mathbb{Z}_{nm}, H = \mathbb{Z}_n$. We claim that the map $f : \mathbb{Z}_{nm} \longrightarrow \mathbb{Z}_n$ given by $x \mapsto x(mod\, n)$ is a homomorphism. Suppose $x, y \in \mathbb{Z}_{nm}$ and write $x = nq_1 + r_1$, $y = nq_2 + r_2$ and $x + y = nq_3 + r_3$. Then

$$(nq_1 + r_1) + (nq_2 + r_2) = nq_3 + r_3$$

so

$$r_1 + r_2 = r_3 \ (mod\, n)$$

It follows that

$$f(x +_G y) = f(x) +_H f(y)$$

so f is a homomorphism.

Exercise 4.31. If we consider the map $\mathbb{Z}_{nm+1} \longrightarrow \mathbb{Z}_n$ given by the same formula, would it still be a homomrphism?

Observation 4.32. For a group G the identity function $\mathrm{id}_G : G \longrightarrow G$ is always a homomorphism. For groups G, H, we always have the **trivial** homomorphism $G \longrightarrow H$ given by $x \mapsto e_H$.

Example 4.33. There is no non-trivial homomorphism $f : \mathbb{Z}_n \longrightarrow \mathbb{Z}$: if there was, let $x \in \mathbb{Z}_n$ be such that $f(x) \neq 0 \in \mathbb{Z}$. In $\mathbb{Z}_n$ he n−fold sum

$$\underbrace{x + x + ... + x}_{n-times} =: nx$$

equals zero but since f is a homomorphism we have

$$0 = f(0) = f(x + x + ... + x) = f(x) + f(x) + ... + f(x)$$

which is an n−fold sum of non-zero elements – contradiction.

Example 4.34. Consider the map

$$f : \mathbb{Z}_{nm} \longrightarrow \mathbb{Z}_n \times \mathbb{Z}_m$$

given by $f(x) = (x\ mod\ n,\ y\ mod\ m)$. We saw that the maps $\mathbb{Z}_{nm} \longrightarrow \mathbb{Z}_n$ $x \mapsto x\ (mod\ n)$ and $\mathbb{Z}_{nm} \longrightarrow \mathbb{Z}_m$ $x \mapsto x\ (mod\, m)$ are group homomorphisms. Since the operation in $\mathbb{Z}_n \times \mathbb{Z}_m$ is coordinat-wise, we get that f is again a group homomrphism.

Observation 4.35. If $f : G \longrightarrow H$ and $g : H \longrightarrow K$ are group homomorphisms, then so is $g \circ f$: for $x, y \in G$, $(g \circ f)(xy) = g(f(xy)) = g(f(x)f(y)) = g(f(x))g((f(y)) = (g \circ f)(x)(g \circ f)(y)$. Dually, a homomorphism $f : G \longrightarrow H$ is surjective iff $\operatorname{Im} f = H$.

Example 4.36. Let G be an abelian group and $H \leq G$ a subgroup. We saw that $G/H = \{xH | x \in G\}$ is a group. The function $q : G \longrightarrow G/H$ given by $q(x) = xH$ is a group homomorphism and the function $\iota : H \longrightarrow G$ given by $\iota(x) = x$ is also a homomorphism. According to the observation above, $q \circ \iota$ is a homomorphism. What is its description?

Definition 4.37. A homomorphism of groups $f : G \longrightarrow H$ is called an **isomorphism** if f is a bijective function between the underlying sets of G and H. In the latter case we write $f : G \xrightarrow{\cong} H$ or simply $G \cong H$.

To get an intuition, suppose $f : G \longrightarrow H$ is an isomorphism and $G = \{x_1, x_2, x_3\}$ (say $x_1 = e_G$). Since f is bijective, we can write $H = \{f(x_1), f(x_2), f(x_3)\}$. The multiplication tables of G and H can be depicted as below:

$\bullet$	x_1	x_2	x_3
x_1	$x_1 \cdot x_1$	$x_1 \cdot x_2$	$x_1 \cdot x_3$
x_2	$x_2 \cdot x_1$	$x_2 \cdot x_2$	$x_2 \cdot x_3$
x_3	$x_3 \cdot x_1$	$x_3 \cdot x_2$	$x_3 \cdot x_3$

$\xrightarrow{f}$

$\bullet$	$f(x_1)$	$f(x_2)$	$f(x_3)$
$f(x_1)$	$f(x_1) \cdot f(x_1)$	$f(x_1) \cdot f(x_2)$	$f(x_1) \cdot f(x_3)$
$f(x_2)$	$f(x_2) \cdot f(x_1)$	$f(x_2) \cdot f(x_2)$	$f(x_2) \cdot f(x_3)$
$f(x_3)$	$f(x_3) \cdot f(x_1)$	$f(x_3) \cdot f(x_2)$	$f(x_3) \cdot f(x_3)$

Since f is a homomorphism, for every $1 \leq i, j \leq 3$, $f(x_i x_j) = f(x_i)f(x_j)$ so f identifies the two multiplication tables. Informally, we can say that G and H are "the same" up to a change of symbols, which is given by f. This is also true when G and H are infinite, though we cannot draw the multiplication table so easily.

Example 4.38. Recall that $\mathbb{Z}/n\mathbb{Z} = \{0 + n\mathbb{Z}, 1 + n\mathbb{Z}, ..., (n-1) + n\mathbb{Z}\}$. Define a map $f : \mathbb{Z}/n\mathbb{Z} \longrightarrow \mathbb{Z}_n$ by $f(k + n\mathbb{Z}) = k$. We calim that f is an isomorphism of groups. To see this, recall that in $\mathbb{Z}/n\mathbb{Z}$,

$$(k + n\mathbb{Z}) + (k' + n\mathbb{Z}) = (k + k')[mod\, n] + n\mathbb{Z}.$$

Thus,

$$f\big((k + n\mathbb{Z}) + (k' + n\mathbb{Z})\big) = f\big((k + k')[mod\, n] + n\mathbb{Z})\big) = (k + k')[mod\, n] = k\ (mod\, n) + k'\ (mod\, n) = f(k + n\mathbb{Z}) + f(k' + n\mathbb{Z})$$

so f is a homomorphism which is clearly bijective.

Example 4.39. Let $\mu_n(\mathbb{C}) = \{w_0, ..., w_{n-1}\}$ be the group of nth roots of unity and define a function $f : \mathbb{Z}_n \longrightarrow \mu_n$ by

$$f(k) = w_k = e^{\frac{2k\pi i}{n}}.$$

To see that this is a homomorphism let $k, k' \in \mathbb{Z}_n$ and write $k + k' = qn + r$. Then in $\mathbb{Z}_n$ we have $k + k' = r$ so that

$$f(k + k') = f(r) = e^{\frac{2r\pi i}{n}}$$

and on the other hand,

$$f(k)f(k') = e^{\frac{2k\pi i}{n}} e^{\frac{2k'\pi i}{n}} = e^{\frac{2(k+k')\pi i}{n}} = e^{2q\pi i} e^{\frac{2r\pi i}{n}} = e^{\frac{2r\pi i}{n}} = f(r) = f(k + k').$$

Note that f is clearly injective and surjective so it is an isomorphism.

Proposition 4.40. *If $f : G \longrightarrow H$ is an isomorphism of groups, the inverse function $f^{-1} : H \longrightarrow G$ is also an isomorphism of groups.*

Proof. Let us recall how f^{-1} is defined. For $y \in H$, there is a unique $x \in G$ such that $f(x) = y$ (since f is bijective). Then $f^{-1}(y) := x$. We know that f^{-1} is bijective, so it remains to show that it is a homomorphism. Suppose $y, y' \in H$ such that $f(x) = y$ and $f(x') = y'$ for some (uniquely determined) $x, x' \in G$. Since f is a homomorphism, we have $f(xx') = f(x)f(x') = yy'$. Thus,

$$f^{-1}(yy') = xx' = f^{-1}(y)f^{-1}(y')$$

as desired. □

Examples 4.41.

1. Let $H = \{0, 5\} \leq \mathbb{Z}_{10} = G$. We claim that there is an isomorphism $\mathbb{Z}_{10}/H \cong \mathbb{Z}_5$. To see this, observe that we can write the (left) cosets as $G/H = \{0+H, 1+H, ..., 4+H\}$ – if we take, for example, the coset $8+H$, then $8-3 = 5 \in H$ so that $3+H = 8+H$. In light of this we define $f : \mathbb{Z}_{10}/H \longrightarrow \mathbb{Z}_5$ by $f(x+H) = x$ for $x = 0, 1, ..., 5$. This is clearly a homomorphism and an isomorphism of the underlying sets, hence an isomorphism of groups.

2. Let G, H be groups. We saw that the Cartesian product $G \times H$ has a subgroup of the form $G \times \{e_H\}$. Clearly, $G \times \{e_H\} \cong G$. If G, H are abelian, we get that the quotient $G \times H/G \times \{e_H\}$ is isomorphic to H. Similarly, we have $G \times H/\{e_G\} \times H \cong G$.

Exercise 4.42. Let $(\mathbb{R}, +, 0)$ be the additive group of the real numbers and $\mathbb{Z} \leq \mathbb{R}$. Show that the groups $\mathbb{R}/\mathbb{Z}$ and $S^1 \subseteq \mathbb{C}$ are isomorphic. Prove your claims.

Definition 4.43. Let $f : G \longrightarrow H$ be a group homomorphism. The **kernel** of f is the set $\ker(f) = \{x \in G | f(x) = e_H\} \subseteq G$. The **image** of f is the set $\text{Im}(f) = \{f(x) | x \in G\} \subseteq H$.

Observation 4.44. A group homomorphism $f : G \longrightarrow H$ is injective iff $\ker f = \{e_G\}$: if f is injective then since $f(e_G) = e_H$ there cannot be any $x \neq e_G \in G$ such that $f(x) = e$. Conversely, if $\ker f = \{e_G\}$ and $x, x' \in G$ are such that $f(x) = f(x')$ then $f(x)^{-1}f(x') = e_H$ and since f is a homomorphism we get so $f(x^{-1}x') = e_H$. Thus, $x^{-1}x' \in \ker f$ and by assumption $x^{-1}x' = e_G$. Thus, $x = x'$ so that f is injective. Clearly, a group homomorphism $f : G \longrightarrow H$ is surjective iff $\text{Im} f = H$.

Example 4.45. Let $f : \mathbb{Z}_5 \longrightarrow \mathbb{Z}_{10}$ be $f(x) = 2x$. Then one readily checks that f is a homomorphism with $\ker f = \{0, 5\}$ and $\text{Im} f = \{0, 2, 4, 6, 8\}$.

Proposition 4.46. *For any group homomorphism $f : G \longrightarrow H$ we have*

1. $\ker(f) \leq G$.

2. *$Im(f) \leq H$.*

Proof.

1. Clearly, $e \in \ker(f)$. If $x, y \in \ker(f)$ then $f(xy) = f(x)f(y) = e$ so $xy \in \ker(f)$ and for any $x \in \ker(f)$, $f(x^{-1}) = f(x)^{-1} = e$ so $x^{-1} \in \ker(f)$.

2. Clearly, $e = f(e) \in \text{Im}(f)$. If $x, y \in \text{Im}(f)$ then there are $a, b \in G$ such that $x = f(a)$ and $y = f(b)$. Thus, $xy = f(a)f(b) = f(ab)$ so that $xy \in \text{Im}(f)$. Similarly, if $x = f(a) \in \text{Im}(f)$ then $x^{-1} = f(a^{-1})$ so that $x^{-1} \in \text{Im}(f)$.

□

Suppose $f : G \longrightarrow H$ is a homomorphism of abelian groups. By Proposition 4.46 we can consider the quotient groups $G/\ker(f)$.

Theorem 4.47 (The first isomorphism theorem)**.** *Let $f : G \longrightarrow H$ be a homomorphism of abelian groups and denote $K = \ker f$. Then there is an isomorphism of groups $G/K \cong Im\,(f)$.*

Proof. Define $f' : G/K \longrightarrow \mathrm{Im}\,(f)$ by $f'(xK) = f(x)$. To see that f' is well-defined, suppose $x, y \in G$ are such that $xK = yK$. Then $xy^{-1} \in K$ so that $f(xy^{-1}) = f(x)f(y)^{-1} = e$ and thus $f(x) = f(y)$ i.e. $f'(xK) = f'(yK)$ and f' is well-defined. Clearly, f' is a group homomorphism since f is and it remains to check that f' is an isomorphism. First, if $a = f(x) \in \mathrm{Im}\, f$ then $f'(xK) = f(x) = a$ so that f' is an surjective. Second, if xK, yK are such that $f'(xK) = f'(yK)$ then $f(x) = f(y)$ so that $f(xy^{-1}) = e$ and we get that $xy^{-1} \in K$ which by Lemma 4.19 means that $xK = yK$. Thus f' is also a injective and we're done. □

Remark 4.48. Theorem 4.47 is true also without the assumption that G, H are abelian. However, defining the quotient group for non-abelian groups is an additional complication that we wish to avoid.

Examples 4.49.

1. Let $f : \mathbb{Z} \longrightarrow \mathbb{Z}_n$ be the *mod* n homomorphism $x \mapsto x\ (mod\, n)$. Then clearly $\mathrm{Im}\, f = \mathbb{Z}_n$. On the other hand,

$$\ker f = \{na | a \in \mathbb{Z}\} = n\mathbb{Z}.$$

   It follows from Theorem 4.47 that $\mathbb{Z}_n \cong \mathbb{Z}/n\mathbb{Z}$.

2. Let $f : \mathbb{Z}_{10} \longrightarrow \mathbb{Z}_2$ be the homomorphism given by $f(x) = x\ (mod\, 2)$. Then $\mathrm{Im}\, f = \mathbb{Z}_2$ and $\ker f = \{0, 2, 4, 6, 8\}$. It follows that $\mathbb{Z}_{10}/\ker f \cong \mathbb{Z}_2$. Note that $\ker f \cong \mathbb{Z}_5$ via the isomorphism $\ker(f) \longrightarrow \mathbb{Z}_5$ given by $x \mapsto \frac{x}{2}$.

4.3 Cyclic groups

Much of our attention in cryptography will be devoted to finite (abelian) groups. In such a context, it's natural to consider the following

Definition 4.50. Let G be a group and $g \in G$ an element. The **order** of g, denoted $o(g)$, is the minimal natural number n such that $g^n = e$. If no such number exists, we set $o(g) = \infty$.

Let $G_1, ..., G_n$ be groups. We define the nth Cartesian product $G = G_1 \times ... \times G_n$ to be the set of all n-tuples $(g_1, ..., g_n)$ with $g_i \in G_i$ for all i. Just like with the two-fold Cartesian product, we can define a binary operation on G by using the operation of G_i in the ith coordinate and this makes G into a group. For the next result, recall that for two natural numbers m, n their **least common multiplier** $\ell = \mathrm{lcm}(m, n)$ is the minimal positive integer ℓ such that $n \mid \ell$ and $m \mid \ell$. By taking a prime decomposition we can write $m = p_1^{s_1} ... p_k^{s_k}$ and $n = p_1^{t_1} ... p_k^{t_k}$ where $s_i, t_i \geq 0$ (note that the same primes appear in both factorisations). Then

$$\mathrm{lcm}(m, n) = p_1^{\min(s_1, t_1)} ... p_k^{\min(s_k, t_k)}.$$

Theorem 4.51 (Cauchy)**.** *Let G be a finite abelian group and p a prime such that $p \mid \#G$. Then there is $g \in G$ of order p.*

Proof. Assume no element of G has order p and we will get a contradiction. No element has order divisible by p: if $g \in G$ has order r and $p \mid r$ then $g^{r/p}$ has order p. Let $G = \{g_1, g_2, ..., g_n\}$ and let g_i have order m_i, so each m_i is not divisible by p. Let $m = \mathrm{lcm}_{i=1,...,n}(m_i)$ be the least common multiple of the m_i's , so m is not divisible by p and $g_i^m = e$ for all i. For $n \in \mathbb{N}$ write $g^{-n} := (g^{-1})^n$. Then because G is abelian, the function

$$f : (\mathbb{Z}_m)^n \longrightarrow G$$

given by

$$f(a_1, ..., a_n) = g_1^{a_1} \cdot ... \cdot g_n^{a_n}$$

is a homomorphism:

$$f(a_1 + b_1, ..., a_n + b_n) = g_1^{a_1+b_1} \cdot ... \cdot g_n^{a_n+b_n} = g_1^{a_1} g_1^{b_1} \cdot ... \cdot g_n^{a_n} g_n^{b_n} = f(a_1, ..., a_n) f(b_1, ..., b_n).$$

This homomorphism is surjective since $f(0, ..., 1, ..., 0) = g_i$ so by the first isomorphism theorem 4.47,

$$G = \mathrm{Im}\,(f) \cong (\mathbb{Z}_m)^n / \ker(f).$$

By Lagrange theorem, $\#G \cdot \# \ker(f) = \#(\mathbb{Z}_m)^n$ so in particular, $\#G \mid m^n$. But $p \mid \#G$ and m^n is not divisible by p – contradiction. □

Remark 4.52. Cauchy theorem can be regarded as a partial converse to Lagrange thoerem and is true also for non-abelian groups. We ommited the proof here since we won't need it in the near future. Moreover, for abelian groups the converse Lagrange theorem holds entirely, ie for every number k that divides the order of the group G, there is a subgroup of G of order k.

Let G be a group an p a prime s.th $p \mid \#G$. By Cauchy theorem, there is an element $g \in G$ of order p. The elements $\{e, g, g^2, ..., g^{n-1}\}$ are all distinct since if there are $1 \le k < l \le n-1$ such that $g^l = g^k$ then $g^{l-k} = e$ with $l - k < n-1$ and so g cannot be of order n. The set $\{e = g^0, g, ..., g^{p-1}\} \subseteq G$ is in fact a subgroup: it is clearly closed under multiplication and $(g^k)^{-1} = g^{p-k}$. Thus we are guarantied to have a subgroup $H \le G$ of order p.

Example 4.53. $\mathbb{Z}_{10}$ has a subgroup of order 5. We identified such earlier – $\{0, 2, 4, 6, 8\}$.

The discussion above naturally leads to

Definition 4.54. Let G be a group and $g \in G$ an element.

1. The group **generated** by g is the set $\langle g \rangle := \{g^n | n \in \mathbb{Z}\}$ which inherits a group structure from G. Note that by definition, $g^{-n} := \left(g^{-1}\right)^n$.
2. A group G is called **cyclic** if there exists $g \in G$ such that $\langle g \rangle = G$.

Remark 4.55. If $g \in G$ has order n then $\langle g \rangle = \{e, g, ..., g^{n-1}\}$

Examples 4.56.

1. Let $n \in \mathbb{N}$. The group $\mathbb{Z}_n$ is cyclic: as a generator we can choose $g = 1$.
2. The group $\mathbb{Z}$ is cyclic. As a generator we can choose either 1 or −1.
3. The group $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic. As a generator we can choose $(1, 1)$.

Cyclic groups have a rather rigid structure. For example:

Observation 4.57. Let $G = \langle g \rangle$ be a cyclic group. Then G is abelian: $g^n g^k = g^{n+k} = g^k g^n$.

Proposition 4.58. *For any $n \in \mathbb{N}$, there exists a cyclic group of order n, and it is unique up to an isomorphism.*

Proof. As we saw, $\mathbb{Z}_n$ is a cyclic group of order n. Suppose $G = \langle e, g, g^2, ..., g^{n-1} \rangle$ is a cyclic group of order n. Then the function $f : \mathbb{Z}_n \longrightarrow G$ by $f(k) = g^k$ is an isomorphism of groups. □

As we saw in Example 4.56 (2), a cyclic group can have more than one generator. In order to account for all generators in $\mathbb{Z}_n$, we will need the following

Definition 4.59. For $n \in \mathbb{N}$, $\varphi(n) = \#\{k | 1 \leq k \leq n \wedge \gcd(n,k) = 1\}$ is the **Euler totient function**.

Proposition 4.60. *An element $k \in \mathbb{Z}_n$ is a generator iff* $\gcd(n,k) = 1$.

Proof. If $1 \leq k \leq n-1$ is a generator then

$$\mathbb{Z}_n = \{0, k, 2k, ..., (n-1)k\} \tag{3}$$

so if $\gamma = \gcd(n,k) > 1$ we can write $n = \gamma s$ and $k = \gamma t$. We get that $sk = s\gamma t = nt$ so $sk = 0 \, (mod\, n)$ and since $s < n-1$ we get a contradiction to 3. Conversely, if $\gcd(n,k) = 1$ we claim that the numbers $\{0, k, 2k, ..., (n-1)k\}$ are all distinct $(mod\, n)$: otherwise, if $sk = tk \ (mod \ n)$ for $1 \leq s < t \leq n-1$ then $(t-s)k = nx$ for some x and since $k \nmid n$ we have $k \mid x$. But then, $(t-s)k = nky$ for some y so $t - s = ny$ which is a contradiction since $1 \leq (t-s) \leq n-1$. □

Corollary 4.61. *Let G be a cyclic group of order n. Then G has exactly $\varphi(n)$ generators. In particular, if $n = p$ is a prime then any $e \neq x \in G$ is a generator.*

Proof. This is the case for $G = \mathbb{Z}_n$ and the number of generators stays the same for isomorphic groups. □

For future reference, it is useful to record:

Theorem 4.62. *For $n \in \mathbb{N}$, the Euler's totient function can be calculated as $\varphi(n) = n \prod_{p|n} (1 - \frac{1}{p})$ where the product is taken over all primes p that divide n. It follows that*

$$n = \sum_{d:d|n} \varphi(d).$$

The next result is useful in the study of cyclic groups

Proposition 4.63. *A subgroup of a finite cyclic group is cyclic.*

Proof. Suppose $G = \langle g \rangle$ is cyclic of order n and $H \leq G$ a subgroup. Denote $H = \{a_0, ..., a_{d-1}\} = \{1, g^{k_1}, ..., g^{k_{d-1}}\}$ where $d \mid n$ with $n = kd$. Since H is of order d, $\left(g^{k_i}\right)^d = g^{k_i d} = 1$ for each i so that $k_i d = nm_i$ for some m_i. But then $k_i d = kdm_i \iff k_i = km_i$. Thus, each a_i is a power of g^k so that $H \subseteq \{1, g^k, g^{2k}, ..., g^{k(d-1)}\}$ and it follows that $H = \{1, g^k, g^{2k}, ..., g^{k(d-1)}\}$ which is cyclic with generator g^k. □

Exercise 4.64.

1. Let G be a group of prime order. Prove that G is cyclic.

2. Show that $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic.

3. Let G, H be groups such that $\#G = \#H = 3$. Is necessarily $G \cong H$?

4. Suppose G, H are groups with $\#G = \#H = 4$. Is necessarily $G \cong H$?

Recall from Proposition 4.13 that given two groups G, H we have their Cartesian product $G \times H$ with coordinate-wise group structure. The next lemma demonstrates how to calculate the order of elements in the Cartesian product

Lemma 4.65. *Let G, H be groups. If $g \in G$ has order n and $h \in H$ has order m then $(g, h) \in G \times H$ has order $\operatorname{lcm}(n, m)$.*

Proof. Note that $(g, h)^N = (e, e) \iff g^N = e \wedge h^N = e \iff o(g)|N \wedge o(h)|N$ so N must be a common multiple of n and m. Thus, the order of (g, h) is the least common multiple of n and m. □

We showed in Exercise 4.64 that $\mathbb{Z}_2 \times \mathbb{Z}_3$ is cyclic. Lemma 4.65 enables us to prove the general case. First, we need a small result from elementary number theory:

Lemma 4.66. *Let n, m be positive integers. Then $\operatorname{lcm}(n, m) = \frac{nm}{\gcd(n,m)}$.*

Proof. Write

$$m = p_1^{k_1} ... p_r^{k_r}$$

$$n = p_1^{l_1} ... p_r^{l_r}$$

where $k_i, l_i \geq 0$ (so the decomposition is over the same set of primes) Observe that

$$\gcd(n, m) = p_1^{\max(k_1, l_1)} ... p_r^{\max(k_r, l_r)}$$

and

$$\operatorname{lcm}(n, m) = p_1^{\min(k_1, l_1)} ... p_r^{\min(k_r, l_r)}.$$

Thus

$$\operatorname{lcm}(n, m) \gcd(n, m) = nm$$

as desired. □

Corollary 4.67. *Let n, m be positive integers. Then $\mathbb{Z}_n \times \mathbb{Z}_m$ is cyclic iff $\gcd(n, m) = 1$. In that case*

$$\mathbb{Z}_n \times \mathbb{Z}_m \cong \mathbb{Z}_{nm}.$$

Proof. Suppose $\gcd(n, m) = 1$ By Lemma 4.65 the order of $(1, 1)$ is $\operatorname{lcm}(n, m)$ which by Lemma 4.66 is $\operatorname{lcm}(n, m) = \frac{nm}{\gcd(n,m)} = nm$. Since $\#\mathbb{Z}_n \times \mathbb{Z}_m = nm$ we get that $\mathbb{Z}_n \times \mathbb{Z}_m = \langle (1, 1) \rangle$ is cyclic.

Conversely, if $\gcd(n, m) > 1$ then the order of any element in $\mathbb{Z}_n$ divides n and the order of any element in $\mathbb{Z}_m$ divides m so that the order of any element in $\mathbb{Z}_n \times \mathbb{Z}_m$ divides $\operatorname{lcm}(n, m) = \frac{nm}{\gcd(nm)}$ which is strictly smaller than nm. Thus in that case $\mathbb{Z}_n \times \mathbb{Z}_m$ is not cyclic. □

Corollary 4.67 in turn leads to an insight on solving a system of equations modulo different integers:

Theorem 4.68. *Let $n, m \in \mathbb{N}$ such that $\gcd(n, m) = 1$. Then for all $r, s \in \mathbb{N}$, the system of equations*

$$\begin{aligned} x = r \ (mod\, n) \\ x = s \ (mod\, m) \end{aligned} \tag{4}$$

has a unique solution in $\mathbb{Z}_{nm}$.

Proof. Consider the map $f : \mathbb{Z}_{nm} \longrightarrow \mathbb{Z}_n \times \mathbb{Z}_m$ given by $a \mapsto (a \; mod \, n, a \; mod \, m)$. If $a, b \in \mathbb{Z}_{nm}$ then since $n \mid nm$ and $m \mid nm$, we have

$$(a + b) \; mod \, nm \mapsto (a \; mod \, n + b \; mod \, n, a \; mod \, m + b \; mod \, m)$$

so f is a group homomorphism. Clearly, $f(1) = (1, 1)$ and since $(1, 1)$ has order nm, f is an isomorphism. Let $[r]_n = r \; mod \; n \in \mathbb{Z}_n$ and $[s]_m = s \; mod \; m \in \mathbb{Z}_m$. Since f is an isomorphism, the element $([r]_n, [s]_m)$ has a unique $x \in \mathbb{Z}_{nm}$ such that $f(x) = (x \; mod \, n, x \; mod \, m) = ([r]_n, [s]_m)$. □

Theorem 4.68 has, in fact, a more general version:

Corollary 4.69 (The Chinese remainder Theorem)**.** *Let $n_1, ..., n_k \in \mathbb{N}$ be such that $\forall i \neq j : \gcd(n_i, n_j) = 1$ and denote $N = n_1 \cdot ... \cdot n_k$. Then for all $a_1, ..., a_k \in \mathbb{N}$, the system of equations*

$$\begin{aligned} x = a_1 \; (mod \, n_1) \\ \vdots \\ x = a_k \; (mod \, n_k) \end{aligned} \tag{5}$$

has a unique solution in $\mathbb{Z}_N$.

Proof. By induction using Theorem 4.68. □

What happens when we want to solve a system of integral equations with non-coprime moduli? We state here a result for the sake of completeness.

Theorem 4.70. *Let $n_1, ..., n_k \in \mathbb{N}$ and denote $N = n_1 \cdot ... \cdot n_k$. Then for all $a_1, ..., a_k \in \mathbb{N}$, the system of equations*

$$\begin{aligned} x = a_1 \; (mod \, n_1) \\ \vdots \\ x = a_k \; (mod \, n_k) \end{aligned} \tag{6}$$

has a solution in $\mathbb{Z}$ if $\forall i \neq j$, $a_i = a_j mod \; \gcd(n_i, n_j)$ and that solution is unique modulo $\operatorname{lcm}(n_1, ..., n_k)$.

4.4 Classification of finite abelian groups

In this section we will prove a classification/structure theorem for finite abelian groups. It will be convenient to employ additive notation $G = (G, +, 0)$ and correspondingly we will denote $G \bigoplus H := G \times H$.

Definition 4.71. For an abelian group G and a prime p, $G(p)$ is the set of all elements whose order is a power of p.

$$G(p) = \{a \in G | \; o(a) = p^n \text{ for some } n \geq 0\}.$$

Observation 4.72. $G(p)$ is a subgroup of G: if $a, b \in G(p)$ of order p^n, p^m ie $p^n a = 0$ and $p^m b = 0$, then $p^m(-b) = -p^m b = 0$ and hence $p^n p^m(a - b) = p^n p^m a - p^n p^m b = 0$. It is called a **primary p-subgroup** of G.

Lemma 4.73. *Let G be an abelian group and a an element of finite order. Then*

$$a = a_1 + ... + a_k$$

with $a_i \in G(p_i)$ where $p_1, ..., p_k$ are the distinct primes in the decomposition of $o(a)$.

Proof. We use induction on the number of distinct primes that divide elements of G of finite order. If $o(a)$ is divisible by at most one prime p_1 then $a \in G(p_1)$.

Now suppose for every element b whose finite order is divisible by at most $k-1$ distinct primes $p_1, p_2, ..., p_{k-1}$, we have $b = b_1 + b_2 + \; + b_{k-1}$ for $b_i \in G(p_i), \;\; i = 1, ..., k-1$.

If $o(a)$ is divisible by k distinct primes $p_1, p_2, ..., p_k$, then there are positive integers $r_1, r_2, ..., r_k$ such that

$$o(a) = p_1^{r_1} \cdot ... \cdot p_k^{r_k}.$$

Let $m = p_2^{r_2} ... p_k^{r_k}$ and $n = p_1^{r_1}$ so that $o(a) = nm$. Then $\gcd(m,n) = 1$ so by 3.3 there are integers u, v such that

$$mu + nv = 1 \iff a = mua + nva.$$

Because $o(a) = mn$ with $n = p_1^{r_1}$ we have $p_1^{r_1} mua = nmua = u(mna) = 0$, so that $mua \in G(p_1)$. Similarly, $m(nva) = v(mna) = v0 = 0$, so the order of nva divides $m = p_2^{r_2} ... p_k^{r_k}$, an integer with $k-1$ distinct prime divisors.

By the induction hypothesis, we have that $nva = a_2 + a_3 + ... + a_k$ for $a_i \in G(p_i)$, $i = 2, 3, ..., k$. With $a_1 = mua \in G(p_1)$, we have by $a = mua + nva$ that $a = a_1 + a_2 + ... + a_k$ as desired. □

Using Lemma 4.73 we can prove our first decomposition result:

Theorem 4.74. *Let G be an abelian group and $\#G = p_1^{r_1} ... p_k^{r_k}$ the prime decomposition of $\#G$. Then*

$$G \cong G(p_1) \bigoplus ... \bigoplus G(p_k).$$

Proof. For $a \in G$, $o(a) \mid \#G$ and hence the primes appearing in the decomposition of a are a subset of $\{p_1, ..., p_k\}$. Thus, by Lemma 4.73 we may write $a = a_1 + ... + a_k$ with $a_i \in G(p_i)$ for $i = 1, ..., k$ (where if $p_i \nmid o(a)$ we set $a_i = 0$).

We now claim that the decomposition $a = a_1 + ... + a_k$ is unique. Suppose that also $a = b_1 + ... + b_k$. Then

$$a_1 - b_1 = (b_2 - a_2) + ... + (b_k - a_k).$$

Since each $G(p_i)$ is a group, $a_1 - b_1 \in G(p_1)$ and $b_i - a_i \in G(p_i)$ for $i = 2, ..., k$. Thus, for $i = 2, ...k$ there exists $s_i \geq 0$ such that $o(b_i - a_i) = p_i^{s_i}$. If we set $m = p_2^{s_2} ... p_k^{s_k}$ then $m(b_i - a_i) = 0$ for all $i = 2, ..., k$ hence also $m(a_1 - b_1) = 0$. Since $m \nmid o(a_1 - b_1) = p_1^{s_1}$, we must have $s_1 = 0$ so $a_1 - b_1 = 0$ ie $a_1 = b_1$. One proceeds in the same way to show that $a_i = b_i$ for $i = 2, ..., k$.

We now define a map $f : G \longrightarrow G(p_1) \oplus ... \oplus G(p_k)$ by setting $f(a) = (a_1, ..., a_k)$ where $a = a_1 + ... + a_k$ is the unique decomposition of a. Uniqueness implies that f is injective. To see that it is surjective, note that if $(a_1, ..., a_k) \in G(p_1) \oplus ... \oplus G(p_t)$, then $f(a_1 + ... + a_k) = (a_1, ..., a_k)$. To see that f is an isomorphism, it is left to check that f is a homomorphism:

$$f((a_1 + ... + a_k) + (b_1 + ... + b_k)) = f((a_1 + b_1) + ... + (a_k + b_k)) = (a_1 + b_1, ..., a_k + b_k) = (a_1, ..., a_k) + (b_1, ..., b_k).$$

□

The order of a primary p-group $G(p)$ is p^n for some n: if it wasn't the case, then there would be some prime $p' \neq p$ that divides the order of the group. Cauchy theorem then implies that there is an element of order p', in contradiction to the definition of $G(p)$.

We say that G is a **p-group** if its order is a power of p (so every primary p-group is a p-group). An element $a \in G$ is of **maximal order** if for every $b \in G$, $o(b) \leq o(a)$. The following lemma is key to our classification theorem but its proof is rather technical, so the reader may skip it.

Lemma 4.75. *Let G be a finite abelian p-group and $a \in G$ an element of maximal order. Then*

$$G \cong \langle a \rangle \bigoplus K$$

for some $K \leq G$.

Proof. Consider the set of subgroups H of G such that $\langle a \rangle \cap H = \{0\}$. This set is nonempty since $H = \{0\}$ satisfies $\langle a \rangle \cap H = \{0\}$. This set is finite because G is finite. In this set there is then a largest subgroup K of G (i.e., with the largest order) for which $\langle a \rangle \cap K = \{0\}$. We are to show that $G = \langle a \rangle + K$. Suppose to the contrary that $G \neq \langle a \rangle + K$, so there is nonzero $b \in G$ such that $b \in \langle a \rangle + K$ Since G is a p-group, there is an integer $j \geq 0$ such that $p^j b = 0$. Thus there is a smallest positive integer $l \leq j$ such that $p^l b \in \langle a \rangle + K$ because $0 \in \langle 0 \rangle + K$. Then the element $c = p^{l-1} b$ is not in $\langle a \rangle + K$ by the minimality of l. However, $pc = p^l b$ is in $\langle a \rangle + K$, so there are $t \in \mathbb{Z}$ and $k \in K$ such that $pc = ta + k$. Since a has maximal order $p^n = o(a)$ then $p^n x = 0$ for all $x \in G$ (convince yourself ;)). Thus $p^{n-1}(ta+k) = p^{n-1}(ta+k) = p^{n-1}pc = 0$. This means that $p^{n-1}ta = -p^{n-1}k$ where $p^{n-1}ta \in \langle a \rangle$ and $p^{n-1}k \in K$. Since $\langle a \rangle \cap K = \{0\}$, then $p^{n-1}ta = 0$. With $o(a) = p^n$ we have that $p^n \mid p^{n-1}t$, so that $p^{n-1}t = p^n m$ for some $m \in \mathbb{Z}$. Cancellation of p^{n-1} on both sides gives $t = pm$. This gives $pc = ta + k = pma + k$, so that $k = pc - pma = p(c - ma)$. Set $d = c - ma$. Then $pd = p(c - ma) = pk \in K$, but $d \notin K$ since if $k' = c - ma \in K$ would imply that $c = ma + k' \in \langle a \rangle + K$, contradicting $c \notin \langle a \rangle + K$. The set $H = \{x + zd | x \in K, z \in \mathbb{Z}\}$ is a subgroup of G that satisfies $K \subseteq H$. Since $d = 0 + 1d \in H$ but $d \notin K$, we have $H \neq K$, so that K is a proper subset of H. Since K is the largest subgroup of G for which $\langle a \rangle \cap K = \{0\}$, then $\langle a \rangle \cap H \neq \{0\}$. Let w be a nonzero element of $\langle a \rangle \cap H$. Then $sa = w = k_1 + rd$ for $k_1 \in K$ and $r, s \in \mathbb{Z}$. We show that $\gcd(p, r) = 1$ by assuming $\gcd(p, r) > 1$ and reaching a contradiction. If $\gcd(p, r) > 1$, then as p is prime, we have $r = py$ for some $y \in \mathbb{Z}$, then as $pd \in K$ we have $0 \neq w = sa = k1 + pyd \in \langle a \rangle \cap K$, a contradiction. Consequently, we have that $\gcd(p, r) = 1$. Then there are integers u, v such that $pu + rv = 1$. Therefore

$$\begin{aligned} c &= 1c \\ &= (pu + rv)c \\ &= u(pc) + v(rc) \\ &= u(ta + k) + v(r(d + ma)) \\ &= u(ta + k) + v(rd + rma) \\ &= u(ta + k) + v(sa - k_1 + rma) \\ &= (ut + vs + rm)a + (uk - vk_1) \in \langle a \rangle + K. \end{aligned} \tag{7}$$

This contradicts $c \notin \langle a \rangle + K$ so that $G = \langle a \rangle + K$. We now construct a map $f : G \longrightarrow \langle a \rangle \bigoplus K$ by taking $x = ra + k \in G$ to $f(x) = (ra, k)$. We leave it for the reader to verify that f is a group isomorphism. □

We can now state our first classification

Theorem 4.76 (structure theorem of finite abelian groups)**.** *Every finite abelian group G is isomorphic to the direct sum of cyclic groups, each of a prime power order. In other words,*

$$G \cong \mathbb{Z}_{p_1^{s_1}} \bigoplus \ldots \bigoplus \mathbb{Z}_{p_s^{s_k}}$$

where $p_1, \ldots, p_k$ are (not necessarily distinct) primes.

Proof. Suppose the prime decomposition of $\#G$ is $\#G = p_1^{n_1} \ldots p_t^{n_k}$. Then

$$G \cong G(p_1) \bigoplus \ldots \bigoplus G(p_k)$$

and by the discussion above, each $G(p_i)$ is a p_i-group, ie of order a power of p_i. Thus, it is enough to show that every p-group H is a direct sum of cyclic groups (whose order must be a power of p), which we'll do by induction on $\#H$. For $\#H = 2$, $H \cong \mathbb{Z}_2$ is a cyclic group.

Now assume the assertion is true for all p-groups whose order is less than $\#H$, and let a be an element of maximal order p^n in H. By Lemma 4.75, $H \cong \langle a \rangle \oplus K$ for some $K \leq H$ and by the inductive hypothesis, K is a direct sum of cyclic groups so H is also such. This finishes the proof. □

Example 4.77. The integer 100 can be written as a product of prime powers in four ways:

$$2 \cdot 2 \cdot 5 \cdot 5, \;\; 2 \cdot 2 \cdot 5^2, \;\; 2^2 \cdot 5 \cdot 5, \;\; 2^2 \cdot 5^2.$$

The structure Theorem of Finite Abelian Groups tells us that every abelian group of order 100 must be isomorphic to one of

$$\mathbb{Z}_2 \bigoplus \mathbb{Z}_2 \bigoplus \mathbb{Z}_5 \bigoplus \mathbb{Z}_5, \;\; \mathbb{Z}_2 \bigoplus \mathbb{Z}_2 \bigoplus \mathbb{Z}_{25} \bigoplus \mathbb{Z}_{16}, \;\; \mathbb{Z}_{16} \bigoplus \mathbb{Z}_5 \bigoplus \mathbb{Z}_5, \;\; \mathbb{Z}_4 \bigoplus \mathbb{Z}_{25}.$$

No two of these finite abelian groups are isomorphic (as they have different numbers of elements of order 2 and 5). Why is $\mathbb{Z}_{100}$ not on the list? By Corollary 4.67 it is $\mathbb{Z}_4 \oplus \mathbb{Z}_{25}$. In fact, this result can be extended into the following

Theorem 4.78. *Let n be a positive integer whose prime decomposition is $n = p_1^{n_1} ... p_k^{n_k}$. Then*

$$\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{n_1}} \bigoplus ... \bigoplus \mathbb{Z}_{p_k^{n_k}}.$$

Proof. By induction on n, suppose the statement is true for all numbers smaller than n. From Corollary 4.67 we have

$$\mathbb{Z}_n \cong \mathbb{Z}_{p_1^{n_1}} \bigoplus \mathbb{Z}_m$$

where $m = p_2^{n_2} ... p_k^{n_k}$. By the inductive hypothesis

$$\mathbb{Z}_m \cong \mathbb{Z}_{p_2^{n_2}} \bigoplus ... \bigoplus \mathbb{Z}_{p_k^{n_k}}$$

so we are done. □

We can now use Theorem 4.78 and the structure theorem of finite abelian groups to get another decomposition of G:

Example 4.79. For the finite abelian group, given as the direct sum of cyclic groups of prime power orders,

$$G \cong \mathbb{Z}_2 \bigoplus \mathbb{Z}_4 \bigoplus \mathbb{Z}_3 \bigoplus \mathbb{Z}_9 \bigoplus \mathbb{Z}_{81} \bigoplus \mathbb{Z}_{25} \bigoplus \mathbb{Z}_{125},$$

we arrange the prime power orders of the cyclic factors by size, with one row for each prime as such:

$$\begin{array}{ccc} & 2 & 2^2 \\ 3 & 3^2 & 3^4 \\ & 5^2 & 5^3 \end{array} \tag{8}$$

Then we use the columns (and Theorem 4.78) to express

$$G \cong \mathbb{Z}_3 \bigoplus (\mathbb{Z}_2 \bigoplus \mathbb{Z}_{3^2} \bigoplus \mathbb{Z}_{5^2}) \bigoplus (\mathbb{Z}_{2^2} \bigoplus \mathbb{Z}_{3^4} \bigoplus \mathbb{Z}_{5^3}) \cong \mathbb{Z}_3 \bigoplus \mathbb{Z}_{450} \bigoplus \mathbb{Z}_{40500}.$$

Notice that $3 \mid 450$ and $450 \mid 40500$ since $40500 = 450 \cdot 90$.

This can always be done.

Theorem 4.80 (structure theorem for finite abelian groups, 2nd version)**.** *Let G be a finite abelian group. Then*

$$G \cong \mathbb{Z}_{n_1} \bigoplus ... \bigoplus \mathbb{Z}_{n_k}$$

where $n_i \in \mathbb{N}$ and for all i $n_i \mid n_{i+1}$.

Proof. Left as an exercise. □

In case the abelian group in question is not finite, the 'closest' property for being finite is given by the following

Definition 4.81. An abelian group G is said to be **finitely generated** if there is a set of elements $\{g_1, ..., g_k\} \subseteq G$, called a **set of generators**, such that for each $x \in G$ there are integers $n_1, ..., n_k \in \mathbb{Z}$ such that

$$x = n_1 g_1 + ... n_k g_k.$$

Examples 4.82.

1. $\mathbb{Z}$ is finitely generated since $\langle 1 \rangle = \mathbb{Z}$.
2. $\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z}$ is finitely generated by $(1,0)$ and $(0,1)$: indeed, if $x = (a,b) \in \mathbb{Z} \times \mathbb{Z}$ then $x = a(1,0) + b(0,1)$.
3. more generally, the group $\mathbb{Z}^r = \underbrace{\mathbb{Z} \times ... \times \mathbb{Z}}_{r-times}$ is finitely generated and a set of generators is given by
$$\{(1,0,...,0), (0,1,0,...,0), ..., (0,...,0,1)\}.$$
4. The group $(\mathbb{Q}, +)$ is not finitely generated. If $\mathcal{A} = \{\frac{p_1}{q_1}, ..., \frac{p_n}{q_n}\}$ is a finite set then the additive subgroup it generates is all fractions whose denominator is the least common multiple of $q_1, ..., q_n$ hence $\mathcal{A}$ cannot be a set of generators.

It turns out that finitely generated abelian groups have a similar structure theorem.

Theorem 4.83 (structure theorem for finitely generated abelian groups)**.** *Let G be a finitely generated abelian group. Then there are positive integers $n_1, ..., n_k, r$ and an isomorphism*

$$G \cong \mathbb{Z}_{n_1} \oplus ... \oplus \mathbb{Z}_{n_k} \oplus \mathbb{Z}^r.$$

The number r is called the ***rank*** *of G.*

4.5 Aside: Permutation group

Let $1 \leq n$ be an integer and consider the set $[n] = \{1, ..., n\}$. A **permutation** of n is a bijective function $\sigma : [n] \longrightarrow [n]$. If $\tau : [n] \longrightarrow [n]$ is also a permutation, the composition $\tau \circ \sigma : [n] \longrightarrow [n]$ is bijective (as composition of such), hence a permutation as well. The identity map $I = \mathrm{id}_{[n]} : [n] \longrightarrow [n]$ is clearly a permutation and the inverse of a permutation $\sigma : [n] \longrightarrow [n]$ is a permutation $\sigma^{-1} : [n] \longrightarrow [n]$ satisfying $\sigma \circ \sigma^{-1} = I$.

Cycle notation of a permutation $\sigma \in S_n$ describes the effect of repeatedly applying the permutation on the elements of the set. It expresses the permutation as a product of cycles; since distinct cycles are disjoint, this is referred to as "decomposition into disjoint cycles".

To write down the permutation σ in cycle notation, one proceeds as follows:

1. Write an opening bracket then select an arbitrary element x of $[n]$ and write it down: $(x(x$.
2. Then trace the orbit of x; that is, write down its values under successive applications of $\sigma :$

$$(\, x \, \sigma(x) \, \sigma(\sigma(x)) \, \ldots$$

3. Repeat until the value returns to x and write down a closing parenthesis rather than x:

$$(\, x \, \sigma(x) \, \sigma(\sigma(x)) \, \ldots)$$

4. Now continue with an element $y \in [n]$, not yet written down, and proceed in the same way:

$$(\, x \, \sigma(x) \, \sigma(\sigma(x)) \, \ldots)(\, y \, \ldots)$$

5. Repeat until all elements of S are written in cycles.

Definition 4.84. A cycle is a permutation $\sigma \in S_n$ with the property that the cycle representation of σ has exactly one cycle. For instance $\sigma = (a_1 a_2 ... a_k)$. We call k the length of the cycle.

Remark 4.85. It may seem that there is ambiguity about an expression such as (164)(29)(8735). Is this one permutation with three cycles, or a product of the three cycles (164), (29), and (8735)? Fortunately, the permutation (164)(29)(8735) is equal to the product of the three cycles (164), (29), and (8735), so there is no trouble.

Definition 4.86. A transposition is a cycle of length 2. So, in cycle notation, a transposition has the form (ab). Note that every transposition is its own inverse: $(ab)(ab) = I$.

Lemma 4.87. *Every permutation can be expressed as a product of transpositions.*

Proof. A quick check reveals that a cycle $(a_1, a_2, ..., a_k)$ can be represented as follows: $(a_1 a_2 a_3 ... a_k) = (a_1 a_k)...(a_1 a_4)(a_1 a_3)(a_1 a_2)$. Since every permutation is a product of cycles, every permutation may be represented as a product of transpositions. □

Example 4.88. To represent the permutation

$$(13584)(2967) \in S_9$$

as a product of transpositions, write

$$(13584)(2967) = (14)(18)(15)(13)(27)(26)(29)$$

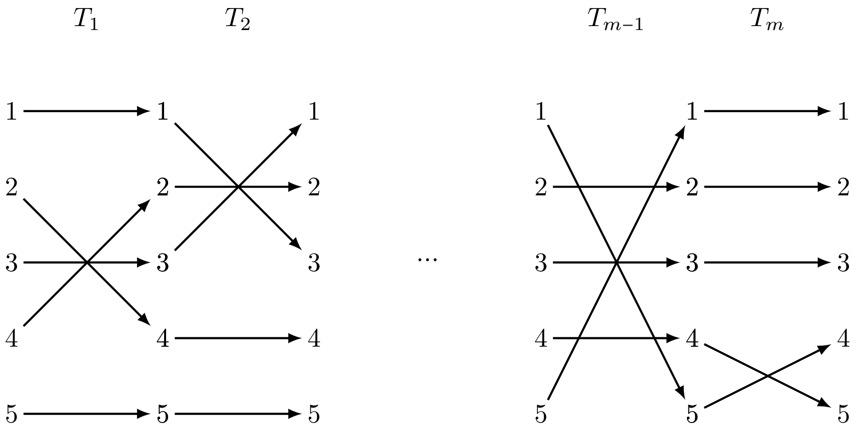
Remark 4.89. Every permutation can be expressed as a product of transpositions in many (actually infinitely many) ways. For instance, the permutation (13584)(2967) from the above example can also be expressed in all of the following ways

$$\begin{aligned} (13584)(2967) &= (72)(38)(17)(28)(47) \\ &= (69)(64)(68)(12)(17)(15)(13)(56)(24) \\ &= (91)(95)(98)(94)(93)(92)(97)(96)(45)(83)(12) \end{aligned} \tag{9}$$

We have just represented a particular permutation as a product of 5, 9, and 11 transpositions— all odd numbers. This is not a coincidence!

Proposition 4.90. *If $T_1, ..., T_m$ are transpositions and $T_1 T_2 T_m = I$, then m is even.*

Proof. Let us think about composition using a figure where each bijection in is represented by arrows from $1, 2, ..., n$ to $1, 2, ..., n$ as follows:



Now, if we start at the rightmost 1 and follow the arrows we go along a path, finally ending up at the leftmost 1 (since the product $T_1T_2T_m$ is the identity). Let's imagine this path as a string which we call strand 1. Similarly, for every $2 \le i \le n$ we have a strand starting and ending at i. The figure may be complicated since the strands may cross one another many times, but nevertheless, we can reason about these crossings. Make the following definitions:

1. Let $c_{i,j}$ be the number of times strand i and strand j cross.
2. Let c be the total number of times one strand crosses another.
3. Let t_k be the number of times one strand crosses another in the transposition T_k.

For any two distinct strands, say i and j, it must be that $c_{i,j}$ is even, since these strands must cross an even number of times in order to end in the same positions in which they begin. It follows from this that the total number of crossings c must also be even. Next let us think about the number of crossings contributed by a single transposition T_k. Suppose (without loss of generality) that $T_k = (ab)$ where $a < b$. If $b = a + 1$ then the only strands crossing in T_k are strand a and $a + 1$ so $t_k = 1$. If $b = a + 2$ then strands a and $a + 2$ cross in T_k but strand $a + 1$ also gets crossed by strand a and $a + 2$ for a total of 3 crossings. More generally, if $b = a + p$ then strands a and $a + p$ will cross one another, and both of these strands will cross all of the strands numbered $a + 1, a + 2, ..., a + p - 1$. This gives a total of $1 + 2(p - 1)$ crossings, so $t_k = 1 + 2(p - 1)$ is odd. The total number of crossings can also be counted by summing the contributions from each transposition, giving us the equation

$$c = t_1 + t_2 + ... + t_m$$

Now c is even, but each tk is odd, and it follows that m must be even, as desired. □

Now we are ready to prove that this even/odd property holds for every permutation.

Theorem 4.91. *For every permutation $\sigma \in S_n$, either every representation of σ as a product of transpositions has an odd number of transpositions, or every such representation has an even number of transpositions.*

Proof. Let $T_1, ..., T_j$ and $U_1, ..., U_k$ be transpositions satisfying

$$\sigma = T_1T_2T_j = U_1U_2U_k$$

To prove the theorem it suffices to show that either j and k are both even, or both odd. For any product of permutations $\tau = \tau_1\tau_2\cdots\tau_m$ the inverse is always given by

$$\tau^{-1} = \tau_m^{-1}\ldots\tau_1^{-1}$$

Since every transposition is its own inverse, we can therefore express σ^{-1} as

$$\sigma^{-1} = U_k U_{k-1}\cdots U_2 U_1$$

Now we have

$$I = \sigma\sigma^{-1} = T_1\cdots T_j U_k\cdots U_1$$

By the previous proposition we deduce that $j+k$ is even, so either j and k are both even, or j and k are both odd, as desired. □

We are ready to make the following

Definition 4.92. Let $\sigma \in S_n$ be a permutation. The **sign** of σ is $\operatorname{sign}(\sigma) = (-1)^m$ where m is the number of transpositions in some (hence any) representation of σ as a product of transpositions.

Example 4.93. If σ is a cycle of length k, say

$$\sigma = (a_1a_2\ldots a_k)$$

then we can express σ as

$$A = (a_1a_k)\ldots(a_1a_3)(a_1a_2).$$

Therefore if a cycle σ is of even length then $\operatorname{sign}(\sigma) = -1$ and if it is of odd length then $\operatorname{sign}(\sigma) = 1$.

5 Fields

5.1 Fields and field homomorphisms

The notion of a group is considered one of the most basic ones in modern Mathematics. In a group we are dealing with a set and a binary operation, that abstracts the properties of multiplication (or addition). However, going back to real numbers, we have in fact **two binary operations**, namely addition and multiplication and these operations are compatible in some sense. In order to abstract these properties for more general cases, we have the following

Definition 5.1. A **field** is the data of a set $\mathbb{F}$ together with two binary operations $+ : \mathbb{F}\times\mathbb{F} \longrightarrow \mathbb{F}$ and $\cdot : \mathbb{F}\times\mathbb{F} \longrightarrow \mathbb{F}$ called addition and multiplication and two specified elements $0, 1 \in \mathbb{F}$ subject to the following conditions:

1. (associativity of + and ·) for every $x, y, z \in \mathbb{F}$, $(x+y)+z = x+(y+z)$ and $(x\cdot y)\cdot z = x\cdot(y\cdot z)$.
2. (commutativity of + and ·) for every $x, y \in \mathbb{F}$, $x+y = y+x$ and $x\cdot y = y\cdot x$.
3. (distributivity) for every $x, y, z \in \mathbb{F}$, $x\cdot(y+z) = x\cdot y + x\cdot z$.
4. (neutral elements wrt + and ·) for every $x \in \mathbb{F}$, $x+0 = x$, $x\cdot 0 = 0$ and $x\cdot 1 = x$.
5. (existence of inverses wrt +) for every $x \in \mathbb{F}$ there is an element $-x$ such that $x+(-x) = 0$.
6. (existence of inverses wrt ·) if $x \neq 0$ there exists an element $x^{-1} \in \mathbb{F}$ such that $x\cdot x^{-1} = 1$.

Remark 5.2. As with groups, we will usually ommit $\cdot$ when multiplying elements in a field, ie write xy instead of $x \cdot y$.

Examples 5.3.

1. The real numbers $\mathbb{R}$ with usual addition and multiplication are a field.
2. The rational numbers $\mathbb{Q}$ with usual addition and multiplication are a field. Note that we need to make sure that addition and multiplication of two rational numbers result in a rational number.
3. The complex numbers $\mathbb{C}$ with addition and multiplication of complex numbers are a field. The non-trivial condition to check is existence of multiplicative inverse: if $z = r(\cos\theta + i\sin\theta) \neq 0$, then $z^{-1} = \frac{1}{r}(\cos(-\theta) + i\sin(-\theta))$.

Exercise 5.4.

1. Prove that $\sqrt{2} \notin \mathbb{Q}$. Hint: suppose by contradiction that $\sqrt{2} \in \mathbb{Q}$ and let $a, b \in \mathbb{Z} \setminus \{0\}$ be such that $\sqrt{2} = \frac{a}{b}$ and $\gcd(a, b) = 1$. Then $a^2 = 2b^2$. Derive a contradiction by dividing to cases where a, b are odd or even.
2. Let $\mathbb{Q}(\sqrt{2}) = \{x + y\sqrt{2} | x, y \in \mathbb{Q}\}$ (note the analogy with $\mathbb{C}$) with addition and multiplication induced from $\mathbb{R}$. Prove that $\mathbb{Q}(\sqrt{2})$ is a field. Write an explicit formula for the inverse of $x + y\sqrt{2}$.

Lemma 5.5. *If $\mathbb{F}$ is a field and $x, y \in \mathbb{F}$ such that $xy = 0$ then either $x = 0$ or $y = 0$.*

Proof. Otherwise, we could multiply by x^{-1} and get $x^{-1}xy = 1 \cdot y = y = 0 \cdot x^{-1} = 0$ – a contradiction. □

Proposition 5.6. *For $n \in \mathbb{N}$, $\mathbb{Z}_n$ with addition and multiplication modolu n is a field iff $n = p$ is prime, and denoted $\mathbb{F}_p$*

Proof. The associativity, commutativity distributivity and neutral elements axioms hold for any $n \in \mathbb{N}$. In addition, inverses for $+$ exist in $\mathbb{Z}_n$ as it is a group. The remaining axiom to check is existence of multiplicative inverses. If n is not prime, then $n = k \cdot l$ for some $1 \leq k, l \leq n - 1$ but then $k \cdot l = 0$ in $\mathbb{Z}_n$ whereas $k, l \neq 0$ – in contradiction to Lemma 5.5. Thus in that case, $\mathbb{Z}_n$ cannot be a field. conversely, suppose $n = p$ is prime. For $0 \neq x \in \mathbb{Z}_p$ we have, by Fermat's little theorem $x^{p-1} = 1 \ (mod\, p)$. Thus $x^{p-2} = x^{-1}$ and we're done. □

What sets apart a field like $\mathbb{Q}$ from the field $\mathbb{F}_p$? The following definition gives such a suggestion:

Definition 5.7. Let $\mathbb{F}$ be a field. If there is a minimal number $n \in \mathbb{N}$ such that $\underbrace{1 + 1 + ... + 1}_{n\text{–}times} = 0$ is called the **characteristic** of $\mathbb{F}$ and we denote $\operatorname{char}\mathbb{F} = n$. If no such number exists, we say that $\mathbb{F}$ is of **characteristic 0** and denote $\operatorname{char}\mathbb{F} = 0$.

Proposition 5.8. *The characteristic of a field $\mathbb{F}$ must be 0 or prime number p.*

Proof. Let $n \in \mathbb{N}$ be the characteristic of $\mathbb{F}$. If n is not prime, then $n = kl$ for some $k, l > 1$. But then $\underbrace{1 + 1 + ... + 1}_{n\text{–}times} = (\underbrace{1 + 1 + ... + 1}_{k\text{–}times})(\underbrace{1 + 1 + ... + 1}_{l\text{–}times}) = 0$ in contradiction to Lemma 5.5. □

As with groups, we would like to have a notion of a map of fields.

Definition 5.9. Let $\mathbb{F}, \mathbb{K}$ be fields and $f : \mathbb{F} \longrightarrow \mathbb{K}$ a function. We say that f is a **field homomorphism** if:

1. $f(0) = 0$ and $f(1) = 1$.

2. for any $x, y \in \mathbb{F}$, $f(x+y) = f(x) + f(y)$.

3. for any $x, y \in \mathbb{F}$, $f(xy) = f(x)f(y)$.

We say that f is a **field isomorphism** if it is furthermore an isomorphism of sets.

Examples 5.10. 1. The inclusions $\mathbb{Q} \hookrightarrow \mathbb{R}$ and $\mathbb{R} \hookrightarrow \mathbb{C}$ are field homomorphisms.

2. There is no field homomorphism $f : \mathbb{F}_p \longrightarrow \mathbb{Q}$. If there were, then $0 = f(0) = f(\underbrace{1+1+...+1}_{p-times}) = \underbrace{f(1)+f(1)+...+f(1)}_{p-times} = \underbrace{1+1+...+1}_{p-times}$ since f is a homomorphism but $\mathbb{Q}$ has characteristic 0 which is a contradiction.

Exercise 5.11. Prove that if $\operatorname{char}\mathbb{F} = p$ and $\operatorname{char}\mathbb{K} = p'$ for $p \neq p'$ then there is no field homomorphism $f : \mathbb{F} \longrightarrow \mathbb{K}$.

Our main basis in cryptography is that of finite fields. Note that a field of characteristic 0 must be infinite since the elements $1, 1+1, 1+1+1, ...$ must all be different. Thus, by Proposition 5.8 a finite field must have characteristic p for some prime number. We saw that $\mathbb{F}_p$ is such a field, and we may ask:

Question 5.12. What are the fields of characteristic p up to isomorphism?

Let $\mathbb{F}$ be a field with p elements (p prime). We can view $(\mathbb{F}_p, +)$ as an abelian group and consider the cyclic group generated by 1: $\langle 1 \rangle = \{1, 1+1, ...\}$. By Lagrange's Theorem 4.20, $|\langle 1 \rangle|$ must divide p. Since p is prime and $\langle 1 \rangle$ has at least two elements $1 \neq 1+1$, it follows that $|\langle 1 \rangle| = p$ so that $\langle 1 \rangle = \mathbb{F}$ and by Proposition 4.58 we have an isomorphism of groups $\langle 1 \rangle \cong \mathbb{Z}_p$. Define a map $f : \mathbb{F}_p \longrightarrow \mathbb{F}$ by $f(0) = 0$ and $f(n) = \underbrace{1+...+1}_{n-times}$. Then it is easy to check that f is an isomorphism of fields so that $\mathbb{F} \cong \mathbb{F}_p$. We have just proved:

Corollary 5.13. *Let p be prime. Then any field with p elements is isomorphic to $\mathbb{F}_p$*

To begin answering Question 5.12 it will be useful to consider the following

Definition 5.14. Let $\mathbb{F}$ be a field. A subset $\mathbb{K} \subseteq \mathbb{F}$ is called a **subfield** if

1. $0_{\mathbb{F}}, 1_{\mathbb{F}} \in \mathbb{K}$.

2. $\mathbb{K}$ is a field under the addition and multiplication defined in $\mathbb{F}$

Examples 5.15.

1. $\mathbb{Q} \subseteq \mathbb{R}$ is a subfield inclusion.

2. $\mathbb{Q} \subseteq \mathbb{Q}(\sqrt{2})$ is a subfield inclusion.

3. $\mathbb{F}_p \subseteq \mathbb{Q}$ is not a subfield inclusion since $\mathbb{F}_p$ is not a field under the addition and multiplication of $\mathbb{Q}$. This is because, e.g., $\underbrace{1+1+...+1}_{p-times} = 0$ in $\mathbb{F}_p$ and $\underbrace{1+1+...+1}_{p-times} = p$ in $\mathbb{Q}$.

Using Definition 5.14 we can identify 'minimal' fields of characteristic p and characteristic 0 as the following propositions show.

Proposition 5.16. *Let p be prime and $\mathbb{F}$ a field of characteristic p. Then $\mathbb{F}$ contains a subfield which is isomorphic to $\mathbb{F}_p$*

Proof. Consider the set

$$\mathbb{K} = \{1, 1+1, ..., \underbrace{1+1+...+1}_{(p)-times} = 0\}.$$

then $\mathbb{K} \subseteq \mathbb{F}$ is a subfield and we can define an isomorphism of fields $f : \mathbb{K} \longrightarrow \mathbb{F}_p$ by $\underbrace{1+1+...+1}_{(k)-times} \mapsto k$. □

Proposition 5.17. *Let $\mathbb{F}$ be a field of characteristic 0. Then $\mathbb{F}$ contains a subfield which is isomorphic to $\mathbb{Q}$.*

Proof. First, note that $\mathbb{F}$ contains the set $\mathbb{F}_{\mathbb{Z}} = \{..., -(1+1), -1, 0, 1, 1+1, ...\}$ which can be thought of as a copy of the integers. Second, $\mathbb{F}$ must contain inverses to all non-zero elements in $\mathbb{F}_{\mathbb{Z}}$ i.e. elements of the form $\frac{1}{b}$ with $b \in \mathbb{F}_{\mathbb{Z}} \setminus \{0\}$. Third $\mathbb{F}$ is closed to multiplication hence must contain all elements of the form $a \cdot \frac{1}{b} = \frac{a}{b}$ where $a \in \mathbb{F}_{\mathbb{Z}}$ and $b \in \mathbb{F}_{\mathbb{Z}} \setminus \{0\}$. The last set is a subfield of $\mathbb{F}$ isomorphic to $\mathbb{Q}$. □

5.2 Vector spaces over fields

To give a partial answer to Question 5.12 in full we will need to take a small digression to the theory of vector spaces.

Definition 5.18. Let $\mathbb{F}$ be a field. A **vector space** V over $\mathbb{F}$ is a set V together with a distinguised element $0_V \in V$ and two binary operations $+ : V \times V \longrightarrow V$, $\cdot : \mathbb{F} \times V \longrightarrow V$ called addition of vectors and multiplication of a vector by a scalar, satisfying the following conditions:

1. (associativity of addition) for all $u, v, w \in V$, $(u+v)+w = u+(v+w)$.
2. (commutativity of addition) for all $v, w \in V$, $v+w = w+v$.
3. (associativity of scalar multiplication) for every $\alpha, \beta \in \mathbb{F}$ and every $v \in V$, $(\alpha\beta)v = \alpha(\beta \cdot v)$.
4. (distributivity) for every $\alpha \in \mathbb{F}$ and $v, w \in V$, $\alpha \cdot (v+w) = \alpha v + \alpha w$.
5. (neutral element) for every $v \in V$, $0_V + v = v$ and $0_{\mathbb{F}} \cdot v = 0_V$.
6. (existence of inverse wrt addition) for every $v \in V$, there is an element $-v \in V$ such that $v + (-v) = 0_V$.

Examples 5.19.

1. Every field $\mathbb{F}$ is a vector space over itself.
2. For any field $\mathbb{F}$, the Cartesian product
$$\mathbb{F}^n := \underbrace{\mathbb{F} \times ... \times \mathbb{F}}_{n-times}$$
is a vector space over $\mathbb{F}$: addition of vectors is defined by
$$(x_1, ..., x_n) + (y_1, ..., y_n) := (x_1 + y_1, ..., x_n + y_n)$$
and scalar multiplication is defined by
$$\alpha \cdot (x_1, ..., x_n) := (\alpha x_1, ..., \alpha x_n).$$
The inverse for addition of vectors is given by
$$-(x_1, ..., x_n) := (-x_1, ..., -x_n)$$
where $-x_i$ is the inverse of x_i with respect to addition in $\mathbb{F}$.

3. The complex numbers $\mathbb{C}$ are a vector space over $\mathbb{R}$: for $\alpha \in \mathbb{R}$ and $z = x + iy \in \mathbb{C}$ we define scalar multiplication as $\alpha z := \alpha x + \alpha y i$.

4. Let $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$ be the field of integers modulo 2 and X be an arbitrary set. Consider the **power set** $P(X) := \{S | S \subseteq X\}$. Then $V = P(X)$ is a vector space over $\mathbb{F}_2$ where $0_V = \varnothing$ and for $S, S' \in V$ (ie $S, S' \subseteq X$) we set $S +_V S' := (S \smallsetminus S') \cup (S' \smallsetminus S)$. For a vector $S \in V$, $-S := X \smallsetminus S$

Exercise 5.20. Let V be a vector space over a field $\mathbb{F}$.

1. Show that the 0 vector in a vector space is unique, ie show that if $v \in V$ is a vector satisfying $v + w = w$ for every $w \in V$ then $v = 0_V$.

2. Show that for every vector $v \in V$ $1_{\mathbb{F}} \cdot v = v$.

Definition 5.21. Let V be a vector space over a field $\mathbb{F}$. A subset $W \subseteq V$ is called a subspace if it is closed under addition of vectors and multiplication of a vector by a scalar. In other words, $W \subseteq V$ is a subspace if the following conditions are satisfied:

1. $0_V \in W$

2. $\forall w, w' \in W : w + w' \in W$.

3. $\forall w \in W, \lambda \in \mathbb{F} : \lambda \cdot w \in W$.

Examples 5.22.

1. Of course, V and $\{0_V\}$ are always subspaces of V , that are referred to as **trivial** subspaces.

2. Let $V = \mathbb{R}^2$ be the Euclidean space of degree 2, viewed as a vector space over $\mathbb{R}$. Then a set $W \subseteq \mathbb{R}^2$ is a non trivial subspace iff its points are precisely a line in $\mathbb{R}^2$ passing through the origin.

3. Let $V = \mathbb{R}^3$ be the Euclidean space of degree 3, viewed as a vector space over $\mathbb{R}$. Then a set $W \subseteq \mathbb{R}^2$ is a non trivial subspace iff its points are precisely a line or a plane in $\mathbb{R}^3$ passing through the origin.

4. Let $f : V \longrightarrow W$ be a linear map. The **kernel** of f, defined by

$$\ker f := \{v \in V | f(v) = 0_w\} \subseteq V,$$

is a subspace of V.

5. Let $f : V \longrightarrow W$ be a linear map. The **image** of f, defined by

$$\operatorname{Im} f := \{w \in W | \exists v \in V : f(v) = w\} \subseteq W,$$

is a subspace of W.

As usual, we would like to have a notion of a map between vector spaces:

Definition 5.23. Let V, W be vector spaces over (the same) field $\mathbb{F}$. A function $f : V \longrightarrow W$ is called a **linear map** (synonyms: vector space homomorphism, linear transformation) if:

1. for every $\alpha \in \mathbb{F}$ and every $v \in V$, $f(\alpha v) = \alpha f(v)$.

2. for every $v, v' \in V$, $f(v + v') = f(v) + f(v')$.

A linear map $f : V \longrightarrow W$ is called a **vector space isomorphism** if f is an isomorphism of the underlying sets of V and W.

Exercise 5.24. Show that $f : V \longrightarrow W$ is a vector space isomorphism iff there exists a linear map $g : W \longrightarrow V$ (also denoted as f^{-1}) such that $f \circ g = \mathrm{id}_W$ and $g \circ f = \mathrm{id}_V$.

Examples 5.25.

1. the function $f : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ given by $f(x, y, z) = (2x + 4y - z, 5x + z)$ is a linear map.
2. the function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ given by $f(x, y) = xy$ is not a linear map: in $\mathbb{R}^2$ we have $(1, 1) + (1, 1) = (2, 2)$ but $f(2, 2) = 4$ whereas $f(1, 1) + f(1, 1) = 2$.
3. if V, W are vector spaces over a field $\mathbb{F}$, the map $V \longrightarrow W$ given by $v \mapsto 0_W$ is a linear map.
4. Consider $\mathbb{C}$ as a vector space over $\mathbb{R}$ and let $f : \mathbb{C} \longrightarrow \mathbb{R}^2$ be defined by $f(x + yi) = (x, y)$. Then f is a linear isomorphism. Note that when $\mathbb{C}$ is viewed as a vector space over $\mathbb{R}$, we have $\mathbb{C} \cong \mathbb{R}^2$ but $\mathbb{C}$ can also be viewed as field whereas $\mathbb{R}^2$ has no apparent structure of a field. Furthermore, we can view $\mathbb{C}$ as a vector space over the field $\mathbb{C}$ and there is no apparent way to view $\mathbb{R}^2$ as a vector space over $\mathbb{C}$.

Definition 5.26. Let V be a vector space over $\mathbb{F}$ and $S = \{v_1, ..., v_n\} \subseteq V$ a finite set of vectors.

1. a **linear combination** of S is a vector of the form $v = a_1 v_1 + ... + a_n v_n = \sum_{i=1}^n a_i v_i$ where $a_1, ..., a_n \in \mathbb{F}$.
2. The set of all linear combinations of S is called the **span** of S and denoted $\mathrm{Span}(S)$. In other words,
$$\mathrm{Span}(S) = \{a_1 v_1 + ... + a_n v_n | a_1, ..., a_n \in \mathbb{F}\}.$$
3. The set S is called **linearly independent** if whenever there are scalars $a_1, ..., a_n$ such that $a_1 v_1 + ... + a_n v_n = 0_V$, it follows that $a_1 = a_2 = ... = a_n = 0_{\mathbb{F}}$. In other words, S is linearly independent if the only linear combination of S that yields the zero vector 0_V is the trivial one, ie the one in which all scalars are $0_{\mathbb{F}}$.

Exercise 5.27. Prove that for any finite set $S \subseteq V$, $\mathrm{Span}\, S \subseteq V$ is a sub-vector space.

We will be mostly interested in a class of vector spaces that are particularly simple:

Definition 5.28. Let V be a vector space over a field $\mathbb{F}$. We say that V is finitely generated (or has **finite dimension**) if there are vectors $S = \{v_1, ..., v_n\} \subseteq V$ such that $\mathrm{Span}\, S = V$. In other words for any $v \in V$ there are scalars $a_1, ..., a_n \in \mathbb{F}$ that satisfy $v = a_1 v_1 + ... + a_n$. In such a case we also say that S is a **generating set** of V.

For example, let $\mathbb{F}$ be a field and consider $\mathbb{F}^n$ as a vector space over $\mathbb{F}$. Take $S = \{(1, 0, ..., 0), (0, 1, 0, ..., 0), ..., (0, ..., 0, 1)\} \in \mathbb{F}^n$. If $v = (x_1, ..., x_n) \in \mathbb{F}^n$ we can write $v = x_1 \cdot (1, 0, ..., 0) + x_2 \cdot (0, 1, 0, ..., 0) + ... + x_n \cdot (0, ..., 0, 1)$ where for all i, $x_i \in \mathbb{F}$. Thus $\mathbb{F}^n$ is a finitely generated vector space over $\mathbb{F}$.

Proposition 5.29. *Let V be a vector space over $\mathbb{F}$. For a finite set $S \subseteq V$ the following conditions are equivalent:*

1. *S is a minimal generating set of V.*
2. *S is a maximal linearly-independent subset of V.*
3. *S is a linearly independent generating set of V.*

Proof. Let us enumerate S as $S = \{v_0, v_1, ..., v_n\}$

1. $(1) \Rightarrow (2)$: Suppose S is a minimal generating set. If S is linearly dependent, then wlog, we can write $v_0 = a_1v_1 + ... + a_nv_n$ for some $a_1, ..., a_n \in \mathbb{F}$ and this means that $S' := S \setminus \{v_0\} = \{v_1, ..., v_n\}$ is a generating set for V: if $v \in V$ then since $\operatorname{Span} S = V$ we can write $v = b_0v_0 + b_1v_1 + ... + b_nv_n$ for some $b_0, ..., b_n \in \mathbb{F}$ and thus $v = b_0(a_1v_1 + ... + a_nv_n) + b_1v_1 + ... + b_nv_n$ which is a linear combination of the vectors in S'. This contradicts the minimality of Sso that S must be linearly independent. If $v \in V \setminus S$ then since $\operatorname{Span} S = V$, v is a linear combination of elements in S so that $S \cup \{v\}$ is linearly dependent. Hence S is a maximal linearly independent subset of V.

2. $(2) \Rightarrow (3)$: if S is not a generating set of V then there is $v \in V \setminus S$ such that v is not a linear combination of elements in S. But then $S \cup \{v\}$ is a linearly independent set in contradiction to the maximality of S.

3. $(3) \Rightarrow (1)$: if S is not a minimal generating set, then wlog the set $S' := S \setminus \{v_0\} = \{v_1, ..., v_n\}$ is a generating set. But then we can write $v_0 = c_1v_1 + ... + c_nv_n$ so that v_0 is linearly dependent on $v_1, ..., v_n$ in contradiction to the assumption that S is linearly independent.

□

Definition 5.30. A (finite) linearly-independent and generating set S with an order $S = \{v_1, ..., v_n\}$ is called a **basis** of V.

Proposition 5.31. *Let V be a finitely generated vector space over $\mathbb{F}$. Then V has a basis.*

Proof. Suppose $S = \{v_1, ..., v_n\}$ is a finite spanning set for V ie $V = \operatorname{Span}(\{v_1, ..., v_n\})$ for $n \in \mathbb{N}$ and prove by induction on n. If $n = 1$, S is linearly independent by definition since if $a_1v_1 = 0$ then $a_1 = 0$ by the axioms of a vector space. Let $n \in \mathbb{N}$ be an integer and suppose the claim is true for all $k < n$. If S is linearly independent, then it is a basis since it is spanning. Otherwise, S is not linearly independent, which means that there are scalars $a_1, ..., a_n$ such that $a_1 + ... + a_n = 0$ and there is $1 \leq k \leq n$ such that $a_k \neq 0$. Then $v_k = a_1/a_kv_1 + ... + a_{k-1}/a_kv_{k-1} + a_{k+1}/a_kv_{k+1} + ... + a_n/a_kv_n$ so that $v_k \in \operatorname{Span}(\{v_1, ..., v_{k-1}, v_{k+1}, ..., v_n\})$ ie $\operatorname{Span}(\{v_1, ..., v_{k-1}, v_{k+1}, ..., v_n\}) = \operatorname{Span}(\{v_1, ..., v_n\}) = V$.

Thus, V is spanned by a set of $n - 1$ vectors and we can revoke the induction hypothesis to conclude that V has a basis. □

Proposition 5.32. *Let V be a vector space over $\mathbb{F}$. A set $S = \{v_1, ..., v_n\}$ is a basis for V iff every element $v \in V$ can be uniquely written as a linear combination of elements in S.*

Proof. Suppose first that S is a basis. If a vector $v \in V$ can be written in two ways as a linear combination of elements in S, then there are scalars $a_1, ..., a_n \in \mathbb{F}$ and $b_1, ..., b_n \in \mathbb{F}$ such that $v = a_1v_v + ... + a_nv_n = b_1v_1 + ... + b_nv_n$. But then $0_V = v - v = (a_1 - b_1)v_1 + ... + (a_n - b_n)v_n$ and since $v_1, ..., v_n$ are linearly independent, this means that $a_i - b_i = 0$ for all i. Thus these two ways coincide. Conversely, if every element can be uniquely written as a linear combination of elements in S then clearly S is a generating set for V. If it was linearly dependant, then there would exist scalars $a_1, ..., a_n \in \mathbb{F}$, not all zero, such that $0_V = a_1v_1 + ... + a_nv_n$. However, we also have $0_{\mathbb{F}}v_1 + ... + 0_{\mathbb{F}}v_n = 0_V$ in contradiction the the uniqueness assumption. Thus, S must be also linearly independent ie a basis for V. □

Proposition 5.33. *If V is a vector space over a field $\mathbb{F}$ having a finite gener- ating set $S = \{v_1, ..., v_n\}$ then any linearly-independent subset of V has at most n elements.*

Proof. Assume that the desired result is false. Then there exists a linearly-independent subset $T = \{w_1, ..., w_{n+1}\}$ of V having $n + 1$ elements, none of which is 0_V (for otherwise T would be linearly dependent). Since S is a generating set for V, there exist scalars $a_1, ..., a_n$, not all equal to $0_{\mathbb{F}}$, satisfying $w_1 = \sum_{i=1}^{n} a_iv_i$. By renumbering the elements of S if necessary, we can in fact assume that $a_1 \neq 0_{\mathbb{F}}$. Then $v_1 = a_1^{-1}w_1 + a_1^{-1}a_2v_2 + ... + a_1^{-1}a_nv_n$ and so $S \subseteq \operatorname{Span} S'$, where $S' := \{w_1, v_2, ..., v_n\}$. But $S' \subseteq V = \operatorname{Span} S$ and so $V = \operatorname{Span} S'$. Now assume that $k < n$ and that we have already shown $V = \text{Span}\{w_1, .., w_k, v_{k+1}, ..., v_n\}$. Then there exist scalars $b_1, ..., b_n$, not all equal to $0_{\mathbb{F}}$, satisfying $w_{k+1} = b_1 w_1 + ... + b_k w_k + b_{k+1} v_{k+1} + ... + b_n v_n$. If the scalars $b_{k+1}, ..., b_n$ are all equal to $0_{\mathbb{F}}$, then we would have shown that T is linearly dependent, which is a contradiction. Therefore at least one of these scalars is nonzero and, by renumbering $\{v_{k+1}, ..., v_n\}$ if necessary, we can assume that $b_{k+1} \neq 0_{\mathbb{F}}$. As above, we then conclude that $v_{k+1} = -b_{k+1}^{-1} b_1 w_1 + ... + -b_{k+1}^{-1} b_k w_k + b_{k+1}^{-1} w_{k+1} - b_{k+1}^{-1} b_{k+2} v_{k+2} - ... - b_{k+1}^{-1} b_n v_n$ and so, as above, $V = \text{Span}\{w_1, ..., w_{k+1}, v_{k+2}, ..., v_n\}$. Thus we can continue in this manner and, after a finite number of steps, we get $V = \text{Span}\{w_1, ..., w_n\}$. But this implies that w_{n+1} is a linear combination of $\{w_1, ..., w_n\}$, contradicting the assumption that T is linearly independent. Thus no such set T can exist. □

Proposition 5.34. *Any two bases of a vector space V finitely generated over its field of scalars $\mathbb{F}$ have the same number of elements.*

Proof. By hypothesis and Proposition 5.31, we know that V has at least one finite basis S. If S' is another basis of V then, by Proposition 5.33, we see that

1. Since S is a generating set for V over $\mathbb{F}$ and S' is linearly independent, the number of elements of S' is at most that of S; and

2. Since S' is a generating set for V over $\mathbb{F}$ and S is linearly independent, the number of elements of S is at most that of S'.

Therefore S and S' have the same number of elements. □

Definition 5.35. Let V be a finitely generated vecto space over $\mathbb{F}$. The number of elements in a basis of V is called the **dimension** of V.

Theorem 5.36 (structure theorem for finite dimensional vector spaces)**.** *Let V be a finitely generated vector space over a field $\mathbb{F}$. Then there is an isomorphism of vector spaces $V \cong \mathbb{F}^n$.*

Proof. Choose a basis $\mathcal{B} = \{v_1, ..., v_n\}$ for V and define a function $f : V \longrightarrow \mathbb{F}^n$ by setting for $v \in V$:

1. if $v = v_i$ for some $1 \leq i \leq n$, $f(v_i) = e_i = (0, ..., 0, 1, 0, ..., 0)$

2. if $v \in V$ arbitrary, then there are unique scalars $a_1, ..., a_n \in \mathbb{F}$ such that $v = a_1 v_1 + ... + a_n v_n$ and we define $f(v) = a_1 f(v_1) + ... + a_n f(v_n) = (a_1, ..., a_n) \in \mathbb{F}^n$

The map f is linear by definition and has an obvious inverse: $f^{-1} : \mathbb{F}^n \longrightarrow V$ sends $(a_1, ..., a_n)$ to

$$f^{-1}(a_1, ..., a_n) = a_1 v_1 + ... + a_n v_n.$$

It is easy to check that $f \circ f^{-1} = \text{id}_{\mathbb{F}^n}$ and $f^{-1} \circ f = \text{id}_V$ so f is indeed an isomorphism. □

Theorem 5.36 has an immediate application to fields of positive characteristic. We saw that one such field is $\mathbb{F}_p$ and asked if there are other examples. Suppose $\mathbb{F}$ is a finite field of characteristic p. If $\alpha \in \mathbb{F}_p$ and $v \in \mathbb{F}$ we can define $\alpha \cdot v := \underbrace{v + v + ... + v}_{\alpha - times} \in \mathbb{F}$. It is easy to verify that this definition, makes $\mathbb{F}$ a vector space over $\mathbb{F}_p$ where addition of vectors is given by addition in $\mathbb{F}$. The vector space $\mathbb{F}$ is finite hence clearly finite dimensional and by Theorem 5.36 we have an isomorphism of vector spaces $\mathbb{F} \cong \mathbb{F}_p^n$. We thus get:

Corollary 5.37. *Let $\mathbb{F}$ be a finite field of positive characteristic p. Then there is an isomorphism of vector spaces $\mathbb{F} \cong \mathbb{F}_p^n$ for some $n \in \mathbb{N}$. In particular $|\mathbb{F}| = p^n$.*

Note that Corollary 5.37 is only a partial answer to Question 5.12 since we only have an isomorphism of vector spaces and not of fields.

5.3 Polynomials over fields

Corollary 5.37 tells us that a finite field of characteristic p must have p^n elements for some $n \in \mathbb{N}$. However, for $n > 1$, it is not clear that such a field actually exists and, if it does, what its explicit structure is. To answer this question, we will need to digress to a discussion on polynomials over a field. Polynomials are also a fundamental object in Cryptography so the material in this section is of more general interest.

Definition 5.38. Let $\mathbb{F}$ be a field. A **polynomial** (in one variable) over $\mathbb{F}$ is a formal expression of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

where $a_0, ..., a_n \in \mathbb{F}$ are called the **coefficients** of f. We assume that $a_n \neq 0$ and write $\deg f = n$. The set of all polynomials over $\mathbb{F}$ is denoted

$$\mathbb{F}[x] := \{f(x) = a_0 + a_1 x + ... + a_n x^n \mid 0 \leq n,\ a_0, ..., a_n \in \mathbb{F},\ a_n \neq 0\} \cup \{0\}.$$

The element $0 \in \mathbb{F}[x]$ is called the **zero polynomial** and by convention $\deg(0) = -\infty$.

Remark 5.39. By definition, two polynomials

$$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n$$

and

$$g(x) = b_0 + b_1 x + b_2 x^2 + ... + b_k x^k$$

are equal iff $n = k$ and for all i, $a_i = b_i$.

Example 5.40. Let $\mathbb{F} = \mathbb{F}_2 = \{0, 1\}$ be the field of two elements. Consider $f(x) = x^2$ and $g(x) = x$. Then $f(x) \neq g(x)$ although they define the same function $\mathbb{F}_2 \longrightarrow \mathbb{F}_2$: $0 \mapsto 0,\ 1 \mapsto 1$.

If

$$f(x) = \sum_{i=0}^{n} a_i x^i$$

and

$$g(x) = \sum_{i=0}^{k} b_i x^i$$

are two polynomials in $\mathbb{F}[x]$ where (without loss of generality) $\deg g = k \leq n = \deg f$ we can define their sum as

$$f(x) + g(x) = \sum_{i=0}^{k} (a_i + b_i) x^i + \sum_{i=k+1}^{n} a_i x^i$$

Example 5.41. Over $\mathbb{F} = \mathbb{R}$,

$$(4x^3 + 2x) + (5x^4 + 3x^3 + 2) = 5x^4 + 7x^3 + 2x + 2.$$

Similarly we can define

$$f(x) \cdot g(x) = \left(\sum_{i=0}^{n} a_i x^i\right)\left(\sum_{i=0}^{k} b_i x^i\right) = \sum_{i=0}^{n+k} c_i x^i$$

where

$$c_k = \sum_{i=0}^{k} a_i b_{k-i}$$

Example 5.42. Let $\mathbb{F} = \mathbb{F}_2$, $f(x) = x^2 - 1$ and $g(x) = x + 1$. Then

$$f(x)g(x) = (x^2 + x - 1)(x + 1) = x^3 + x^2 + x^2 + x - x - 1 = x^3 + 2x^2 - 1 = x^3 - 1.$$

where the last equality comes from the fact that $2 = 0$ in $\mathbb{F}_2$.

Although a polynomial $f(x) = \sum_i a_i x^i$ over a field $\mathbb{F}$ is a formal expression, every such polynomial defines a function $\mathbb{F} \longrightarrow \mathbb{F}$ given by $t \mapsto \sum_i a_i t^i$. We will be particularly interested in values t that are mapped to $0_{\mathbb{F}}$:

Definition 5.43. Let $\mathbb{F}$ be a field and $f = f(x) = \sum_i a_i x^i \in \mathbb{F}[x]$ a polynomial. An element $t \in \mathbb{F}$ is a called a **zero** (or root) of f if $\sum a_i t^i = 0$ (in $\mathbb{F}$).

Example 5.44. Let $f(x) = x^n - 1 \in \mathbb{C}[x]$ be a polynomial over the complex numbers. Then the set of all zeros of f is precisely the group μ_n of nth roots of unity. Note that when we think of f over the real numbers, i.e. $f(x) \in \mathbb{R}[x]$ the numbers of zeros changes: for example when $n = 3$ we have only one zero of f over $\mathbb{R}$ and 3 zeros of f over $\mathbb{C}$.

We have defined two binary operations

$$+ : \mathbb{F}[x] \times \mathbb{F}[x] \longrightarrow \mathbb{F}[x]$$

and

$$\cdot : \mathbb{F}[x] \times \mathbb{F}[x] \longrightarrow \mathbb{F}[x].$$

We note that the zero polynomial 0 can be viewed as a unit element with respect to + and the polynomial 1 can be viewed as a unit element with respect to $\cdot$. If $f(x) = \sum_i a^i x^i$ then $-f(x) := \sum_i -a_i x^i$ is clearly an additive inverse to $f(x)$ i.e. $f(x) + (-f(x)) = 0_{\mathbb{F}[x]}$. However, in general, there is no multiplicative inverse to $f(x)$: if $\deg f \geq 1$ then for any $g(x) \in \mathbb{F}[x]$ we have $\deg(f \cdot g) = \deg f + \deg g \geq 1$ whereas $\deg(1) = 0$ so there is no $g(x)$ such that $f(x)g(x) = 1$. Note that the data of $\mathbb{F}[x]$, the elements $0, 1 \in \mathbb{F}[x]$ and the binary operations $+, \cdot$ **almost form a field**: the only axiom not satisfied is that of a multiplicative inverse.

Remark 5.45. As can be easily seen, a non-zero polynomial $f(x) \in \mathbb{F}[x]$ is invertible iff $\deg f = 0$ (i.e. $f(x) = c$ for $c \in \mathbb{F}^\times$) and in that case we call $f(x)$ a **unit** in $\mathbb{F}[x]$: if $f(x) \in \mathbb{F}[x]$ is a polynomial of degree $1 \leq n$, then for any polynomial $g(x) \in \mathbb{F}[x]$, $\deg(f(x)g(x)) \geq n$ hence we cannot have $f(x)g(x) = 1$ for degree reasons.

Definition 5.46. A set R together with elements $0, 1 \in R$ and two binary operations

$$+ : R \times R \longrightarrow R$$

and

$$\cdot : R \times R \longrightarrow R$$

is called a **ring** if it satisfies all the axioms of a field in Definition 5.1 except (possibly) axiom (6) – existence of inverses wrt $\cdot$.

Examples 5.47.

1. Any field $\mathbb{F}$ is also a ring.

2. For $\mathbb{F}$ a field, $\mathbb{F}[x]$ is a ring.

3. The integers $\mathbb{Z}$ with ordinary addition and multiplication are a ring.

4. The integers modolu n, $\mathbb{Z}_n$ with addition and multiplication modolu n are a ring.

5.4 Euclidean Algorithm for Polynomials

Let $\mathbb{F}$ be a field and $f(x), g(x) \in \mathbb{F}[x]$ be non-zero polynomials with $n = \deg f \geq \deg g = k$. Write

$$f(x) = \sum_{i=0}^{n} a_i x^i$$

and

$$g(x) = \sum_{i=0}^{k} b_i x^i$$

with $a_n, b_k \in \mathbb{F}^\times$.

Set $n_0 = n$. Then for $c_0 := \frac{a_n}{b_k}$ we have

$$f(x) - c_0 x^{n_0 - k} g(x) = r_1(x) = (a_{n-1} - b_{k-1} c_1) x^{n-1} + ...$$

where $r_1(x)$ is a polynomial with $n_1 = \deg r_1 < n = n_0$. If $n_1 \geq k$ we can similarly write

$$r_1(x) - c_1 x^{n_1 - k} g(x) = r_2(x)$$

where $n_2 = \deg r_2 < n_1$ and $c_1 \in \mathbb{F}^\times$. Since the degrees of r_i's strictly decrease in each iteration, this process can last only finitely many steps s and we get a set of polynomials $\{r_i(x)\}_{i=1}^{s}$ and coefficients $\{c_i\}_{i=1}^{s-1} \subseteq \mathbb{F}^\times$ such that:

$$f(x) = c_0 x^{n_0 - k} g(x) + r_1(x),$$

$$r_1(x) = c_1 x^{n_1 - k} g(x) + r_2(x),$$

$$...$$

$$r_{s-1}(x) = c_{s-1} x^{n_{s-1} - k} g(x) + r_s(x)$$

where $\deg r_s < k$. Combining the above equations we get

$$f(x) = g(x) \left(c_0 x^{n_0 - k} + c_1 x^{n_1 - k} + ... + c_{s-1} x^{n_{s-1} - k} \right) + r_s(x).$$

Let us denote

$$q(x) = c_0 x^{n_0 - k} + c_1 x^{n_1 - k} + ... + c_{s-1} x^{n_{s-1} - k}$$

and $r(x) = r_s(x)$. We are ready to state:

Theorem 5.48. *Let $\mathbb{F}$ be a field and $f(x), g(x) \in \mathbb{F}[x]$ be non-zero polynomials with $\deg f \geq \deg g$. Then there exist unique polynomials $q(x), r(x) \in \mathbb{F}[x]$ such that*

$$f(x) = g(x) q(x) + r(x)$$

and with $\deg r < \deg g$.

Proof. The discussion above proved existence so it remains to prove uniqueness. Suppose $q'(x), r'(x)$ also satisfy

$$f(x) = g(x) q'(x) - r'(x)$$

with $\deg r' < \deg g$. Then

$$g(x) q(x) - r(x) = g(x) q'(x) - r'(x)$$

so that

$$g(x)\left(q(x) - q'(x)\right) = r'(x) - r(x)$$

But

$$\deg(r - r') < \deg g \le \deg(g(q - q'))$$

and to avoid contradiction we must have

$$q(x) - q'(x) = 0$$

after which it follows that $r'(x) - r(x) = 0$ and we are done. □

Terminology 1. *Under the notation of 5.48 we call q the* ***quotient*** *of f by g and call r the* ***remainder****. We also denote $f\%g = r$.*

The above terminology is meant to sharpen the analogy between modular arithmetic of integers and polynomials. Before diving into 'mod-$g(x)$ arithmetic', let us draw an important

Corollary 5.49. *Let $\mathbb{F}$ be a field and $f(x) \in \mathbb{F}[x]$. If $\alpha \in \mathbb{F}$ is a zero of $f(x)$ then there exists a polynomial $q(x) \in \mathbb{F}[x]$ such that $f(x) = (x - \alpha)q(x)$.*

Proof. Apply Theorem 5.48 with $g(x) = (x - \alpha)$ to get $q(x), r(x) \in \mathbb{F}[x]$ such that

$$f(x) = (x - \alpha)q(x) + r(x)$$

and $\deg r < \deg(x - \alpha) = 1$ so that $r(x) = c$ is a constant. Since α is a zero of f we get $r(\alpha) = 0$ so $r(x) = 0$ is the zero polynomial. □

Corollary 5.50. *Let $\mathbb{F}$ be a field. Then a polynomial $f(x) \in \mathbb{F}[x]$ of degree n has at most n roots.*

Proof. If f has $n + 1$ roots (or more) then a repeated application of Corollary 5.49 yields a decomposition

$$f(x) = \prod_{i=1}^{n+1}(x - \alpha_i)g(x)$$

and the right-hand-side has degree $\ge n + 1$ in contradiction to the fact that $\deg f = n$. □

We know that for every two distinct points in $\mathbb{R}^2$, we have a unique polynomial of degree 1 that passes through them. It is not hard to see that the formula for such a line is valid in any field. In fact, any $n + 1$ distinct points in $\mathbb{F} \times \mathbb{F}$ determine a unique polynomial in $\mathbb{F}[x]$ that 'passes' through them as the next result shows:

Theorem 5.51 (Lagrange interpolation polynomial)**.** *Let $\mathbb{F}$ be a field and $(x_1, y_1), ..., (x_{n+1}, y_{n+1}) \in \mathbb{F} \times \mathbb{F}$ a set of $n+1$ pairs of elements of $\mathbb{F}$ such that $\forall i \ne j : x_i \ne x_j$. Then there exists a unique polynomial $f(x) \in \mathbb{F}[x]$ of degree at most n such that $f(x_i) = y_i$ for $i = 1, ..., n + 1$.*

Proof. Fix $1 \le i \le n + 1$. We have $\prod_{j \ne i}(x_i - x_j) \ne 0$ since all factors are non-zero. Define a polynomial

$$D_i(x) = \frac{\prod_{j \ne i}(x - x_j)}{\prod_{j \ne i}(x_i - x_j)}.$$

Then $D_i(x_i) = 0$ and $D_i(x_j) = 1$ for every $j \ne i$.

We define

$$f(x) = \sum_{i=1}^{n+1} y_i D_i(x).$$

Then $f(x_i) = y_i \cdot 1 = y_i$ and f is of degree $\leq n$ as a sum of degree n polynomials. For uniqueness, suppose by contradiction that $g(x)$ is another polynomial of degree $\leq n$ such that $g(x_i) = y_i$ for $i = 1, ..., n+1$. Then $h(x) = f(x) - g(x)$ is a polynomial of degree $\leq n$ (as a sum of such) but for every i, $g(x_i) = f(x_i) - g(x_i) = y_i - y_i = 0$ so f has $n+1$ distinct roots in contradiction to Corollary 5.50. □

The discussion (or proof) preceding Theorem 5.48 gives in fact an algorithm to perform long division of polynomials.

Examples 5.52. Let $\mathbb{F} = \mathbb{Q}$.

1. Take $f(x) = 6x^3 - 2x^2 + x + 3$ and $g(x) = x^2 - x + 1$. Then long division $f(x)\%g(x)$ yields the following

$$
\begin{array}{rr}
 & 6x + 4. \\ \cline{2-2}
x^2 - x + 1\big) & 6x^3 - 2x^2 \quad + x + 3 \\
 & -6x^3 + 6x^2 - 6x \qquad\quad \\ \cline{2-2}
 & 4x^2 - 5x + 3 \\
 & -4x^2 + 4x - 4 \\ \cline{2-2}
 & -x - 1
\end{array}
$$

Thus
$$6x^3 - 2x^2 + x + 3 = (x^2 - x + 1)(6x + 4) - x - 1.$$

2. Take $f(x) = 2x^4 - x^2 + x + 3$ and $g(x) = x^2 - 1$. Then long division $f(x)\%g(x)$ yields the following

$$
\begin{array}{rr}
 & 2x^2 \qquad + 1. \\ \cline{2-2}
x^2 - 1\big) & 2x^4 \quad - x^2 + x + 3 \\
 & -2x^4 + 2x^2 \qquad\quad \\ \cline{2-2}
 & x^2 + x + 3 \\
 & -x^2 \qquad + 1 \\ \cline{2-2}
 & x + 4
\end{array}
$$

Thus
$$2x^4 - x^2 + x + 3 = (x^2 - 1)(2x^2 + 1) + x + 4.$$

Definition 5.53. Let $g(x) \in \mathbb{F}[x]$ be a polynomial of degree n. The set of possible **remainders** of mod-$g(x)$ division can be described as
$$\mathcal{R}_{\mathbb{F},n} = \left\{a_0 + a_1 x + ... + a_{n-1}x^{n-1} | a_0, ..., a_{n-1} \in \mathbb{F}\right\}$$
(note that here we allow all coefficients, including a_{n-1} to be 0).

Mod-$g(x)$ arithmetic is derived from polynomial arithmetic as follows. Let
$$r(x) = f(x) mod\, g(x)$$
and
$$s(x) = h(x) mod\, g(x).$$
Then there are quotient polynomials $q(x), t(x)$ such that
$$r(x) = f(x) - q(x)g(x)$$
and
$$s(x) = h(x) - t(x)g(x).$$

Thus

$$f(x) + h(x) = r(x) + s(x) - (q(x) + t(x))g(x)$$
$$f(x)h(x) = r(x)s(x) - (q(x)s(x) + t(x)r(x))g(x) + q(x)t(x)g(x)g(x).$$

It follows that

$$f(x) + h(x) = (r(x) + s(x))\, mod\, g(x)$$

and

$$f(x)h(x) = (r(x)s(x))\, mod\, g(x)$$

Observe that mod-$g(x)$ addition on $\mathcal{R}_{\mathbb{F},n}$ for

$$r(x) = a_0 + a_1 x + ... + a_{n-1}x^{n-1}$$

and

$$s(x) = b_0 + b_1 x + ... + b_{n-1}x^{n-1}$$

is given by adding coefficients 'component-wise':

$$(r(x) + s(x))\, mod\, g(x) = (a_0 + b_0) + (a_1 + b_1)x + ... + (a_{n-1} + b_{n-1})x^{n-1}$$

When $\mathbb{F} = \mathbb{F}_p$ is the field of p elements, we have $|\mathcal{R}_{\mathbb{F},n}| = p^n$ and the discussion above endows $\mathcal{R}_{\mathbb{F}_p,n}$ with the structure of a vector space that is isomorphic to $\mathbb{F}_p^n$. The addition of vectors arises from the addition of polynomials $mod\, g(x)$ and scalar multiplication arises from multiplying a polynomial by a scalar and taking $mod\, g(x)$ of the result.

Let $\mathbb{F}$ be a field and $g(x) \in \mathbb{F}[x]$ with $\deg g = n$. We will denote by $\mathbb{F}_{g(x)}$ the set $\mathcal{R}_{\mathbb{F},n}$ with mod-$g(x)$ arithmetic. From the discussion above, $\mathbb{F}_{g(x)}$ is a ring which is isomorphic as a vector space over $\mathbb{F}$ to $\mathbb{F}^n$. We are interested in polynomials $g(x)$ for which $\mathbb{F}_{g(x)}$ is a field.

Definition 5.54. Let $\mathbb{F}$ be a field. A polynomial $g(x)$ is **prime** (or irreducible) if it cannot be decomposed as $g(x) = a(x)b(x)$ for non-constant polynomials $a(x), b(x) \in \mathbb{F}[x]$.

Examples 5.55.

1. Clearly, any degree 1 polynomial $f(x) = x + a$ for $a \in \mathbb{F}$ is prime.

2. For $\mathbb{F} = \mathbb{F}_3$, the polynomial $g(x) = x^2 + 1$ is prime. If it wasn't we would have $g(x) = x^2 + 1 = a(x)b(x)$ where $\deg a = \deg b = 1$, i.e. $a(x) = \alpha_1 x + \alpha_0$ and $b(x) = \beta_1 x + \beta_0$ with $\alpha_1, \beta_1 \in \mathbb{F}_p^\times$. But then $a = -\frac{\alpha_0}{\alpha_1}$ is a root of g. However, we can easily see that $g(x)$ has no root over $\mathbb{F}$.

3. the polynomial $g(x) = x^2 + 1$ of the previous example is not prime over the field $\mathbb{F}_5$: since $3^2 + 1 = 0\ (mod\, 5)$ and $2^2 + 1 = 0\ (mod\, 5)$ we have $x^2 + 1 = (x - 3)(x - 2)$.

The terminology of Definition 5.54 is meant to suggest an analogy between numbers and polynomials. To flesh out the analogy, it will be convenient to restrict attention to **monic** polynomials i.e. polynomials of the form $x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$. Of course, any polynomial is a product of a monic polynomial and a non-zero field element. Let us start with

Proposition 5.56 (prime factorisation of polynomials)**.** *Let $f(x) \in \mathbb{F}[x]$ be a monic polynomial. Then there is $k \geq 1$ and monic prime polynomials*

$$a_1(x), ..., a_k(x) \in \mathbb{F}[x]$$

such that

$$f(x) = a_1(x) \cdot ... \cdot a_k(x).$$

Furthermore, this factorisation is unique up to re-ordering of the factors $a_i(x)$.

Proof. Clearly such factorisation must exist: if f is not prime we can write $f(x) = g(x)h(x)$ with $\deg g, \deg h < \deg f$ and continue factoring until we get monic prime factors. It remains to prove uniqueness. Suppose by contradiction that there is a polynomial with non-unique (monic) prime factorisation and let n be the minimal degree for which such polynomial $f(x)$ exist. Thus we can write $f(x)$ in two ways

$$a_1(x) \cdot ... \cdot a_k(x) = b_1(x) \cdot ... \cdot b_l(x) \tag{10}$$

where a_i, b_j are (monic) prime and $k, l \geq 1$. Now, $a_1(x)$ cannot appear on the right hand side of 10 since we could then cancel it from both sides and obtain contradiction to the minimality of n. Similarly, $b_1(x)$ cannot appear on the left hand side. Without loss of generality, assume $\deg b_1 \leq \deg a_1$. By Theorem 5.48 we can write $a_1(x) = b_1(x)q(x) + r(x)$. Since $a_1(x)$ is prime, $r(x) \neq 0$ and $\deg r < \deg b_1 \leq \deg a_1$. Now, $r(x)$ has a prime factorisation $r(x) = \alpha \cdot r_1(x) \cdot ... \cdot r_m(x)$ with $\alpha \in \mathbb{F}^\times$ and $b_1(x)$ cannot divide any of the $r_i(x)$ since it has a higher degree. Substituting that to 10 we get

$$(q(x)b_1(x) + \alpha r_1(x)...r_m(x))\, a_2(x)...a_k(x) = b_1(x)...b_l(x).$$

Define $f'(x) = r_1(x)...r_m(x)a_2(x)...a_k(x)$. Then f' is monic as a product of such and $\deg f' < \deg f$. However, we can write

$$f'(x) = r_1(x)...r_m(x)a_2(x)...a_k(x) = \alpha^{-1} b_1(x)\,(b_2(x)...b_l(x) - q(x)a_2(x)...a_k(x))$$

and these are two factorisations of f' in which the prime polynomial b_1 appear in one but not the other. This is a contradiction to the minimality of n. □

In light of Proposition 5.56 we can extend the analogy between polynomials and numbers with the following

Definition 5.57. Let $f(x), g(x)$ be monic polynomials. The **greatest common divisor** of f and g is the polynomial of maximal degree $\gamma(x)$ that divides both $f(x)$ and $g(x)$. We denote $\gamma = \gcd(f, g)$.

Remark 5.58. Note that the polynomial $\gamma = \gcd(f, g)$ of Definition 5.57 is unique by the uniqueness of factorisation to prime polynomials (Proposition 5.56).

Proposition 5.59. *For any field* $\mathbb{F}$, $\gcd(x^n - 1, x^m - 1) = x^{\gcd(n,m)} - 1$.

Proof. Without loss of generality, assume $n \leq m$ and use induction on m. If $m = 1$ then $n = 1$ and the claim is trivial. Suppose the claim is true for all $m' < m$. The case $n = m$ is trivial so we can further assume $n < m$. Then

$$x^m - 1 - x^{m-n}(x^n - 1) = x^{m-n} - 1$$

so that a polynomial g dividing both $x^m - 1$ and $x^n - 1$ must divide $x^{m-n} - 1$. By induction,

$$\gcd(x^{m-n} - 1, x^n - 1) = x^{\gcd(m-n,n)} - 1$$

but $\gcd(m - n, n) = \gcd(m, n)$ and

$$x^m - 1 = x^{m-n}(x^n - 1) + x^{m-n} - 1$$

so

$$\gcd(x^m - 1, x^n - 1) = \gcd(x^{m-n} - 1, x^n - 1) = x^{\gcd(m-n,n)} - 1 = x^{\gcd(m,n)} - 1.$$

□

5.5 Discrete Fourier Transform

In this section we will discuss an algorithm to efficiently multiply polynomials over finite fields, called the Discrete Fourier Transform (DFT). Fourier Analysis revolves around a notion of "convolution" of functions and is widely applicable in Signal Processing. The DFT algorithm, which is considered one of the most popular algorithms in applied mathematics, uses techniques from Fourier Analysis applied in an algebraic context.

Let $\mathbb{F} = \mathbb{F}_p$ be a prime field and $f(x), g(x) \in \mathbb{F}[x]$ be two polynomials. If we write

$$f(x) = \sum_{i=0}^{\deg f} f_i x^i \quad g(x) = \sum_{i=0}^{\deg g} g_i x^i$$

then the standard formula for their multiplication is

$$f(x)g(x) = \sum_{k=0}^{\deg f + \deg g} \sum_{i=0}^{k} f_i g_{k-i} x^i$$

which has complexity $\mathcal{O}(\deg f \cdot \deg g)$.

To improve complexity, suppose for simplicity that $f, g \in \mathbb{F}_{<n}[x]$ have degree less than n and that $\mathbb{F}$ contains a primitive nth root of unity w, ie $\mu_n = \langle w \rangle$. These assumptions are no loss of generality since, as we will see in Section 5.7, there is always an integer N such that $\mu_n \subseteq \mathbb{F}_{p^N}$ and since $\mathbb{F}_p \subseteq \mathbb{F}_{p^N}$ is a subfield inclusion, we can consider f, g as polynomials over $\mathbb{F}_{p^N}$. Note that our assumption means that $w^0, ..., w^{n-1}$ are n distinct elements in the base field $\mathbb{F}$.

We will represent a polynomial $f \in \mathbb{F}_{<n}[x]$ as a vector $\vec{f} = (f_0, ..., f_{n-1})$ where by convention, $f_k = 0$ for all $\deg f < k$. Thus, the map $\vec{-} : \mathbb{F}_{<n}[x] \longrightarrow \mathbb{F}^n$ gives an isomorphism of vector spaces and we will often interchange between these two representations.

Definition 5.60. The **Discrete Fourier Transform** of $f \in \mathbb{F}_{<n}[x]$ is the vector

$$\mathrm{DFT}_w(f) = (f(w^0), ..., f(w^{n-1})) \in \mathbb{F}^n.$$

Thus we can regard DFT as a map $\mathrm{DFT} : \mathbb{F}^n \longrightarrow \mathbb{F}^n$.

Next, consider the **Vandermonde** matrix

$$V_w = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w^1 & w^2 & \dots & w^{n-1} \\ 1 & w^2 & w^4 & \dots & w^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \dots & w^{(n-1)^2} \end{pmatrix}.$$

Lemma 5.61. *For any $f \in \mathbb{F}_{<n}[x]$, we have $\mathrm{DFT}_w(f) = V_w \cdot \vec{f}^T$.*

Proof. This is a direct computation. For example,

$$(V_w \vec{f}^T)_1 = f_0 + ... + f_{n-1} = f(w^0) = \mathrm{DFT}_w(f)_1$$

and

$$(V_w \vec{f}^T)_2 = f_0 \cdot 1 + f_1 \cdot w + ... + f_{n-1} w^{n-1} = f(w) = \mathrm{DFT}_w(f)_2.$$

□

Lemma 5.62. *The matrix V_w is invertible and $V_w^{-1} = \frac{1}{n} V_{w^{-1}}$.*

Proof. This is another direct computation, based on the fact that $1, w, w^2, ..., w^{n-1}$ is a geometric series hence its sum is given by the formula

$$S_n = 1 + w + ... + w^{n-1} = \frac{1 - w^n}{1 - w} = 0$$

since $w^n = 1$. □

Corollary 5.63. *The map $\mathrm{DFT}_w : \mathbb{F}^n \longrightarrow \mathbb{F}^n$ is invertible and $\mathrm{DFT}_w^{-1} = \frac{1}{n} \mathrm{DFT}_{w^{-1}}$.*

Proof. Follows from the two lemmas above. □

Next, we define a notion of convolution for our context.

Definition 5.64. Let $f, g \in \mathbb{F}_{<n}[x]$ be two polynomials.

1. The **convolution** of f and g, is defined as the polynomial

$$f \star g := fgmod\,(x^n - 1) \in \mathbb{F}_{<n}[x].$$

2. the **pointwise product** of $\vec{f}$ and $\vec{g}$ is the vector

$$\vec{f} \cdot \vec{g} := (f_0 g_0, ..., f_{n-1} g_{n-1}) \in \mathbb{F}^n.$$

A key property of the DFT is that it relates pointwise product with convolution as the following theorem states.

Theorem 5.65. *For $f, g \in \mathbb{F}_{<n}[x]$, we have*

$$\mathrm{DFT}_w(f \star g) = \mathrm{DFT}_w(f) \cdot \mathrm{DFT}_w(g).$$

Proof. By definition and Euclidean division 5.48, there is a polynomial $q \in \mathbb{F}_{<n}[x]$ such that

$$f \star g = fg + q(x^n - 1).$$

For every $0 \le i \le n - 1$ we have:

$$(f \star g)(w^i) = f(w^i)g(w^i) + g(w^i)(w^{in} - 1) = f(w^i)g(w^i)$$

where the last equality holds since $w^{in} = 1$ (as $w \in \mu_n$). The result now follows from the definiton of DFT_w. □

Corollary 5.66. *Suppose $f, g \in \mathbb{F}_{<n/2}[x]$ then:*

1. *$fg = f \star g$.*
2. *$f \star g = \mathrm{DFT}_w^{-1}(\mathrm{DFT}_w(f)\,\mathrm{DFT}_w(g))$.*
3. *$fg = \frac{1}{n} \mathrm{DFT}_{w^{-1}}(\mathrm{DFT}_w(f) \cdot \mathrm{DFT}_w(g))$.*

Corollary 5.67.

1. *Follows immediately from our assumption that $f, g \in \mathbb{F}_{<n/2}[x]$.*
2. *Follows from Theorem 5.65 and Corollary 5.63.*

3. follows from the previous two parts.

Let us now turn attention to the problem of finding a minimal extension $\mathbb{F}$ of $\mathbb{F}_p$ that contains a primitive nth root of unity w. From the existence of algebraic closure $\overline{\mathbb{F}}_p$ (see Section 5.7), we know that there exists $1 \leq N$ such that

$$w \in \mathbb{F}_{p^N} \iff \mu_n \subseteq \mathbb{F}_{p^N}.$$

If w, N are as above then $n = \operatorname{ord}_{\mathbb{F}^\times_{p^N}}(w)$ and by Lagrange theorem 4.20, $n \mid p^N - 1$ which is thus a necessary condition. The condition $n \mid p^N - 1$ is also sufficient: as $\mathbb{F}^\times_{p^N}$ is cyclic (Theorem 5.82), $\forall d \mid p^N - 1\ \exists x \in \mathbb{F}^\times_{p^N}$ with $\operatorname{ord}(x) = d$.

Of course, for efficiency, we would like to find the minimal N such that $n \mid p^N - 1$. In order to get a more compact formulation, let us digress a bit to talk about

Definition 5.68. Let $1 \leq n$ be an integer. The **multiplicative group of integers modulo** n is

$$\mathbb{Z}_n^\times = \{x \in \mathbb{Z}_n | \exists d : x^d = 1(mod\, n)\}.$$

In other words, $\mathbb{Z}_n^\times$ is the subset of $\mathbb{Z}_n$ consisting of all elements that have a multiplicative inverse. Clearly, $\mathbb{Z}_n^\times$ is an abelian group and $\mathbb{Z}_n^\times = \mathbb{Z}_n$ iff n is prime.

Lemma 5.69.

$$\mathbb{Z}_n^\times = \{x \in \mathbb{Z}_n | \gcd(x, n) = 1\}$$

so

$$|\mathbb{Z}_n^\times| = \varphi(n)$$

where $\varphi(n)$ is Euler's totient function 4.59.

Proof. By Euclidean division $\gcd(x, n) = 1$ iff there are y, q such that $xy + qn = 1 \iff xy = 1(mod\, n)$. □

Coming back to DFT, our assumptions are that $n = 2^d$ and p is a prime. If we denote $\overline{p} = p(mod\, n)$ then by Lemma 5.69 $\overline{p} \in \mathbb{Z}_n^\times$. Then, $N = \operatorname{ord}_{\mathbb{Z}_n^\times}(\overline{p})$ is the minimal integer such that $p^N - 1 = (mod\ n)$. Let us record this for future reference

Corollary 5.70. *Let p be a prime and n an integer with $\gcd(p, n) = 1$. Then the minimal field extension $\mathbb{F}_p \subseteq \mathbb{F}_{p^N}$ containing an nth root of unity is obtained for $N = \operatorname{ord}_{\mathbb{Z}_n^\times}(p)$.*

We are ready to state

Theorem 5.71. *Algorithm 5.5 outputs* $\mathrm{DFT}_w(f)$.

Proof. We need to verify that the kth entry of the output is $f(w^k)$ for all $0 \leq k < n$. For the even values $k = 2i$, we have

$$\begin{aligned} f(w^{2i}) &= g(w^{2i}) + (w^{2i})^{n/2} \cdot h(w^{2i}) \\ &= g(w^{2i}) + h(w^{2i}) = r(w^{2i}). \end{aligned} \tag{11}$$

For the odd values $k = 2i + 1$:

$$\begin{aligned} f(w^{2i+1}) &= \sum_{0 \leq j < n/2} f_j w^{2i+1} j + \sum_{0 \leq j < n/2} f_{n/2+j} w^{(2i+1)(n/2+j)} \\ &= \sum_{0 \leq j < n/2} g_j w^{2ij} w^j + \sum_{0 \leq j < n/2} h_j w^{2ij} w^{in} w^{n/2} w^j \\ &= \sum_{0 \leq j < n/2} (g_j - h_j) w^j w^{2ij} = \sum_{0 \leq j < n/2} s_j w^{2ij} = s(w^{2i}). \end{aligned} \tag{12}$$

□

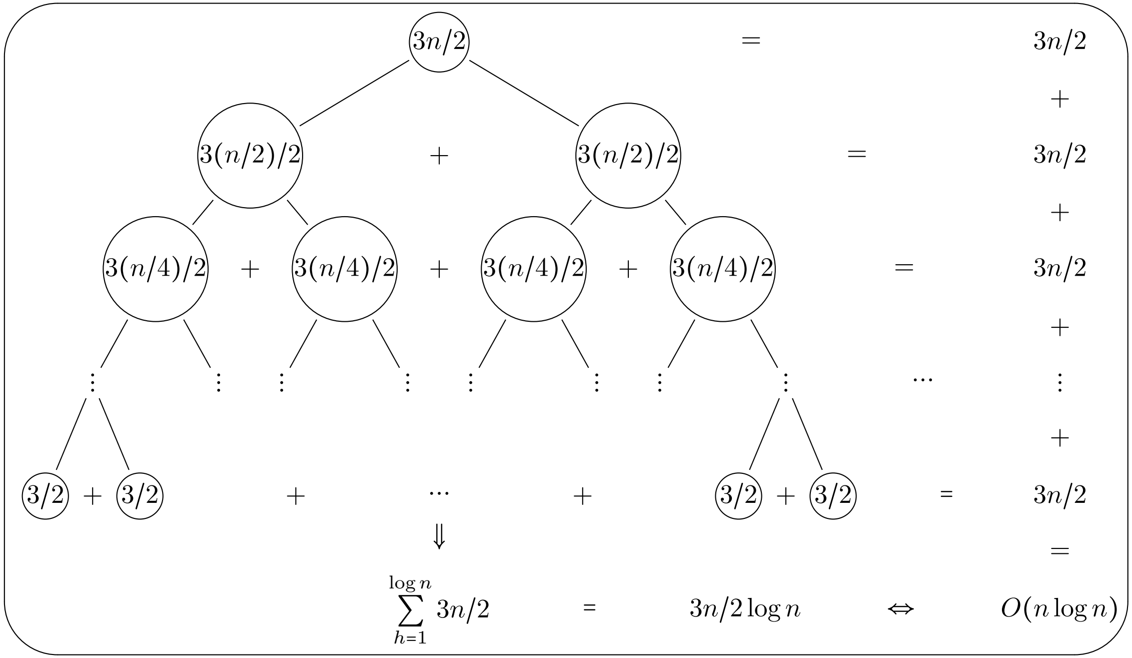
Algorithm 1 Algorithm: Discrete Fourier Transform (DFT) for polynomials over a finite field.

```
Input: a prime field $\mathbb{F}_p$, an integer $n = 2^d$, a finite field extension $\mathbb{F} = \mathbb{F}_{p^N}$ containing $n$th root of unity $w$, and vectors $\vec{f} = (f_0, ..., f_{n-1}) \in \mathbb{F}_p^n \leftrightarrow f \in \mathbb{F}_p^{<n}[x]$ and $(w^0, w^1, ..., w^{n-1}) \in \mathbb{F}^n$.
Output: $\mathrm{DFT}_w(f)$.
If $n = 1$ return $(f_0)$.
Write $f(x) = g(x) + x^{n/2}h(x)$ where $g, h \in \mathbb{F}_p^{<n/2}[x]$.
Compute the vectors in $\mathbb{F}^n$ $\vec{r} = \vec{g} + \vec{h}$ $\vec{s} = (\vec{g} - \vec{h})(w^0, ..., w^{n/2-1})$
Recursively compute $\mathrm{DFT}_{w^2}(r), \mathrm{DFT}_{w^2}(s)$
return
    $(r(w^0), s(w^0), r(w^2), s(w^2), ..., r(w^{n/2}), s(w^{n/2})) \in \mathbb{F}^n.$
```

Lastly, let us analyse the complexity of the DFT algorithm above. Let $T(n)$ be the number of field operations required to compute $\mathrm{DFT}_w(f)$. Then $T(n)$ satisfies the relation

$$T(n) = 2T(n/2) + 3n/2.$$

To find a closed formula for $T(n)$, let us draw the corresponding **recursion tree**:


$3n/2$ = $3n/2$
+
$3(n/2)/2$ + $3(n/2)/2$ = $3n/2$
+
$3(n/4)/2$ + $3(n/4)/2$ + $3(n/4)/2$ + $3(n/4)/2$ = $3n/2$
+
⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋮ ⋯ ⋮
+
$3/2$ + $3/2$ + ⋯ + $3/2$ + $3/2$ = $3n/2$
⇓ =
$\sum_{h=1}^{\log n} 3n/2$ = $3n/2 \log n$ ⇔ $O(n \log n)$

The tree is of height $\log n$ and each level requires $3n/2$ operations, so $T(n) = O(n \log n)$. Note that this is a substantial complexity improvement over the naive way to multiply polynomials, which is $O(n^2)$.

5.6 Classification of finite fields

We are now ready to state

Theorem 5.72. *Let $\mathbb{F}$ be a field and $g(x)$ a prime polynomial in $\mathbb{F}[x]$ with $\deg g = n$. Then $\mathbb{F}_{g(x)}$ together with mod-$g(x)$ arithmetic is a field.*

The proof of Theorem 5.72 is remarkably similar to the proof that $\mathbb{F}_p$ is a field: using the extended Euclidean algorithm. We thus need a version of the extended Euclidean algorithm for polynomials:

Theorem 5.73. *Let $\mathbb{F}$ be a field and $a(x), b(x) \in \mathbb{F}[x]$ be polynomials. Then there exist polynomials $s(x), t(x), f(x)$ such that*

$$a(x)s(x) + b(x)t(x) = f(x)$$

where $f(x) = \gcd(a(x), b(x))$.

The proof of Theorem 5.73 in turn, is remarkably similar to that of the Extended Euclidean Algorithm for integers, using Theorem 5.48 and we will omit it.

Proof of Theorem 5.72. As discussed above, we only need to show multiplicative inverse. Let $r(x) \in \mathbb{F}_{g(x)}$. Using Theorem 5.73 with $a(x) = r(x)$ and $b(x) = g(x)$ we get polynomials s, t, γ such that

$$r(x)s(x) + g(x)t(x) = \gamma(x).$$

Since $g(x)$ is prime, $1 = \gamma(x) = \gcd(r(x), g(x))$ so

$$r(x)s(x) = 1 mod\, g(x).$$

□

Corollary 5.74. *Let $\mathbb{F} = \mathbb{F}_p$. Then every prime polynomial $g(x) \in \mathbb{F}_p$ of degree n gives rise to a field with p^n elements.*

Proof. The set $\mathbb{F}_{g(x)}$ with mod-$g(x)$ arithmetic is a field by Theorem 5.72. By construction, the set $\mathbb{F}_{g(x)}$ is in one-to-one correspondence with the set of remainder polynomials $\mathcal{R}_{\mathbb{F},n} = \{r(x) = a_0 + a_1x + ... + a_{n-1}x^{n-1} | a_0, ..., a_{n-1} \in \mathbb{F}_p\}$ whose size is p^n. □

Example 5.75. Let us construct a field with $2^2 = 4$ elements. We first observe that the polynomial $g(x) = x^2 + x + 1$ is prime over $\mathbb{F}_2$ since it has no roots. Thus, $\mathbb{F}_{g(x)}$ is a field with $2^{\deg g} = 4$ elements. There are four possible remainder polynomials $\{0, 1, x, x + 1\}$. Addition is componentwise *mod* 2. For multiplication, note that $x * x = x^2 = x + 1$ $(mod\, x^2 + x + 1)$. Also, $x * x * x = x^3 = 1$ $(mod\, x^2 + x + 1)$ so the three non-zero elements $\{1, x, x + 1\}$ form a cyclic group under mod-$g(x)$ multiplication.

The complete mod-$g(x)$ addition and multiplication tables are given as follows.

$+$	0	1	x	$x+1$
0	0	1	x	$x+1$
1	1	0	$x+1$	x
x	x	$x+1$	0	1
$x+1$	$x+1$	x	1	0

$\times$	0	1	x	$x+1$
0	0	0	0	0
1	0	1	x	$x+1$
x	0	x	$x+1$	1
$x+1$	0	$x+1$	1	x

Consider now a prime field $\mathbb{F} = \mathbb{F}_p$ and a prime polynomial $g(x) \in \mathbb{F}[x]$. Observe that $\mathbb{F}_{g(x)} = \{a_0 + a_1x + ... + a_{n-1}x^{n-1} | a_0, ..., a_{n-1}\}$ contains a "copy" of $\mathbb{F}$ given by the constants $\{a_0 | a_0 \in \mathbb{F}\}$ so we may view the field $\mathbb{F}_{g(x)}$ as an extension of $\mathbb{F}$. Let us make it precise

Definition 5.76. Let $\mathbb{F}$ be a field and $\mathbb{F}'$ a field containing an isomorphic copy of $\mathbb{F}$. Then we call $\mathbb{F}'$ an **extension** of $\mathbb{F}$. We say that the degree of the extension is $[\mathbb{F} : \mathbb{F}'] = \dim_{\mathbb{F}} \mathbb{F}'$ where dim is the dimension of $\mathbb{F}'$ as a vector space over $\mathbb{F}$.

Since $\mathbb{F}_{g(x)}$ contains a copy of $\mathbb{F}$ and the coefficients of g are all in $\mathbb{F}$, we may view $g(x)$ as a polynomial over $\mathbb{F}_{g(x)}$, and we denote $g(X) \in \mathbb{F}_{g(x)}[X]$. With this in mind, we have:

Corollary 5.77. *Let $\mathbb{F} = \mathbb{F}_p$ be a prime field and $g(x) \in \mathbb{F}[x]$ be a prime polynomial. Then $\mathbb{F}_{g(x)}$ is a field extension of $\mathbb{F}$ and $[\mathbb{F}_{g(x)} : \mathbb{F}] = \deg g$. Moreover, the polynomial $g(X) \in \mathbb{F}_{g(x)}[X]$ admits a root in $\mathbb{F}_{g(x)}$.*

Proof. The first part is just a restatement of the discussion above. For the second part, observe that $\mathbb{F}_{g(x)} \cong \mathbb{F}^n$ as a vector space since we can identify a remainder polynomial $r(x) = a_0 + a_1x + ... + a_{n-1}x^{n-1}$ with an n-tuple $(a_0, ..., a_{n-1}) \in \mathbb{F}^n$. For the third part, consider the remainder polynomial $r(x) = x \in \mathbb{F}_{g(x)}$. Then when we substitute $r(x)$ in $g(X)$ we get $g(x)$ which is equal to 0 in $\mathbb{F}_{g(x)}$. Thus, $r(x)$ is a root of $g(X)$. □

Remark 5.78. The field $\mathbb{F}_{g(x)}$ is denoted in the literature as $\mathbb{F}_p[x]/\langle g(x)\rangle$, and the notation comes from viewing $\langle g(x)\rangle$ as n ideal in the ring $\mathbb{F}_p[x]$. We chose to avoid ideals and ring quotients in these notes to keep things on an elementary level.

In order to prove the existence of finite fields of order p^n we need to show that there exist a prime polynomial over $\mathbb{F}_p$ of degree n. It is easy to show a slightly weaker statement:

Proposition 5.79. *Let p be a prime. Then for any $N \in \mathbb{N}$ there exist $n > N$ and a prime polynomial over $\mathbb{F}_p$ of degree n.*

Proof. Assume by contradiction that there is $N \in \mathbb{N}$ for which the statement is false, i.e there is no prime polynomial of degree $\geq N$. Note that the set of polynomials of degree $\leq N$ in $\mathbb{F}_p[x]$ is finite since $\mathbb{F}_p$ has only finitely many elements. It follows that the set of all (monic) prime polynomials over $\mathbb{F}_p$ is finite, say, $\{p_1(x), ..., p_k(x)\}$. Consider the polynomial $p(x) = 1 + p_1(x)...p_k(x)$. Clearly $p_i(x) \nmid p(x)$ for all $1 \leq i \leq k$ (since the remainder is 1) but this is a contradiction to the prime factorisation of Proposition 5.56. □

Exercise 5.80. Describe a field with $3^2 = 9$ elements together with its addition and multiplication tables.

We have proved via Theorem 5.72 and Proposition 5.79 that for any prime p there exist finite fields of characteristic p of arbitrary size. For $n = 1$ we showed that such a field is unique up to isomorphism. Can we say the same for $n > 1$? The answer lies in the following

Theorem 5.81 (Structure theorem for finite fields)**.** *Let p be prime. Then for any $n \in \mathbb{N}$ there exist a field of characteristic p with exactly $q = p^n$ elements. Moreover, such a field is unique up to isomorphism and denoted $\mathbb{F}_q$.*

The proof of Theorem 5.81 is beyond the scope of these notes. Let us finish this section with a result that may be useful for future considerations.

Theorem 5.82. *Let p be prime and $q = p^k$. The multiplicative group of the field of q elements $G = \mathbb{F}_q^\times$ is cyclic.*

Proof. We denote $n = q - 1$ so that $|G| = n$. By Lagrange Theorem 4.20, the order d of an element $a \in G$ must divide n since $< a >$ is a subgroup of G of order d. Let d be a divisor of n and denote by $\psi(d)$ the number of elements in G of order d. If $a \in G$ is of order d, then $H =< a >= \{1, a, ..., a^{d-1}\} \leq G$ is a cyclic group of order d so that every $x \in H$ satisfies $x^d = 1$. Since $\mathbb{F} = \mathbb{F}_q$ is a field, the polynomial $f(x) = x^d - 1$ can have at most d roots and it follows that it has exactly d roots – the elements of H. Note that not all elements in H have order d. Rather, the number of elements of H of order d are precisely its generators. By Corollary 4.61 the number of such generators is precisely $\varphi(d)$ and it follows that $\psi(d) = 0$ or $\psi(d) = \varphi(d)$. By Theorem 4.62 we have

$$n = \sum_{d:d|n} \varphi(d)$$

and since G is a finite group, we have

$$n = \sum_{d:d|n} \psi(d).$$

Thus,

$$\sum_{d:d|n} \varphi(d) = \sum_{d:d|n} \psi(d)$$

and it follows that $\psi(n) = \varphi(n)$ so G has at least one generator of order n. □

5.7 Algebraic Closure

Let p be prime, $q = p^n$. If $\beta \in \mathbb{F}_q^\times$ then $\langle\beta\rangle \leq \mathbb{F}_q^\times$ is a subgroup of the multiplicative field of $\mathbb{F}_q$ so that $|\langle\beta\rangle| \mid q-1$. It follows that $\beta^{q-1} = 1$ in $\mathbb{F}_q$ ie β is a root of the polynomial $f(x) = x^q - x$ over $\mathbb{F}_q$. Since f can have at most q roots, it follows that the elements of $\mathbb{F}_q$ are precisely the roots of f and we can thus write f as a product of distinct linear factors

$$f(x) = \prod_{\beta\in\mathbb{F}_q} (x - \beta). \tag{13}$$

Let p be a prime, $q = p^n$ and consider the field $\mathbb{F}_q$ given by Theorem 5.81. We revoke a small result from elementary Number Theory

Lemma 5.83. *If p is prime and $m \mid n$ then $p^m - 1 \mid p^n - 1$.*

Theorem 5.84. *Let $q = p^n$. Then for every $m \mid n$ there is a unique subfield of $\mathbb{F}_q$ given by the roots of $x^{p^m} - x$ over $\mathbb{F}_q$. Moreover, every subfield of $\mathbb{F}_q$ is obtained in this form.*

Proof. Let us start with the last statement. If $K \subseteq \mathbb{F}_q$ is a subfield then $\mathbb{F}_q$ is a vector space over K which is finite dimensional (since $\mathbb{F}_q$ is finite). Hence, by Theorem 5.36 we have an isomorphism of vector spaces over K, $\mathbb{F}_q \cong K^d$. It follows that $|K| \mid |\mathbb{F}_q| = p^n$ so that $|K| = p^m$ with $m \mid n$. Let us turn to uniqueness. If $K \subseteq \mathbb{F}_q$ is a subfield of order p^m then $K^\times$ is a cyclic group of order $p^m - 1$ hence its elements are roots of $x^{p^m-1} - 1$ over $\mathbb{F}_q$. Thus the elements of K must be precisely the roots of $x^{p^m} - x$. Lastly, we prove existence. Let $m \mid n$. We know from Equation 13 that the elements of $\mathbb{F}_q^\times$ can be identified with the roots of the polynomial $x^{p^n-1} - 1$. By Lemma 5.83 $p^m - 1 \mid p^n - 1$, and by Proposition 5.59 the polynomial $x^{p^m-1} - 1$ divides $x^{p^n-1} - 1$ in $\mathbb{F}_p[x]$ hence $x^{p^m-1} - 1$ decomposes to linear factors over $\mathbb{F}_q$. Let K be the set of roots of $x^{p^m} - x$ over $\mathbb{F}_q$, or the set of roots of $x^{p^m-1} - 1$ together with 0, so that $|K| = p^m$. Clearly, the elements of K are closed under multiplication since $(\alpha\beta)^{p^m-1} = \alpha^{p^m-1}\beta^{p^m-1} = 1 \cdot 1 = 1$. Similarly, the elements of K are closed under inverse. To see clossness to addition, observe that $(\alpha+\beta)^{p^m} = (\alpha+\beta)$ for any field of characteristic p by Lemma 4.10. Thus, $K \subseteq \mathbb{F}_q$ is indeed a subfield. □

Theorem 5.84 allows us to identify $\mathbb{F}_{p^m}$ as a subfield of $\mathbb{F}_{p^n}$ whenever $m \mid n$. In particular, we have a sequence of subfield inclusions

$$\mathbb{F}_p = \mathbb{F}_{p^{1!}} \subseteq \mathbb{F}_{p^{2!}} \subseteq \ldots \subseteq \mathbb{F}_{p^{n!}} \subseteq \ldots$$

This in turn, enables us to give the following

Definition 5.85. Let $\mathbb{F}$ be a field with $\operatorname{char}\mathbb{F} = p$. The **algebraic closure** of $\mathbb{F}$, denoted $\overline{\mathbb{F}}$ is the union

$$\overline{\mathbb{F}} = \bigcup_{n \in \mathbb{N}} \mathbb{F}_{p^{n!}}$$

of the subfield inclusions described above.

Proposition 5.86. *Let $\mathbb{F}$ be a field of characteristic p. Then $\overline{\mathbb{F}}$ admits a structure of a field.*

Proof. Note first that by Theorem 5.84, any field $\mathbb{F}_{p^n}$ can be identified as a subfield of $\mathbb{F}_{p^{n!}}$ (since $n \mid n!$) and hence is contained in $\overline{\mathbb{F}}$. In addition for every n, $\mathbb{F}_p$ is a subfield of $\mathbb{F}_{p^{n!}}$. Thus we define the neutral elements by $0, 1 \in \mathbb{F}_p$. To define addition and multiplication let $x, y \in \overline{\mathbb{F}}$. Then there are n, k such that $x \in \mathbb{F}_{p^{n!}}$ and $y \in \mathbb{F}_{p^{k!}}$. But then, $x, y \in \mathbb{F}_{p^{n!k!}}$ and we define their addition and multiplication to be the one in $\mathbb{F}_{p^{n!k!}}$. If $0 \neq x \in \overline{\mathbb{F}}$ then $x \in \mathbb{F}_{p^{n!}}$ and thus $-x$ and x^{-1} are defined in $\overline{\mathbb{F}}$ as they are in $\mathbb{F}_{p^{n!}}$. Associativity, commutativity and distributivity are satisfied in all $\mathbb{F}_{p^{n!}}$ hence also in $\overline{\mathbb{F}}$. □

Let $\mathbb{F}$ be a field and $g(x) \in \mathbb{F}[x]$ a prime polynomial. The set $\mathbb{F}_{g(x)}$ with mod-$g(x)$ addition and multiplication is a field by Theorem 5.72 and contains $\mathbb{F}$ as the set of constant remainder polynomials. Thus, we can consider g as a polynomial over $\mathbb{F}_{g(x)}$. Consider the quotient map $\pi : \mathbb{F}[x] \longrightarrow \mathbb{F}_{g(x)}$ given by $\pi(f(x)) = f(x) mod\ g(x)$. Note that when g is considered as a polynomial over $\mathbb{F}_{g(x)}$, $g(\pi(x)) = \pi(g(x)) = 0 mod\ g(x)$ so that $\pi(x)$ is a root of g. In light of this, we are ready to prove the main property of the algebraic closure:

Theorem 5.87. *Let $g(x) \in \overline{\mathbb{F}}_p[x]$ be a polynomial of degree n. Then $g(x)$ decomposes to linear factors*

$$g(x) = \prod_{i=1}^{n} (x - \alpha_i)$$

for $\alpha_i \in \overline{\mathbb{F}}_p$.

Proof. It is enough to prove the theorem for prime polynomials, so assume g is prime. Since g has finitely many coefficients, they all must lie in some field $K = \mathbb{F}_q$. By the discussion above, when we consider g as a polynomial over $K_{g(x)}$ we get $g(x) = (x - \alpha)g_1(x)$ for some polynomial $g_1(x)$ over $K_{g(x)}$ with $\deg g_1 = \deg g - 1$. But $K_{g(x)} \cong \mathbb{F}_{q^n}$ hence is contained in $\overline{\mathbb{F}}_p$. Thus, as polynomials over $\overline{\mathbb{F}}_p$ we get a decomposition $g(x) = (x - \alpha)g_1(x)$. Repeat this procedure for $g_1, g_2, \ldots$ to decompose $g(x)$ to linear factors. □

In light of Theorem 5.87, we can decompose the polynomial $x^n - 1$ into linear factors $x^n - 1 = \prod_{i=1}^{n}(x - \alpha_i)$ over $\overline{\mathbb{F}}_p$. The considerations in Example 4.6 (5), apply in this case as well so that the collection of all distinct roots of $x^n - 1$ forms a group under multiplication.

Definition 5.88. Let p be prime. The group of n**th roots of unity** over $\overline{\mathbb{F}}_p$, denoted $\mu_n = \mu_n(\mathbb{F}_p)$ is the set of roots of the polynomial $x^n - 1$ in $\overline{\mathbb{F}}_p$.

Since we are in positive characteristic, the order of $\mu_n(\overline{\mathbb{F}}_p)$ may be different than n.

Proposition 5.89. *If $p \nmid n$, then the group of nth roots of unity μ_n is cyclic of order n.*

Proof. The case $n = 1$ is trivial so suppose $n \geq 2$. Since $p \nmid n$, the polynomial $x^n - 1$ and its derivative nx^{n-1} have no common roots so the linear factors in the decomposition $x^n - 1 = \prod_{i=1}^{n}(x - \alpha_i)$ are all distinct so that $|\mu_n(\overline{\mathbb{F}}_p)| = n$. Moreover, $\mu_n(\overline{\mathbb{F}}_p)$ is a subgroup of $\mathbb{F}_q^\times$ for some q and the latter is cyclic by Theorem 5.82. By Proposition 4.63, a subgroup of a cyclic group must by cyclic and the result follow. □

Throughout this section, we developed the algebraic closure of finite fields. In fact, algebraic closure is a construction that exists for any field and is characterized by the property of Theorem 5.87, ie. that over an algebraic closure any polynomial splits to linear factors. For the sake of completeness, we record

Theorem 5.90 (fundamental theorem of algebra)**.** *The algebraic closure of the real numbers $\mathbb{R}$ are the complex numbers $\mathbb{C}$, i.e. $\overline{\mathbb{R}} = \mathbb{C}$. In particular, every polynomial $f(x) \in \mathbb{C}[x]$ admits a decomposition*

$$f(x) = \prod_i (x - \alpha_i).$$

Proof. We will not use this theorem much since we mostly work over finite fields. A proof can be found, for example, in [FR]. □

6 Elliptic Curves

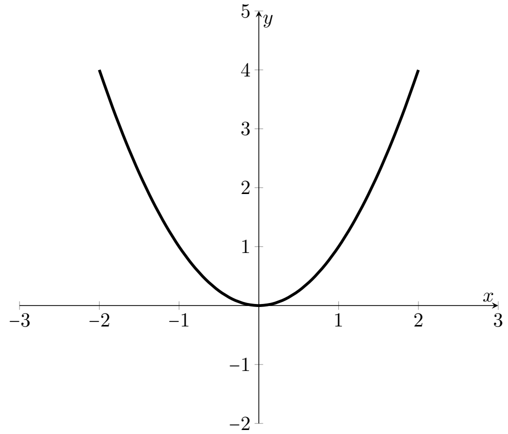
Let $\Bbbk$ be a field. We saw that a polynomial in one variable $f(x) \in \Bbbk[x]$ must have finitely many roots. We can however consider polynomials in more variables

Definition 6.1. A **polynomial in two variables** x, y over a field $\Bbbk$ is a formal expression of the form

$$f(x,y) = a_{0,0} + a_{1,0}x + a_{0,1}y + a_{1,1}xy + a_{2,1}x^2y + ... + a_{n,k}x^ny^k = \sum_{0\leq i\leq n, 0\leq j\leq k} a_{i,j}x^iy^j$$

where $a_{n,k} \neq 0$. Here we define $\deg f = n + k$. We will denote the set of all such polynomials by $\Bbbk[x,y]$. A **root** (or **zero**) of $f(x,y)$ is a point $(x_0, y_0) \in \Bbbk^2$ such that $f(x_0, y_0) = 0$. The set of all such roots is called the **solution set** of f.

Example 6.2. For $\Bbbk = \mathbb{R}$ we can take $f(x,y) = y - x^2$. The set of roots of f can be depicted as the parabola in the $X - Y$ plane $y = x^2$:



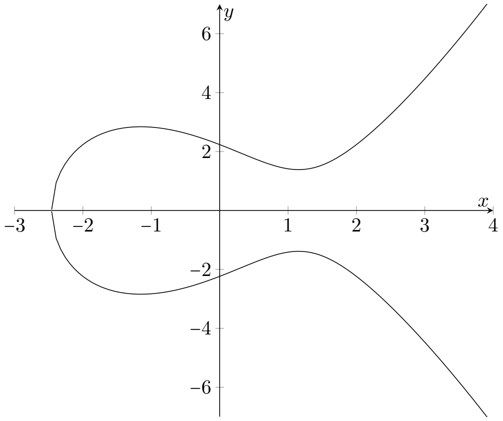


Figure 1: An elliptic curve defined over $\mathbb{R}$ by the equation $y^2 = x^3 - 4x + 5$

Example 6.2 indicates a general phenomenon. When our base field $\Bbbk$ has some geometry (as is the case with $\Bbbk = \mathbb{R}$ or $\Bbbk = \mathbb{C}$), the solution set of a polynomial in two variables has a geometry as well – that of a one dimensional surface, aka a curve. Thus we can expect that the solution set of an n-dimensional polynomial will have a geometry of an $(n-1)$-dimensional surface. It turns out we can use the geometry over $\mathbb{R}$ and $\mathbb{C}$ to get intuition on solution sets of polynomials over finite fields! In modern Mathematics, this approach is usually referred to as **Algebraic Geometry**. We will be restricting attention to solution sets of polynomials in two variables. More specifically

Definition 6.3. Let $\Bbbk$ be a field of characteristic $\neq 2, 3$. An **elliptic curve** E defined over $\Bbbk$ (denoted $E/\Bbbk$) is a polynomial of the form

$$E : y^2 - x^3 - Ax - B \tag{14}$$

where $A, B \in \Bbbk$ satisfy

$$\delta(E) := 4A^3 + 27B^2 \neq 0.$$

Notation 6.4. We will sometime write $E_{A,B}$ to denote an elliptic curve of the form $E/\Bbbk : y^2 = x^3 + Ax + B$.

Remark 6.5. Henceforth, unless explicitly mentioned otherwise, we will assume that our base field $\Bbbk$ has $\operatorname{char} \Bbbk \neq 2, 3$.

Remark 6.6. The form appearing in 14 is called the **short Weierstrass form**. When $\operatorname{char} \Bbbk = 2$ or $\operatorname{char} \Bbbk = 3$ the definition of an elliptic Curve over $\Bbbk$ is expressed in the **long Weierstrass form**:

$$E : y^2 + a_1xy + a_2y = x^3 + a_3x + a_4$$

such that $\Delta(E) \neq 0$. Note that here $\Delta(E)$ is the discriminant of E which has a slightly different formula than in the case of the short Weirstrass form.

The condition

$$4A^3 + 27B^2 \neq 0$$

of Definition 6.3 might seem odd but in fact is equivalent to a geometric property:

Proposition 6.7. *Let $E : y^2 = x^3 + Ax + B$ be a polynomial equation defined over $\mathbb{R}$. Then E has a well-defined and unique tangent line at every point iff E is an elliptic curve, ie $4A^3 - 27B^2 \neq 0$.*

Proof. The solution set of E can be depicted as a curve in $\mathbb{R}$ and this curve in turn can be expressed as the union of graphs of two functions $y = \pm\sqrt{x^3 + Ax + B}$. When $y \neq 0$, a tangent line with respect to E must have a slope given by the condition $\frac{dy}{dx} = 0$ ie

$$\pm\frac{3x^2 + A}{2\sqrt{x^3 + Ax + B}} = 0.$$

Thus, we have a well-defined tangent line when $y \neq 0$ and $x^3 + Ax + B > 0$. If $x^3 + Ax + B < 0$ then there is no value y such that (x, y) is in the solution set; this leaves us with the case $x^3 + Ax + B = 0$. For this case we take a different method, namely that of implicit differentiation $\frac{d}{dx}$ of both sides:

$$\frac{d(y^2)}{dx} = 2y\frac{dy}{dx} = 3x^2 + A \iff \frac{dy}{dx} = \frac{3x^2 + A}{2y}.$$

Suppose $3x^2 + A \neq 0$ but $y = 0$. Then we can reflect our curve along the line $y = x \subseteq \mathbb{R}^2$ which means, algebraically, that we interchange $y \leftrightarrow x$, ie the nominator being zero and the denominator being non-zero – resulting in a tangent line of slope 0. Thus, flipping x, y again we get a tangent line parallel to the y-axis. Lastly, suppose $y = 0$ and $3x^2 + A = 0$. Then necessarily $A < 0$ and we can substitute $A = -A$ to rewrite the equation as $y^2 = x^3 - Ax + B$ with $A > 0$. In that case, $3x^2 - A = 0$ means that $x = \pm\sqrt{\frac{A}{3}}$. Since $x^3 - Ax + B = 0$ we get

$$(\sqrt{A/3})^3 - A\sqrt{A/3} + B = 0$$

$$\iff \frac{2A^{\frac{3}{2}}}{3\sqrt{3}} = B$$

$$\iff \frac{4A^3}{27} = B^2 \iff 4A^3 - 27B^2 = 0.$$

Substituting back $A = -A$ in the last equation shows us that we have a well-defined tangent line in the remaining case iff $4A^3 + 27B^2 \neq 0$ and this finishes the proof. □

In other words, Proposition 6.7 says that for a curve $E : y^2 = x^3 + Ax + B$ over $\mathbb{R}$ the condition $\Delta(E) = 4A^3 + 27B^2 \neq 0$ amount to a certain **smoothness**. As alluded earlier, in Algebraic Geometry one can extend the notion of smoothness to curves (or varieties) over arbitrary fields and in particular finite fields. We will not flesh out the definition of smoothness over finite fields here but merely point out that it remains equivalent to the condition $\Delta(E) = 4A^3 + 27B^2 \neq 0$ for all fields $\Bbbk$ with $\operatorname{char}\Bbbk \neq 2, 3$.

Let us examine the condition $\Delta(E) = 4A^3 + 27B^2 \neq 0$ further. Given a polynomial $f(x)$ of degree n over a field $\Bbbk$ we can ask when does $f(x)$ admit roots of multiplicity higher than 1 in the algebraic closure $\overline{\Bbbk}$.

Definition 6.8. Let $f(x) \in \Bbbk[x]$ be a polynomial of the form $f(x) = a_n x^n + ... + a_1 x + a_0$ and let $\alpha_1, ..., \alpha_n$ be its (not necessarily distinct) roots in the algebraic closure $\overline{\Bbbk}$. The **discriminant** of f is defined to be

$$\Delta(f) = a_n^{2n-2} \prod_{i<j} (\alpha_i - \alpha_j)^2.$$

What is relevant to us is the observation that f has roots of multiple degree iff $\Delta(f) = 0$. Since $f(x) = a_n \prod_{i=1}^n (x - \alpha_i)$ it follows that we can write the discriminant of f in terms of the coefficients $a_0, ..., a_n$.

Example 6.9.

1. A quadratic $f(x) = ax^2 + bx + c$ has the well-known discriminant

$$\Delta(f) = b^2 - 4ac$$

2. A cubic $f(x) = ax^3 + bx^2 + cx + d$ has discriminant

$$\Delta(f) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd$$

3. A quartic $f(x) = ax^4 + bx^3 + cx^2 + dx + e$ has discriminant

$$\begin{aligned}\Delta(f) =& 256a^3e^3 - 192a^2bde^2 - 128a^2c^2e^2 + 144a^2cd^2e \\ & - 27a^2d^4 + 144ab^2ce^2 - 6ab^2d^2e - 80abc^2de \\ & + 18abcd^3 + 16ac^4e - 4ac^3d^2 - 27b^4e^2 + 18b^3cde \\ & - 4b^3d^3 - 4b^2c^3e + b^2c^2d^2 .\end{aligned} \tag{15}$$

Remark 6.10. For a polynomial $f(x) = a_nx^n + ... + a_1x + a_0$, there is a formula for $\Delta(f)$ given by a determinant of a $2n \times 2n$ matrix built from the coefficients $a_0, ..., a_n$.

It follows from Example 6.9 that in the case of a cubic of the form $f(x) = x^3+Ax+B$ we have $\Delta(f) = -(4A^3+27B^2)$.

Thus, that the condition $\Delta(E) = 4A^3 + 27B^2 \neq 0$ is equivalent to the condition that the polynomial $x^3 + Ax + B$ has no roots of multiplicity > 1.

6.1 The group law on an elliptic curve

Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve over a field $\Bbbk$.

Definition 6.11. The **$\Bbbk$-rational** points of E are the set

$$E(\Bbbk) = \{(x, y) \in \Bbbk^2 | y^2 = x^3 + Ax + B\} \cup \{\mathcal{O}\}$$

where $\mathcal{O} = \infty$ is considered as a 'point at infinity' of E. If $\Bbbk \subseteq \mathbb{L}$ is any field inclusion, then the $\mathbb{L}$-rational points are simply

$$E(\mathbb{L}) = \{(x, y) \in \mathbb{L}^2 | y^2 = x^3 + Ax + B\} \cup \{\mathcal{O}\}.$$

The reason for including $\mathcal{O}$ will become clear later, but for now it is useful to regard it as a point sitting simultaneously at the top and bottom of the y-axis so that lines parallel to the y-axis pass through $\mathcal{O}$.

Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve defined over $\mathbb{R}$. Start with two points $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ on E as depicted in Figure 6.1. Draw the line L that intersects P and Q. We will see below (since E is a cubic) that L intersects E in a third point $R = (x_3, y_3)$. Since the graph of E is symmetric around the x-axis, the point $R' = (x_3, -y_3)$ must also lie on the curve E and we define $P + Q := R'$.

Assume first that $P \neq Q$ and that none of them is the point at infinity $\mathcal{O}$. Then the slope of the line L is

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

If $x_1 = x_2$ then L is vertical and we'll treat this case later, so suppose $x_1 \neq x_2$. Then L is given by $L : y = m(x-x_1)+y_1$. To find the intersection with E, we substitute that to get

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

which can be re-arranged in the form

$$0 = x^3 - m^2x^2 + ...$$

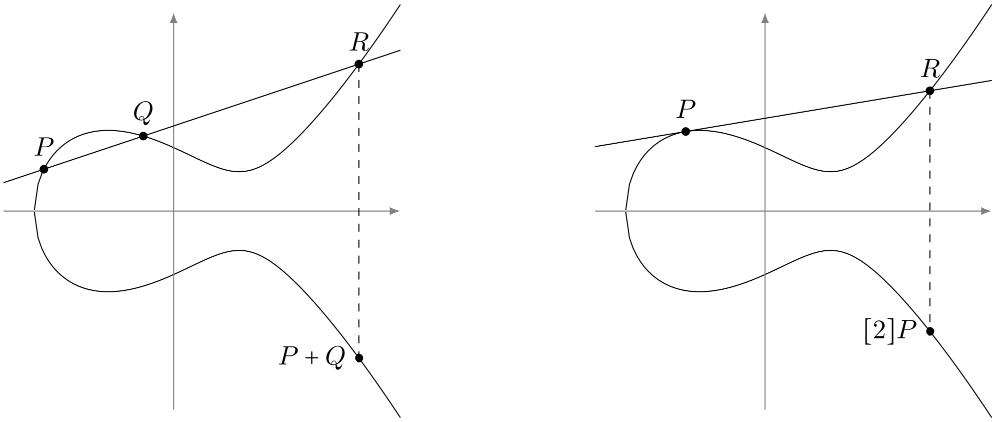


Figure 2: An elliptic curve defined over $\mathbb{R}$, and the geometric representation of its group law.

the roots of the last cubic are the x-coordinates of the intersection points of L with E and we know that x_1, x_2 are two such roots. If we decompose this cubic to linear factors over $\overline{\mathbb{R}} = \mathbb{C}$ (Theorem 5.90) as $x^3 - m^2x^2 + ... = (x - x_1)(x - x_2)(x - x_3)$ then since $x_1, x_2 \in \mathbb{R}$, x_3 must also be a real number. Note that for any monic cubic with three roots,

$$x^3 + ax^2 + bx + c = (x - r)(x - s)(x - t) = x^3 - (r + s + t)x^2 + ...$$

so in our case $x_3 = m^2 - x_1 - x_2$ and thus $y_3 = m(x_3 - x_1) + y_1$. Now reflect along the x-axis to get

$$P + Q = R' = (m^2 - x_1 - x_2, m(x_1 - x_3) - y_1).$$

In the case $x_1 = x_2$ but $y_1 \neq y_2$, the line L is vertical hence intersects E at $\mathcal{O}$ and reflecting $\mathcal{O}$ along the x-axis yields $\mathcal{O}$ again so we define $P + Q = \mathcal{O}$.

Lastly, if $P = Q$, the line that passes through P, Q can be thought of the limit line when P, Q get closer to each other, i.e. the tangent line to E at $P = Q$ (recall that we showed such a line always exist on elliptic curve). To find the slope of the tangent line, we use implicit differentiation

$$\frac{d}{dx}(y^2) = 2y\frac{dy}{dx} = 3x^2 + A \Rightarrow m = \frac{dy}{dx} = \frac{3x_1^2 + A}{2y_1}.$$

If $y_1 = 0$ the line is vertical and we set $P + Q = \mathcal{O}$ as before. Therefore, assume $y_1 \neq 0$. The equation for L is $L : y = m(x - x_1) + y_1$ and as before we obtain a cubic equation $0 = x^3 - m^2x^2 + ...$ this time, we know only one root, but it's a double root so we proceed as before to get $x_3 = m^2 - 2x_1$ and

$$P + P = (m^2 - 2x_1, m(x_1 - x_3) - y_1). \tag{16}$$

Finally, suppose $Q = \mathcal{O}$. The line through P and $\mathcal{O}$ is the vertical line that intersects E at P. The third intersection point of this line with E is the reflection of P along the x-axis and when we reflect this point along the x-axis we get P back. Thus, we define $P + \mathcal{O} = P$ for all points on E, so in particular $\mathcal{O} + \mathcal{O} := \mathcal{O}$.

Observation 6.12. The formulas for addition of points on E described above make sense for an elliptic curve $E/\Bbbk$ defined over any field $\Bbbk$. The only amendment we need to do is to replace $\overline{\mathbb{R}} = \mathbb{C}$ with $\overline{\Bbbk}$ in order to decompose a polynomial into linear factors (by Theorem 5.87). We thus extend our definition of addition of points to this more general case.

Theorem 6.13. *Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve defined over a field $\Bbbk$ with* $\operatorname{char}\Bbbk \neq 2,3$. *Then the $\Bbbk$-rational points*

$$E(\Bbbk) = \{(x,y) \in \Bbbk \times \Bbbk | y^2 = x^3 + Ax + B\} \cup \{\mathcal{O}\},$$

together with addition of points defined above, form an abelian group. The neutral element is $\mathcal{O}$ and the inverse of a point $P = (x_1, y_1)$ is given by $-P = (x_1, -y_1)$.

Proof sketch. It is easy to see from the definition of addition of points that $P + (-P) = \mathcal{O}$. However, proving that addition of points on E is associative is a tedious chase of equations and we will omit it from the current notes. A full proof may be found in [Wash]. See also Terrance Tau's blog for an intuitive explanation. □

Remark 6.14. The definition of addition of points on an elliptic curve over a general field illustrates a typical reasoning in Algebraic Geometry: one first makes construction over $\mathbb{R}$ using geometric insights, and then extend it to arbitrary fields by analogy.

We finish this section with the following

Observation 6.15. Let $E/\Bbbk = E_{A,B}$ be an elliptic curve defined over a field $\Bbbk$. If $\Bbbk \subseteq \Bbbk'$ is a subfield inclusion, then $E(\Bbbk) \subseteq E(\Bbbk')$ is a subgroup inclusion.

Proof. We have

$$\begin{aligned} E(\Bbbk) &= \{(x,y) \in \Bbbk \times \Bbbk \,| y^2 = x^3 + Ax + B\} \cup \{\mathcal{O}\} \\ &\subseteq \{(x,y) \in \Bbbk' \times \Bbbk' \,| y^2 = x^3 + Ax + B\} \cup \{\mathcal{O}\} = E(\Bbbk') \end{aligned} \tag{17}$$

since $\Bbbk \subseteq \Bbbk'$. The group operation in both sides agrees since it is defined in terms of the field operations and $\Bbbk \subseteq \Bbbk'$ is a subfield inclusion. □

6.2 Projective coordinates

We said that the $\Bbbk$-rational points of an elliptic curve $E(\Bbbk)$ include a 'point at infinity' $\mathcal{O}$. In this section we will formalize that matter.

Let $\Bbbk$ be a field and consider the set $\Bbbk \times \Bbbk \times \Bbbk \setminus \{(0,0,0)\}$ of triples of elements in $\Bbbk$ with the origin removed. Define an equivalence relation on this set by setting for each $(x,y,z) \in \Bbbk^3$ and nonzero scalar $\lambda \in \Bbbk^\times$:

$$(x,y,z) \sim (\lambda x, \lambda y, \lambda z).$$

To see why this is an equivalence relation, recall Definition 2.26. Note that for reflexivity we can take $\lambda = 1$, and get $(x,y,z) \sim (x,y,z)$.

For transitivity, if

$$(x_1,y_1,z_1) \sim (x_2,y_2,z_2) \text{ and } (x_2,y_2,z_2) \sim (x_3,y_3,z_3)$$

then

$$(x_2,y_2,z_2) = \lambda(x_1,y_1,z_1) \text{ and } (x_3,y_3,z_3) = \lambda'(x_2,y_2,z_2)$$

so that

$$(x_3,y_3,z_3) = \lambda'\lambda(x_1,y_1,z_1)$$

and we get $(x_1,y_1,z_1) \sim (x_3,y_3,z_3)$ hence $\sim$ is transitive (symmetry is left for the reader). Informally speaking, an equivalence class of $\sim$ can be viewed as a line in $\Bbbk^3$ that passes through the origin.

Definition 6.16. The **two-dimensional projective plane** over $\Bbbk$ is the quotient set

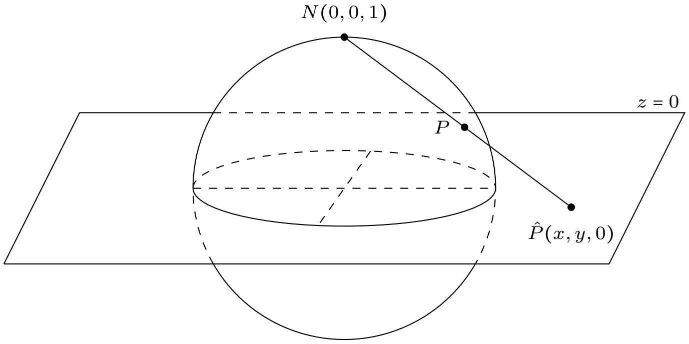
$$\mathbb{P}^2_{\Bbbk} = \Bbbk^3 \setminus \{(0,0,0)\}/\sim$$

by the equivalence relation described above.

We denote the equivalence class of a point (x,y,z) by $(x:y:z)$.

As an intuition for the definition of $\mathbb{P}^2_{\Bbbk}$ consider $\Bbbk = \mathbb{R}$. A point $(x:y:z) \in \mathbb{P}^2_{\mathbb{R}}$ corresponds to the collection $\{(\lambda x, \lambda y, \lambda z) | \lambda \in \mathbb{R}\}$ which can be considered as a line through the origin in $\mathbb{R}^3$. The lines on the x-y plane are the "points at infinity" which correspond to directions. A representative of this line may be taken to be a unit vector on this line i.e. a point on the 2-dimensional sphere S^2. Antipodal points on the sphere are identified with the same point on the projective plane.

Thus, we can identify $\mathbb{P}^2_{\mathbb{R}} \simeq S^2/\sim$ where $\sim$ identifies antipodal points. Under this identification, we can view $\mathbb{R}^2$ as points in $S^2/\sim$ via the **stereographic projection** shown in picture below. Here each point $\hat{P}$ on the plane is identified with two points on the sphere $P, -P$ by drawing a line that passes through $\hat{P}$ and either the north or south pole, and taking the intersection points $P, -P$ of this line with the sphere. One can find explicit formulas for this projection in terms of "spherical coordinates" ie in terms of Sine and Cosine.



A polynomial in three variables $F(x,y,z)$ over $\Bbbk$ is a sum of terms $a_{ijk}x^iy^jz^k$, called **monomials**, where $a_{ijk} \in \Bbbk$. Such polynomial is called **homogeneuos** of degree n if its monomials are all of the form $a_{ijk}x^iy^jz^k$ with $i+j+k=n$. If $F(x,y,z)$ is a homogeneous polynomial of degree n, then for any $\lambda \in \Bbbk^\times$, $F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x,y,z)$. Thus, for such polynomials, $F(x,y,z) = 0$ if and only if for any $\lambda \in \Bbbk^\times$, $F(\lambda x, \lambda y, \lambda z) = 0$. It follows that if $F(x,y,z)$ is homogeneous of some degree and $(x_1,y_1,z_1) \sim (x_2,y_2,z_2)$ then $F(x_1,y_1,z_1) = 0$ if and only if $F(x_2,y_2,z_2) = 0$.

Therefore, a zero of such polynomial F in $\mathbb{P}^2_{\Bbbk}$ is well-defined as it does not depend on the representative of the equivalence class.

Remark 6.17. The polynomial $F(x,y,z) = x^2 + 2y - 3z$ is not homogeneous and thus the considerations described above fail. For example, $F(1,1,1) = 0$ so we might be tempted to say that F has a zero at $(1:1:1)$, but $F(2,2,2) = 2 \neq 0$ whereas $(1:1:1) = (2:2:2)$.

Definition 6.18. If $f(x,y)$ is any polynomial in two variables of degree n, then

$$F(x,y,z) = z^n f(\frac{x}{z}, \frac{y}{z})$$

is called the **homogenization** of f.

Note the homogenization $F(x,y,z)$ of a polynomial $f(x,y)$ of degree n is a homogeneous polynomial of degree n and $F(x,y,1) = f(x,y)$. For example, if

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

then

$$F(x,y,z) = y^2z - x^3 - Axz^2 - Bz^3$$

is a homogeneous polynomial of degree 3.

Example 6.19. We can now understand what it formally means that two parallel lines intersect at infinity. Let

$$\ell_1 : y = mx + b_1$$

$$\ell_2 : y = mx + b_2$$

be two non-vertical parallel lines with $b_1 \neq b_2$. Their homogenization is

$$y = mx + b_1z$$

$$y = mx + b_2z$$

and to check their intersection in $\mathbb{P}^2_{\Bbbk}$ we look for zeros of the homogeneous polynomial

$$y - mx - b_1z - (y - mx - b_2z) = (b_2 - b_1)z$$

and see that $z = 0$. This means that $y = mx$ and since we cannot have $x = y = 0$ we get that a representative of the intersection of ℓ_1 and ℓ_2 in $\mathbb{P}^2_{\Bbbk}$ is $(x : mx : 0) = (1 : m : 0)$ which is a point at infinity.

Similarly, if $\ell_1 : x = b_1$ and $\ell_2 : x = b_1$ are two distinct parallel vertical lines, their intersection in $\mathbb{P}^2_{\Bbbk}$ is $(0 : 1 : 0)$ which is also one of the points at infinity.

We now look at an elliptic curve $E : y^2 = x^3 + Ax + B$ defined over a field $\Bbbk$. Its homogeneous form is $y^2z = x^3 + Axz^2 + Bz^3$. The points (x,y) on the original curve in $\Bbbk \times \Bbbk$ correspond to points $(x : y : 1)$ in $\mathbb{P}^2_{\Bbbk}$. To see what points on the projective version of E lie at infinity, set $z = 0$ and obtain $x^3 = 0$ ie. $x = 0$. Thus, the coordinate y can be any non-zero element of $\Bbbk$. In other words, the point $(0 : y : 0) = (0 : 1 : 0)$ is the only point at infinity lying on the projective version of E.

6.3 Rational functions on an elliptic curve

Throughout this section, we let $\Bbbk$ be a field with $\operatorname{char} \Bbbk \neq 2, 3$. Recall our notation $E_{A,B}$ for an elliptic curve $E : y^2 = x^3 + Ax + B$ defined over $\Bbbk$.

Recall form section 5.4 that for a polynomial $g(x) \in \Bbbk[x]$ we can define mod-g arithmetic. A standard notation for mod-g arithmetic is $\Bbbk[x]/(g = 0)$. Similarly, if $g(x,y) \in \Bbbk[x,y]$ we can the analog of mod-g arithmetic is the set $\Bbbk[x,y]/(g = 0) = \Bbbk[x,y]/\sim$ where $a(x,y) \sim b(x,y)$ iff $a(x,y) - b(x,y) = f(x,y)g(x,y)$ for some $f(x,y) \in \Bbbk[x,y]$.

Definition 6.20. For an elliptic curve $E/\Bbbk : y^2 = x^3 + Ax + B$ we define the set of **polynomials** over E to be

$$\Bbbk[E] := \Bbbk[x,y]/(y^2 - x^3 - Ax - B = 0).$$

Remark 6.21. It can be shown that the defining polynomial of an elliptic curve $g(x,y) = y^2 - x^3 - Ax - B$ is irreducible in $\Bbbk[x,y]$ ie cannot be written as $g = a \cdot b$ for non-constant polynomials $a, b \in \Bbbk[x,y]$. However, unlike in $\Bbbk[x]$, $\Bbbk[x,y]$ does not admit a well-defined gcd. For example, if $f(x,y) = x$ and $g(x,y) = y$ we would want to say that $\gcd(f,g) = 1$, implying (by Bezout identity) that there are polynomials $a(x,y), b(x,y)$ such that $xa(x,y) + yb(x,y) = 1$ (as polynomials). However, if we substitute $(x,y) = (0,0)$ we get $0 = 1$, a contradiction. Thus, $\Bbbk[E]$ is not a field.

By definition, we can replace every term y^2 in a polynomial $f \in \Bbbk[E]$ with $x^3 + Ax + B$ without changing the equivalence class of f. Thus, f can be written in a **canonical form** as $f(x,y) = v(x) + yw(x)$ for some $v, w \in \Bbbk[x]$.

Exercise 6.22. Show that the canonical form is unique.

Definition 6.23. Let $f \in \Bbbk[E]$ be given in canonical form $f(x,y) = v(x) + yw(x)$. The **conjugate** of f is $\overline{f} = v(x) - yw(x)$ and the **norm** of f is

$$N_f = f \cdot \overline{f} = v(x)^2 - y^2 w(x)^2 = v(x)^2 - (x^3 + Ax + B)w(x)^2 \in \Bbbk[x] \subseteq \Bbbk[E].$$

Exercise 6.24. Show that $N_{fg} = N_f N_g$ for any $f, g \in \Bbbk[E]$.

As pointed out in Remark 6.21, $\Bbbk[E]$ is not a field. In order to make it such, we have the following

Definition 6.25. For an elliptic curve $E/\Bbbk$ the set of **rational functions** on E is the quotient set

$$\Bbbk(E) \coloneqq \Bbbk[E] \times \Bbbk[E] / \sim$$

where $(f,g) \sim (h,k) \iff f \cdot k = h \cdot g \in \Bbbk[E]$. To check if equality holds, we can write $f \cdot k$ and $h \cdot g$ in canonical forms and compare coefficients. We denote the equivalence class of (f,g) by $\frac{f}{g}$. For $r \in \Bbbk(E)$ and a finite point $P \in E(\Bbbk)$ we say that r is **finite** at P if there exists a representation $r = \frac{f}{g}$ with $f, g \in \Bbbk[E]$ such that $g(P) \neq 0$. In this case, we define $r(P) = \frac{f(P)}{g(P)}$. Otherwise, we write $r(P) = \infty$.

Remark 6.26. For $r = \frac{f}{g} \in \Bbbk(E)$ we can write

$$\frac{f}{g} = \frac{f\overline{g}}{g\overline{g}} = \frac{f\overline{g}}{N_g}$$

and write $f\overline{g}$ in canonical form $(f\overline{g})(x,y) = v(x) + yw(x)$. We get

$$r(x,y) = \frac{f(x,y)}{g(x,y)} = \frac{(f\overline{g})(x,y)}{N_g(x)} = \frac{v(x)}{N_g(x)} + y\frac{w(x)}{N_g(x)}$$

which we will refer to as the **canonical form** for r.

Our next matter is defining the value of a rational function r at the point at infinity ie to give meaning to the expression $r(\mathcal{O})$. In the situation of a rational function in one variable, ie an expression of the form $r(x) = \frac{f(x)}{g(x)}$ with $f(x), g(x) \in \Bbbk[x]$ we typically (as in calculus) compare the degrees of f and g in order to get a meaningful value $r(\infty)$. For example, if $r(x) = \frac{x}{x^2+1}$ we would say that $r(\infty) = 0$ whereas if $r(x) = \frac{x^2}{x+1}$ we would say that $r(\infty) = \infty$. The situation in $\Bbbk[E]$ is more subtle since we have $y^2 = x^3 + Ax + B$ which means that the degree of y should be $\frac{2}{3}$ of the degree of x. Since we want to keep degrees as integers, we set $\deg(y) = 3$ and $\deg(x) = 2$ in $\Bbbk[E]$. The classical degree of a polynomial $f \in \Bbbk[x]$ will be denoted $\deg_x(f)$. This motivates the following

Definition 6.27. Let $f \in \Bbbk[E]$ and write it in canonical form $f(x,y) = v(x) + yw(x)$. The **degree** of f is

$$\deg(f) \coloneqq \max\{2 \cdot \deg_x(v), 3 + 2 \cdot \deg_x(w)\}.$$

Remark 6.28. Recall that $\deg_x(0) = -\infty$ and $\deg_x(c) = 0, \ \forall c \in \Bbbk^\times$.

Proposition 6.29. *Let $E = E_{A,B}$ be an elliptic curve defined over $\Bbbk$ and denote $s(x) = x^3 + Ax + B$. For $f, g \in \Bbbk[E]$:*

1. $\deg(f) = \deg_x(N_f)$.

2. $\deg(f \cdot g) = \deg(f) + \deg(g)$.

Proof.

1. Write f in canonical form $f(x,y) = v(x) + yw(x)$, then $N_f = v(x)^2 - s(x)w(x)^2$. Since $\deg_x(v^2)$ and $\deg_x(w^2)$ are even and $\deg_x(s)$ is odd, it follows that

$$\begin{aligned}\deg_x(N_f) = \deg_x(v^2 - sw^2) &= \max\{\deg_x(v^2), \deg_x(s) + \deg_x(w^2)\} \\ &= \max\{2\deg_x(v), 3 + 2\deg_x(w)\} = \deg(f).\end{aligned} \tag{18}$$

2. We can easily calculate

$$\begin{aligned}\deg(fg) = \deg_x(N_{fg}) = \deg_x(N_f N_g) &= \deg_x(N_f) + \deg_x(N_g) \\ &= \deg(f) + \deg(g).\end{aligned} \tag{19}$$

□

It makes no sense to talk about the degree of the nominator of a rational function $r \in \Bbbk(E)$ since that depends on the representation $r = \frac{f}{g} = \frac{h}{k}$. However, if $r = \frac{f}{g} \in \Bbbk[E]$, the quantity $\deg(f) - \deg(g)$ does not depend on the representation since if $\frac{f}{g} = \frac{h}{k}$ we have $fk = hg$ and by Proposition 6.29, $\deg(f) - \deg(g) = \deg(h) - \deg(k)$.

Definition 6.30. Let $r = \frac{f}{g} \in \Bbbk(E)$ be a rational function and distinguish the following cases:

1. If $\deg(f) < \deg(g)$: set $r(\mathcal{O}) = 0$.
2. If $\deg(f) > \deg(g)$: say that r is not finite at $\mathcal{O}$.
3. If $\deg(f) = \deg(g)$ and $\deg(f)$ is **even**: write both f and g in canonical form, so that they both have leading terms ax^d and bx^d (respectively) with $a, b \in \Bbbk^\times$ and $d = \frac{\deg(f)}{2}$, and we set $r(\mathcal{O}) = \frac{a}{b}$.
4. If $\deg(f) = \deg(g)$ and $\deg(f)$ is **odd**: write both f and g in canonical form, so that they both have leading terms $ax^d y$ and $bx^d y$ (respectively), $a, b \in \Bbbk^\times$ and $\deg(f) = \deg(g) = 3 + 2d$, and we set $r(\mathcal{O}) = \frac{a}{b}$.

Remark 6.31. For $r = \frac{f}{g} \in \Bbbk(E)$, it may seem natural to define $\deg(r) = \deg(f) - \deg(g)$ so that the value $r(\mathcal{O})$ would depend on the sign of $\deg(r)$. However, this differs from the usual definition of a degree of a rational function in Algebraic Geometry so we avoid defining the degree of a rational function altogether.

Example 6.32. Consider $E = E_{A,B}$ and $\Bbbk(E)$. For

$$r(x,y) = \frac{x^3 + 2x + y + 2x^4 y}{x + x^2 + 5xy^3}$$

one can write

$$r(x,y) = \frac{x^3 + 2x + y + 2x^4 y}{x + x^2 + 5xy(x^3 + Ax + B)} = \frac{(x^3 + 2x) + y(1 + 2x^4)}{(x + x^2) + y(5x^4 + 5Ax^2 + 5Bx)}.$$

The last representative has nominator of degree $\max\{2 \cdot 3, 3 + 2 \cdot 4\} = 11$, and denominator of degree $\max\{2 \cdot 3, 3 + 2 \cdot 4\} = 11$ which are both odd. Thus $r(\mathcal{O}) = \frac{2}{5}$.

Exercise 6.33. For $r, s \in \Bbbk(E)$ with $r(\mathcal{O}), s(\mathcal{O})$ finite, we have $(rs)(\mathcal{O}) = r(\mathcal{O})s(\mathcal{O})$ and $(r + s)(\mathcal{O}) = r(\mathcal{O}) + s(\mathcal{O})$.

6.3.1 Zeros and poles

Definition 6.34. Let $E/\Bbbk$ be an elliptic curve and let $r \in \Bbbk(E)$ be a rational function. We say that r has a **zero** in $P \in E$ if $r(P) = 0$ and that r has a **pole** in P if $r(P)$ is not finite.

The goal of this section is to define the **multiplicity** of zeros and poles. The motivation comes from functions in one variable. Consider $E = E_{1,0} : y^2 = x^3 + x$ and $P = (0,0) \in E$. Then P is a zero of the functions x and y. However, between these two functions there is a relation: $x = y^2 - x^3$. In the analytic sense, when $x \longrightarrow 0$, the term x^3 can be neglected so we would like to say that the function x has a zero at P whose multiplicity is twice that of the function y at P. This is formalised in the following

Definition 6.35. Let $E/\Bbbk$ be an elliptic curve and $P \in E$. A rational function $u \in \Bbbk(E)$ with $u(P) = 0$ is called a **uniformizer** at P if:
$\forall r \in \Bbbk(E) \smallsetminus \{0\},\ \ \exists d \in \mathbb{Z}, s \in \Bbbk(E)$ finite at P with $s(P) \neq 0$ such that

$$r = u^d \cdot s.$$

Remark 6.36. As we will see soon, uniformizers exist for all points in $E(\Bbbk)$ (though different points may require different uniformizers). However, even for a fixed point P there is usually more than one uniformizer at P. The uniformizers we present below are simply one common choice.

Proposition 6.37. *Let $E/\Bbbk$ be an elliptic curve and $P = (a,b) \in E$ a finite point with $2P \neq \mathcal{O}$. Then the function $u(x,y) = x - a$ is a uniformizer at P.*

Proof. First note that $u(P) = 0$. Now let $r \in \Bbbk(E) \smallsetminus \{0\}$ be arbitrary. If r has neither zero nor pole at P, we can take $d = 0$ and $r = s$.

Suppose r has a zero at P, ie $r(P) = 0$. Note that if we have proved that u is a uniformizer in the case P is a zero, then for $r \in \Bbbk(E)$ with a pole at P, $\frac{1}{r}$ has a zero at P so that there exists $d \in \mathbb{Z}$ and $s \in \Bbbk(E)$ which is finite and non-zero at P such that $\frac{1}{r} = u^d s$. But then $r = u^{-d}\frac{1}{s}$ shows that u is a uniformizer for $\frac{1}{r}$.

Thus, we assume $r(P) = 0$ so we can write $r\frac{f}{g}$ with $f(P) = 0$ and $g(P) \neq 0$. If we can decompose $f = u^d s$ as above then

$$r = \frac{f}{g} = \frac{u^d s}{g} = u^d \frac{s}{g}$$

with $\frac{s}{g}(P) \neq 0$ and finite so we are done.

Set $s_0(x,y) = f(x,y)$ and repeat the following process (starting from $i = 0$) while $s_i(P) = 0$:

Write $s_i(x,y) = v_i(x) + y w_i(x)$ in canonical form. Distinguish the cases $\overline{s}_i(P) = 0$ and $\overline{s}_i(P) \neq 0$ (recall that $\overline{s}_i = v_i(x) - y w_i(x)$ is the conjugate of s_i.)

$\underline{\overline{s}_i(P) = 0}$: Since $y(P) = b \neq 0$, the system of linear equations

$$\begin{aligned} v_i(a) + b w_i(a) &= 0 \\ v_i(a) - b w_i(a) &= 0 \end{aligned} \tag{20}$$

has a unique solution (e.g. its rank equals 2) of the form $v_i(a) = w_i(a) = 0$.

Thus, we can write

$$s_i(x,y) = v_i(x) + y w_i(x) = (x-a)v_{i+1}(x) + (x-a) y w_{i+1}(x) = (x-a)s_{i+1}(x) \tag{21}$$

for $s_{i+1}(x) := v_{i+1}(x) + y w_{i+1}(x)$ with some polynomials $v_{i+1}(x), w_{i+1}(x) \in \Bbbk[x]$.

$\underline{\overline{s}_i(P) \neq 0}$: multiply s_i by $1 = \frac{\overline{s}_i}{\overline{s}_i}$ to get

$$s_i(x,y) = \frac{N_{s_i}}{\overline{s}_i}.$$

Now, $s_i(P) = 0$ and $\overline{s}_i(P) \neq 0$ implies that $N_{s_i}(a) = 0$ so we can write

$$N_{s_i}(x) = (x-a)n(x)$$

for some $n(x) \in \Bbbk[x]$.

We now set

$$s_{i+1}(x) = \frac{n(x)}{\overline{s}_i(x,y)}$$

(which is finite at P), and we again get

$$s_i(x,y) = \frac{N_{s_i(x)}}{\overline{s}_i(x)} = \frac{(x-a)n(x)}{\overline{s}_i(x,y)} = (x-a)s_{i+1}(x,y).$$

If the process terminates, we get $f(x,y) = (x-a)^i s_i(x,y)$ where $s \coloneqq s_i$ is finite and non-zero with $u(x,y) = x-a$ and $d = i$ so we are done.

Since s_i is a rational function and not a polynomial, it is not clear that this process indeed terminates. Let us show it anyhow.

$$\begin{aligned} N_f(x) &= N_{u^i s_i}(x) \\ &= ((x-a)^i v_i(x))^2 - y^2((x-a)^i w_i(x))^2 \\ &= (x-a)^{2i}(v_i(x)^2 - y^2 w_i(x)^2) \\ &= (x-a)^{2i} N_{s_i} \end{aligned} \tag{22}$$

and we see that i is bounded since $\deg(N_f) = 2i + \deg(N_{s_i})$ and since $\deg(N_{s_i}) > 0$. Thus, there can only be finitely many iterations i and this finishes the proof.

□

Lemma 6.38. *Let $E/\Bbbk = E_{A,B}$ be an elliptic curve over a field $\Bbbk$ such that $E(\Bbbk)$ contains all point of order two (e.g. $\Bbbk$ algebraically closed). Let $P \in E$ such that $2P = \mathcal{O}$. Then the rational function $u_P(x,y) = u(x,y) = y$ is a uniformizer at P.*

Proof. Since E is an elliptic curve, $s_E(x) = x^3 + Ax + B$ has three distinct roots $\alpha_1, \alpha_2, \alpha_3$ ie

$$s_E(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3).$$

We saw that if P is of order 2 then (without loss of generality) $P = (\alpha_1, 0)$ so that $u(P) = 0$. Let $r = \frac{f}{g} \in \Bbbk(E) \setminus \{0\}$ be such that $r(P) = 0$. Then we can assume that the presentation $r = \frac{f}{g}$ is such that $f(P) = 0$ and $g(P) \neq 0$. Write f in canonical form $f(x,y) = v(x) + yw(x)$ which means $v(\alpha_1) = 0$ so $v(x) = (x-\alpha_1)v_1(x)$ for some $v_1(x)$ with $\deg v_1 < \deg v$. We can thus write

$$\begin{aligned} f(x,y) = (x-\alpha_1)v_1(x) + yw(x) &= \frac{(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)v_1(x) + yw_1(x)}{(x-\alpha_2)(x-\alpha_3)} \\ &= \frac{y^2 v_1(x) + yw_1(x)}{(x-\alpha_2)(x-\alpha_3)} = y\frac{yv_1(x) + w_1(x)}{(x-\alpha_2)(x-\alpha_3)} = u(x,y)W(x,y) \end{aligned} \tag{23}$$

where $w_1(x) = w(x)(x-\alpha_2)(x-\alpha_3)$ and $W(x,y) = \frac{yv_1(x)+w_1(x)}{(x-\alpha_2)(x-\alpha_3)}$. Note that $W(P)$ is finite . If $W(P) \neq 0$ we are done since we can take

$$s(x,y) = W(x,y)/g(x,y)$$

and write $r(x,y) = u(x,y)^1 \cdot s(x,y)$. Otherwise, we repeat the process with W/g instead of r. More specifically, we take

$$r'(x,y) = \frac{W(x,y)}{g(x,y)} = \frac{yv_1(x)+w_1(x)}{g(x,y)(x-\alpha_2)(x-\alpha_3)}$$

and by assumption $r'(P) = 0$. This means that $w_1(\alpha_1) = 0$ so that

$$w_1(x) = (x-\alpha_2)(x-\alpha_3)w(x) = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)w_2(x)$$

for some $w_2(x)$ such that $\deg w_2 < \deg w$. We can thus write

$$\begin{aligned} W(x,y) &= \frac{yv_1(x)+w_1(x)}{(x-\alpha_2)(x-\alpha_3)} \\ &= \frac{yv_1(x)+(x-\alpha_1)(x-\alpha_2)(x-\alpha_3)w_2(x)}{(x-\alpha_2)(x-\alpha_3)} \\ &= \frac{yv_1(x)+y^2w_2(x)}{(x-\alpha_2)(x-\alpha_3)} \\ &= y\frac{v_1(x)+yw_2(x)}{(x-\alpha_2)(x-\alpha_3)} =: yW_1(x,y) \end{aligned} \tag{24}$$

where

$$W_1(x,y) = \frac{v_1(x)+yw_2(x)}{(x-\alpha_2)(x-\alpha_3)}.$$

As before, $W_1(P)$ is finite and if $W_1(P) \neq 0$ we are done since we can take

$$s'(x,y) = W_1(x,y)/g(x,y)$$

and get $r' = u^1 s'$ with $s'(P) \neq 0$ so that $r = u^2 s'$. As can be seen form the above, repeating the process produces a sequence of polynomials $v_1, v_2, ...$ and $w_1, w_2, ..$ such that for n odd $\deg v_n < \deg v_{n-2}$ and for n even $\deg w_n < \deg w_{n-2}$. It follows that the process must terminate since v and w have only finitely many roots.

Lastly, if $r(P) = \infty$, i.e. r has a pole at P , then $\frac{1}{r}(P) = 0$ so that by the argument above, $\frac{1}{r} = u^d w$ with $w(P) \neq 0$ and thus $r = u^{-d}\frac{1}{w}$ as required. □

Lemma 6.39. *Let $E/\Bbbk$ be an elliptic curve. Then the function $u(x,y) = \frac{x}{y}$ is a uniformizer at $\mathcal{O}$.*

Proof. Since $\deg(y) = 3 > \deg(x) = 2$, it follows from Definition 6.30 that $u(\mathcal{O}) = 0$. Let $r = \frac{f}{g} \in \Bbbk(E) \setminus \{0\}$ be such that $r(\mathcal{O}) = 0$ or $r(\mathcal{O})$ is not finite. This means, by Definition 6.30 that $d = \deg(f) - \deg(g) \neq 0$. We would like to take $s(x,y) = \left(\frac{y}{x}\right)^d r(x,y)$ since then we have

$$u(x,y)^d s(x,y) = \left(\frac{x}{y}\right)^d \left(\frac{y}{x}\right)^d r(x,y) = r(x,y)$$

However, in order for this to work we need to show that $s(\mathcal{O})$ is finite and non-zero. We have

$$s(x,y) = \frac{y^d f(x,y)}{x^d g(x,y)}$$

and because

$$\begin{aligned}\deg(y^d f(x,y)) - \deg(x^d g(x,y)) &= \deg(y^d) + \deg(f) - (\deg(x^d) + \deg(g)) \\ &= 3d + \deg(f) - 2d - \deg(g) = 0\end{aligned} \tag{25}$$

we get from Definition 6.30 that $s(P)$ is finite and non-zero.

□

Theorem 6.40. *Let $E/\Bbbk$ be an elliptic curve. Then any point on $E(\Bbbk)$ has a uniformizer and the number d of Definition 6.35 does not depend on its choice.*

Proof. The previous claims ensure the existence of a uniformizer for every point $P \in E(\Bbbk)$. It is left to show that the integer d does not depend on the choice of uniformizer. Let $P \in E(\Bbbk)$ and let $u, u' \in \Bbbk(E)$ be uniformizers at P. Then we can write $u = u'^a p$ and $u' = u^b q$ for $p, q \in \Bbbk(E)$ such that $p(P), q(P) \neq 0, \infty$. We thus get

$$u = u'^a p = (u^b q)^a p = u^{ab} q^a p \iff 1 = u^{ab-1} q^a p.$$

If $ab \neq 1$ then evaluating at P gives $1 = 0 \cdot q^a(P)p(P) = 0$ which is a contradiction. Thus $ab = 1$ ie. $a = b = \pm 1$. If $a = b = -1$ we get

$$u = u'^{-1} p \iff uu' = p$$

which cannot be true since $p(P) \neq 0$ whereas $u(P) = u'(P) = 0$ so we must have, $a = b = 1$.

If $r \in \Bbbk(E) \setminus \{0\}$ then since u and u' are uniformizers at P we get $r = u^d s$ and $r = u^{d'} s'$ for some $d, d' \in \mathbb{Z}$ and $s, s' \in \Bbbk(E)$ such that $s(P), s'(P) \neq 0, \infty$. But then

$$u^d s = u'^{d'} s' = (uq)^{d'} s' = u^{d'} q^{d'} s'$$

which yields

$$u^{d-d'} = \frac{q^{d'} s'}{s}.$$

If $d \neq d'$ we get a contradiction since the LHS evaluated at P is zero while the RHS evaluated at P is non-zero. Thus, $d = d'$ as desired.

□

Definition 6.41. Let $E/\Bbbk$ be an elliptic curve, $P \in E(\Bbbk)$ and $u \in \Bbbk(E)$ a uniformizer at P. For $r \in \Bbbk(E) \setminus \{0\}$ a rational function with $r = u^d \cdot s$ with $s(P) \neq 0, \infty$, we say that r has **order** d at P and write

$$\mathrm{ord}_P(r) = d.$$

The **multiplicity of a zero** of r is the order of r at that point and the **multiplicity of a pole** of r is the order of r at that point.

Observation 6.42. Let $E/\Bbbk$ be an elliptic curve and $r \in \Bbbk(E)$ a rational function. If $P \in E(\Bbbk)$ a point which is neither a zero or a pole of r, then $\mathrm{ord}_P(r) = 0$.

Proof. Pick a uniformizer u and set $s(x,y) = r(x,y)$. Then $s(P)$ is finite and non-zero and $r = u^0 s$. □

Example 6.43. Let $E = E_{A,B}$ be an elliptic curve and $P = (a,b) \in E(\Bbbk)$, finite and not of order 2. We want to calculate the orders of $r(x,y) = x - a$ at all points $Q \in E(\Bbbk)$ where $r(Q)$ is zero or not finite (otherwise, $\operatorname{ord}_Q(r) = 0$ by Observation 6.42). Note that $Q = P = (a,b)$ and $Q = -P = (a,-b)$ are both zeros of r. Since in this case r itself is a uniformizer (with $s(x,y) = 1$) we get that $\operatorname{ord}_Q(r) = 1$.

When $Q = \mathcal{O}$, we have a pole for r. We take as a uniformizer $u(x,y) = \frac{x}{y}$ and $s(x,y) = \frac{x^3 - ax^2}{y^2}$ (note that $s(Q) = 1$) and get

$$u(x,y)^{-2}s(x,y) = \frac{y^2}{x^2} \cdot \frac{x^3 - ax^2}{y^2} = x - a = r(x,y)$$

so that $\operatorname{ord}_{\mathcal{O}}(r) = -2$.

Example 6.44. Let $E/\Bbbk : y^2 = x^3 + Ax + B$ be an elliptic curve and let $r(x,y) = y \in \Bbbk(E)$. From Example 6.95, we know that a point $P \in E(\Bbbk)$ is of order two iff $P = (\alpha, 0)$ where α is a root of $x^3 + Ax + B$. According to Lemma 6.38, the rational function $u(x,y) = y$ is a uniformizer for points of order two, and we thus get $r = u^1 \cdot 1$ deducing that at point P of order two, r has a zero of multiplicity 1. According to Lemma 6.39, $u(x,y) = \frac{x}{y}$ is a uniformizer at $\mathcal{O}$. The rational function $s(x,y) = \frac{x^3 y}{y^3}$ satisfy $s(\mathcal{O}) \neq 0, \infty$ by Definition 6.30 since $\deg(x^3 y) - \deg(y^3) = 6 + 3 - 3 \cdot 3 = 0$. But then we get

$$u(x,y)^{-3}s(x,y) = \left(\frac{x}{y}\right)^{-3} \frac{x^3 y}{y^3} = y = r(x,y)$$

so we see that $\operatorname{ord}_{\mathcal{O}}(r) = -3$.

6.4 Divisors

Examples 6.44 shows an interesting phenomenon: if $\Bbbk$ is **algebraically closed** then $x^3 + Ax + B$ has 3 distinct roots $\alpha_1, \alpha_2, \alpha_3$ and thus $E = E_{A,B}$ has 3 points of order 2:

$$\{P_1, P_2, P_3\} = \{(\alpha_1, 0), (\alpha_2, 0), (\alpha_3, 0)\} \subseteq E(\Bbbk).$$

The function $r(x,y) = y \in \Bbbk(E)$ has 3 zeros, namely P_1, P_2, P_3, each of order 1 and a pole at $\mathcal{O}$ of order 3. In other words if we sum the orders of all points (recall that order of a point which is not a zero or a pole is 0 by Observation 6.42), we get $\sum_{P \in E(\Bbbk)} \operatorname{ord}_P(r) = 1 + 1 + 1 - 3 = 0$. In fact, the same is true for Example 6.43 since in that case $\sum_{Q \in E(\Bbbk)} \operatorname{ord}_Q(r) = 1 + 1 - 2 = 0$. This motivates the following

Definition 6.45. Let $E/\Bbbk$ be an elliptic curve. A **divisor** on E is an expression

$$D = \sum_{P \in E(\Bbbk)} n_P[P]$$

where $\forall P,\ n_P \in \mathbb{Z}$ and only finitely many n_P's are non-zero. The **degree** of a divisor D is

$$\deg(D) = \sum_{P \in E(\Bbbk)} n_P.$$

The **sum** of a divisor D is

$$\operatorname{sum}(D) = \sum_{P \in E(\Bbbk)} n_P \cdot P \in E(\Bbbk).$$

Observation 6.46. The set of all divisors on E forms a group: If $D = \sum_{P\in E(\Bbbk)} n_P[P]$ and $D' = \sum_{P\in E(\Bbbk)} m_P[P]$ then

$$D + D' := \sum_{P\in E(\Bbbk)} (n_P + m_P)[P].$$

The unit element is the divisor $\sum_{P\in E(\Bbbk)} 0\cdot[P]$ and the inverse of a divisor $D = \sum_{P\in E(\Bbbk)} n_P[P]$ is the divisor $-D = \sum_{P\in E(\Bbbk)} -n_P[P]$.

Definition 6.47. Let $E/\Bbbk$ be an elliptic curve and $r \in \Bbbk(E)\smallsetminus\{0\}$ a rational function. The **associated divisor** of r is defined to be

$$\operatorname{div}(r) := \sum_{P\in E(\Bbbk)} \operatorname{ord}_P(r)[P].$$

Lemma 6.48. *Let $E/\Bbbk$ be an elliptic curve and $r, s \in \Bbbk(E)\smallsetminus\{\mathcal{O}\}$ rational functions.*

1. $\operatorname{div}(rs) = \operatorname{div}(r) + \operatorname{div}(s)$.
2. $\operatorname{div}(\frac{r}{s}) = \operatorname{div}(r) - \operatorname{div}(s)$.

Proof. Both claims follow from the claim that for any $P \in E(\Bbbk)$, $\operatorname{ord}_P(rs) = \operatorname{ord}_P(r) + \operatorname{ord}_P(s)$ which in turn follows from the fact that is u is a uniformizer at P with degrees d_r, d_s for r, s respectively, then the degree of u for rs is $d_r + d_s$. Analogously, $\operatorname{ord}_P(\frac{r}{s}) = \operatorname{ord}_P(r) - \operatorname{ord}_P(s)$ □

There is another astonishing fact emerging from Examples 6.43 and 6.44: for a rational functions $r \in \Bbbk(E)$, we have in the group $E(\Bbbk)$:

$$\sum_{P\in E(\Bbbk)} \operatorname{ord}_P(r)\cdot P = \mathcal{O}.$$

In Example 6.43 we have

$$\sum_{Q\in E(\Bbbk)} \operatorname{ord}_Q(r)\cdot Q = 1\cdot P + 1\cdot(-P) + (-2)\cdot\mathcal{O} = \mathcal{O}$$

where as in Example 6.44 we have

$$\sum_{P\in E(\Bbbk)} \operatorname{ord}_P(r)\cdot P = P_1 + P_2 + P_3 + (-3)\cdot\mathcal{O} = P_1 + P_2 + P_3 = \mathcal{O}$$

Note that the last equality comes from the fact that $P_1+P_2+P_3$ must have order at most 2 in $E(\Bbbk)$ since $2(P_1+P_2+P_3) = 2P_1 + 2P_2 + 2P_3 = \mathcal{O}$ but if, for example, $P_1 + P_2 + P_3 = P_1$ we get $P_2 + P_3 = \mathcal{O} \iff P_2 = -P_3$ and we know that this is not true (since $-P_3 = P_3$), hence $P_1 + P_2 + P_3 = \mathcal{O}$. This leads us to a key

Theorem 6.49. *Let $E/\Bbbk$ be an elliptic curve over an algebraically closed field $\Bbbk$.*

1. *Let r and r' be rational functions on E. If $\operatorname{div}(r) = \operatorname{div}(r')$ there exists a non-zero constant $c \in \Bbbk$ such that $r = cr'$.*
2. *Let $D = \sum_{P\in E(\Bbbk)} n_P[P]$ be a divisor. Then there exists a rational function $r \in \Bbbk(E)$ such that $\operatorname{div}(r) = D$ if and only if:*
   - $\deg(D) = 0$.
   - $\operatorname{sum}(D) = 0$.

In particular, if a rational function $r \in \Bbbk(E)$ has no zeros and no poles, it is constant.

Example 6.50. Let $E = E_{A,B}$ be an elliptic curve. Suppose $P \in E(\Bbbk)$ has order m, ie $mP = \mathcal{O}$. By Theorem 6.49 there exists a rational function f_P such that $\operatorname{div}(f_P) = m[P] - m[\mathcal{O}]$. The case $m = 2$ is particularly simple. We saw that points of order 2 are of the form $P = (\alpha, 0)$ where α is a root of $x^3 + Ax + B$. As we saw in Example 6.43, $\operatorname{div}(x - \alpha) = 2[P] - 2[\mathcal{O}]$.

For our last result in this section, let us define the **support** of a divisor $\sum_P n_P[P]$ to be the points $P \in E(\Bbbk)$ such that $n_P \neq 0$. We will need the following

Definition 6.51. Let $r \in \Bbbk(E) \smallsetminus \{0\}$ be a rational function and $D = \sum_P n_P[P]$ be a divisor whose support does not include zeros or poles of r. Then the function r **evaluated** at D is

$$r(D) := \prod_{P \in E(\Bbbk)} r(P)^{n_P}.$$

6.5 Isogenies

Let $E/\Bbbk$ be an elliptic curve. We discussed the notion of rational functions on E and saw that they give rise to functions $E(\Bbbk) \longrightarrow \mathbb{P}^1_{\Bbbk} = \Bbbk \cup \{\infty\}$. Since $\mathbb{P}^1$ is a curve, one may wonder if it is possible to extend the notion of rational function to a map $E(\Bbbk) \longrightarrow E(\Bbbk)$ between the elliptic curve to itself. This is indeed possible as we will see below.

Definition 6.52. Let

$$E/\Bbbk = E_{A,B} : y^2 = x^3 + Ax + B$$

be an elliptic curve. A **rational map** $\rho : E \longrightarrow E$ is a pair $\rho = (r, s)$ where $r, s \in \Bbbk(E)$ are rational functions on E such that for all $P \in E(\Bbbk)$,

$$s(P)^2 = r(P)^3 + Ar(P) + B.$$

In particular, $r(P) = \infty$ if and only if $s(P) = \infty$.

A rational map $\rho = (r, s) : E \longrightarrow E$ induces a map $\rho : E(\Bbbk) \longrightarrow E(\Bbbk)$ given by $P \mapsto (r(P), s(P))$ if $r(P), s(P) \neq \infty$ and $P \mapsto \mathcal{O}$ if $r(P) = s(P) = \infty$.

Example 6.53. Let $E/\Bbbk$ be an elliptic curve and let $1 \leq n$. The map $[n] : E(\Bbbk) \longrightarrow E(\Bbbk)$ given by $[n](P) := nP$ is a rational map. The construction of Section 6.1, gives rational functions $r, s \in \Bbbk(E)$ such that $[n] = (r, s)$.

Example 6.54. Let E be an elliptic curve and $Q \in E(\Bbbk)$. The map $\tau_Q : E \longrightarrow E$ given by $\tau_Q(P) = P + Q$ is a rational map.

As usual, we have:

Proposition 6.55. *Let $E/\Bbbk$ be an elliptic curve and $\rho, \tau : E \rightrightarrows E$ be two rational maps. Then $\tau \circ \rho : E \longrightarrow E$ is a rational map.*

Proof. Suppose $\rho = (r, s)$ and $\tau = (u, v)$. Then

$$(\tau \circ \rho)(x, y) = \Big(u\big(r(x, y), s(x, y)\big), v\big(r(x, y), s(x, y)\big)\Big)$$

which is a pair of rational functions since each coordinate is a substitution of rational functions into rational functions. Since for every $P \in E(\Bbbk)$ we have $\big(r(P), s(P)\big) \in E(\Bbbk)$ (and similarly for u, v), we have $(\tau \circ \rho)(P) \in E(\Bbbk)$, so that the pair of rational functions representing $\tau \circ \rho$ satisfies the equation of E, ie $\tau \circ \rho$ is again a rational map. □

Rational maps have a rather rigid structure, as the following proposition shows:

Proposition 6.56. *Let $E/\mathbb{k}$ be an elliptic curve over field $\mathbb{k}$ and $\rho = (r,s) : E \longrightarrow E$ be a rational map. If ρ is non-constant, then it induces a surjective map $E(\overline{\mathbb{k}}) \longrightarrow E(\overline{\mathbb{k}})$.*

Proof. We first show the analogous assertion for rational functions $E \longrightarrow \mathbb{P}^1$. If $r \in \mathbb{k}(E)$ is a non-constant rational function, then it must have a zero by Theorem 6.49. For $x_0 \in \mathbb{k}$, the same argument as above, applied to $r - x_0$, implies that $r - x_0$ has a zero, so there is $P \in E(\mathbb{k})$ such that $r(P) = x_0$. Thus, r is surjective.

Now consider the rational map $\rho = (r,s) : E(\mathbb{k}) \longrightarrow E(\mathbb{k})$. If r is constant, then since $\forall P \in E(\mathbb{k})$

$$s(P)^2 = r(P)^3 + Ar(P) + B,$$

we get that s can take at most two values, namely the roots of $r(P)^3 + Ar(P) + B$, hence must be constant by the argument above.

Otherwise, r is surjective. It has a zero hence a pole by Theorem 6.49. In particular there is $P \in E(\mathbb{k})$ such that $r(P) = \infty$ so that $s(P) = \infty$ as well and $\rho(P) = \mathcal{O}$. Let $Q \in E(\mathbb{k})$. The map $\tau_{-Q} : E \longrightarrow E$ is a rational map by Example 6.54, and the map $\tau_{-Q} \circ \rho$ is a rational map as a composition of such (Proposition 6.55). The same argument as above, shows that there is $P' \in E(\mathbb{k})$ such that $(\tau_{-Q} \circ \rho)(P') = \mathcal{O}$. But then, $\rho(P') = Q$ so that ρ is surjective. □

We now proceed to study a specific rational map that plays a key role in the study of elliptic curves over finite fields. Throughout the rest of this section, let $\mathbb{F}_q$ be a finite field of characteristic p, so that $q = p^k$ for some $1 \leq k$.

Definition 6.57. The map $\Phi_q : \overline{\mathbb{F}}_q \longrightarrow \overline{\mathbb{F}}_q$ given by $\Phi_q(x) = x^q$ is called the **Frobenius endomorphism**. If $E/\mathbb{F}_q$ is an elliptic curve defined over $\mathbb{F}_q$, Φ_q act on the coordinates of points in $E(\overline{\mathbb{F}}_q)$ by

$$\begin{aligned} \Phi_q(x,y) &= (x^q, y^q) \\ \Phi_q(\mathcal{O}) &= \mathcal{O} \end{aligned} \tag{26}$$

Proposition 6.58. *Let $E/\mathbb{F}_q$ be an elliptic curve defined over $\mathbb{F}_q$ and $(x,y) \in E(\overline{\mathbb{F}}_q)$. Then*

1. *$\Phi_q(x,y) = E(\overline{\mathbb{F}}_q)$ so that Φ_q defines a rational map $\Phi_q : E \longrightarrow E$ that we call (with slight abuse) the Frobenius endomorphism.*

2. *$(x,y) \in E(\mathbb{F}_q) \iff \Phi_q(x,y) = (x,y)$.*

Proof. Note first that $\overline{\mathbb{F}}_q$ is a field of characteristic p.

1. Recall from field theory that in a field of characteristic p, for any a, b, $(a+b)^p = a^p + b^p$. Thus, $(a+b)^{p^2} = ((a+b)^p)^p = (a^p + b^p)^p = a^{p^2} + b^{p^2}$. By trivial induction on k, we get that for any $a, b \in \overline{\mathbb{F}}_q$, $(a+b)^q = a^q + b^q$. Recall also that for any $a \in \mathbb{F}^q$ $a^q = a$ since the multiplicative group $\mathbb{F}_q^\times$ is cyclic group of order $q-1$. Since $(x,y) \in E(\overline{\mathbb{F}}_q)$,

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

where $A, B \in \mathbb{F}_q$ since E is defined over $\mathbb{F}_q$. Thus

$$\begin{aligned} (y^2)^q &= (x^3 + Ax + B)^q \\ \iff (y^q)^2 &= (x^q)^3 + A^q x^q + B^q \\ \iff (y^q)^2 &= (x^q)^3 + Ax^q + B \\ \iff (x^q, y^q) &\in E(\overline{\mathbb{F}}_q). \end{aligned} \tag{27}$$

2. Note that for $a \in \overline{\mathbb{F}}_q$, $a^q = a \iff a \in \mathbb{F}_q$ since this all elements $a \in \mathbb{F}_q$ satisfy it and the polynomial $x^q - x$ can have at most q distinct roots in $\overline{\mathbb{F}}_q$. Thus,

$$(x,y) \in E(\mathbb{F}_q) \iff x,y \in \mathbb{F}_q \iff \Phi(x,y) = (x,y).$$

□

Corollary 6.59. *Let $E/\mathbb{F}_q$ be an elliptic curve and $\Phi_q : E \longrightarrow E$ the Frobenius endomorphism. Then for any rational map $g : E(\mathbb{F}_q) \longrightarrow \mathbb{P}^1_{\mathbb{F}_q}$ we have a commutative diagram*

$$\begin{tikzcd}
E(\mathbb{F}_q) \arrow[r, "\Phi_q"] \arrow[d, "g"] & E(\mathbb{F}_q) \arrow[d, "g"] \\
\mathbb{P}^1_{\mathbb{F}_q} \arrow[r, "\Phi_q"] & \mathbb{P}^1_{\mathbb{F}_q}
\end{tikzcd}$$

where the bottom Φ_q sends ∞ to ∞.

Proof. Follows immediately from Proposition 6.58 □

It is natural to require that a rational map between elliptic curve will be compatible with the corresponding group structures. This is a repeating theme when studying mathematical objects – one would like to have a notion of a structure-preserving map (e.g. group homomorphism, linear map of vector spaces etc.). This is made precise by the following

Definition 6.60. Let E, E' be two elliptic curves defined over $\Bbbk$. A rational map $\alpha : E' \longrightarrow E$ is called an **isogeny** if it induces a group homomorphism $\alpha : E(\overline{\Bbbk}) \longrightarrow E'(\overline{\Bbbk})$ ie $\alpha(P_1 + P_2) = \alpha(P_1) + \alpha(P_2)$. An isogeny $\alpha : E \longrightarrow E$ from E to itself is called an **endomorphism**.

Remark 6.61. By definition, an isogeny α induces a group homomorphism $E(\mathbb{L}) \longrightarrow E'(\mathbb{L})$ for any intermediate $\Bbbk \subseteq \mathbb{L} \subseteq \overline{\Bbbk}$ since the restriction of a group homomorphism to subgroups is again a group homomorphism.

Remark 6.62. Since a composition of rational maps is again a rational map, and a composition of group homomorphisms is a group homomorphism, the same holds for isogenies. Let $\alpha, \beta : E \rightrightarrows E$ be two endomorphisms. Their sum is the endomorphism $\alpha + \beta : E \longrightarrow E$ given by $(\alpha + \beta)(P) := \alpha P + \beta P$. Thus, the set of all endomorphism of E, denoted $\mathrm{End}(E)$, admits the structure of a ring with addition as described above and multiplication being composition. Note that this ring is generally **not commutative**.

Isogeny is the natural notion for a morphism between elliptic curves. Building on that, we make the following

Definition 6.63. Let $E/\Bbbk, E'/\Bbbk$ be two elliptic curves (over the same field). An isogeny $\alpha : E \longrightarrow E'$ is called an **isomorphism** if there exists an isogeny $\alpha^{-1} : E' \longrightarrow E$ such that $\alpha \circ \alpha^{-1} = \mathrm{id}_{E'}$ and $\alpha^{-1} \circ \alpha = \mathrm{id}_E$.

Exercise 6.64. For any elliptic curve E, multiplication by n is an endomorphism $[n] : E \longrightarrow E$. Show that $[n] + [m] = [m + n]$.

Another interesting example is given in the following

Proposition 6.65. *Let $E/\mathbb{F}_q$ be an elliptic curve. Then the Frobenius map $\Phi_q : E \longrightarrow E$ is an endomorphism.*

Proof. Clearly, Φ_q is a rational map. Let $(x_1, y_1), (x_2, y_2) \in E(\overline{\mathbb{F}}_q)$ be two points with $x_1 \neq x_2$. Their sum is (x_3, y_3) with

$$\begin{aligned} x_3 &= m^2 - x_1 - x_2 \\ y_3 &= m(x_1 - x_3) - y_1 \\ m &= \frac{y_2 - y_1}{x_2 - x_1}. \end{aligned} \tag{28}$$

Raise all equations by q to get

$$\begin{aligned} m' &= \frac{y_2^q - y_1^q}{x_2^q - x_1^q} \\ x_3^q &= (m')^2 - x_1^q - x_2^q \\ y_3 &= m'(x_1^q - x_3^q) - y_1^q. \end{aligned} \tag{29}$$

(recall that in a field of characteristic p, $(x+y)^q = x^q + y^q$ for every $q = p^n$). This says that

$$\Phi_q(x_3, y_3) = \Phi_q(x_1, y_1) + \Phi_q(x_2, y_2).$$

The case of $x_1 = x_2$ is checked similarly. □

Let $\alpha : E/\Bbbk \longrightarrow E'/\Bbbk$ be an isogeny and denote $\alpha(x, y) = (R_1(x, y), R_2(x, y))$ where R_1, R_2 are rational functions on E. Recall that a rational function R on E admits a canonical form

$$R(x, y) = \frac{q_1(x) + q_2(x)y}{q_3(x)}. \tag{30}$$

Since α a group homomorphism, $\alpha(x, -y) = \alpha(-(x, y)) = -\alpha(x, y)$. This means that $R_1(x, y) = R_1(x, -y)$ and $R_2(x, -y) = -R_2(x, y)$ and thus when R_1 is written in the form 30, $q_2(x) = 0$ whereas when R_2 is written in the form 30, $q_1(x) = 0$. We conclude that α has a canonical form

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

for rational functions $r_1(x) = \frac{p(x)}{q(x)}$ and $r_2(x) = \frac{u(x)}{v(x)}$ where each quotient is reduced, ie has no common (polynomial) factors. If $q(x) = 0$ for some $(x, y) \in E(\overline{\Bbbk})$, we set $\alpha(x, y) = \mathcal{O}$, and otherwise, we have shown that $v(x) \neq 0$ so $\alpha(x, y)$ is well-defined.

We will be interested in two main properties of isogenies.

Definition 6.66. Let $\alpha : E/\Bbbk \longrightarrow E'/\Bbbk$ be an isogeny.

- The **degree** of α is $\deg(\alpha) = \max\{\deg(p), \deg(q)\}$ and when $\alpha = 0$ we set $\deg(\alpha) = 0$.
- α is called **seperable** if the formal derivative $r_1'(x)$ is not identically zero (equivalently if $p(x)'q(x) - p(x)q'(x)$ is not identically zero).

Example 6.67. Clearly $\deg \Phi_q = q$ as it is given by polynomials of degree q. However, Φ_q is not seperable since $r_1(x) = x^q$ so that $r_1'(x) = qx^{q-1} = 0$.

Exercise 6.68. Show that multiplication by 2 is a seperable map. What is its degree? does your argument generalises to $2 < n$?

we have the following basic

Lemma 6.69. *Let $E/\Bbbk$ be an elliptic curve and $E \underset{\alpha'}{\overset{\alpha}{\rightrightarrows}} E$. Then*

$$\deg(\alpha \circ \alpha') = \deg(\alpha)\deg(\alpha').$$

Proof. Let $(R(x), S(x)y), (R'(x), S'(x)y)$ be the canonical forms of α, α' respectively. Then $(R''(x), S''(x)y) := (R(R'(x)), S(R'(x))S'(x)y)$ is a representation of $\alpha \circ \alpha'$. Moreover, $R(R'(x))$ is reduced since $R(x)$ and $R'(x)$ are. Thus, $(R''(x), S''(x)y)$ can be made a canonical form without changing $R(R'(x))$ (but possibly removing common factors in $S(R'(x))S'(x)$). Observe that for two polynomials $p(x), q(x)$, $\deg p(q(x)) = \deg p \deg q$. Suppose our rational functions take the reduced form $R(x) = u(x)/v(x)$ and $R'(x) = u'(x)/v'(x)$. Then $R(R') = u(u'/v')/v(u'/v')$ so the degree of the numerator is $\deg u \cdot \max\{\deg u', \deg v'\}$ and the degree of the denominator is $\deg v \cdot \max\{\deg u', \deg v'\}$ (as rational functions). It follows that

$$\deg(\alpha \circ \alpha') = \max\{\deg u \max\{\deg u', \deg v'\}, \deg v \max\{\deg u', \deg v'\}\} = \deg \alpha \deg \alpha'.$$

□

Corollary 6.70. *Let $E/\Bbbk$ be an elliptic curve and $\alpha : E \longrightarrow E$ an isomorphism. Then* $\deg \alpha = 1$.

Proof. By definition, there exists an isogeny α^{-1} such that $\alpha \circ \alpha^{-1} = \mathrm{id}$. Since $\deg(\mathrm{id}) = 1$, Lemma 6.69 implies $\deg \alpha = 1$. □

The notion of seperability may seem opaque at first glance. The following proposition helps clarify the matter.

Proposition 6.71. *Let $\alpha : E \longrightarrow E'$ be a non-zero isogeny. If α is seperable then $\#\ker(E(\overline{\Bbbk}) \xrightarrow{\alpha} E'(\Bbbk)) = \deg(\alpha)$ and otherwise $\#\ker(\alpha) < \deg(\alpha)$.*

Proof. Since α is a non-zero homomorphism, it is not constant and hence by Proposition 6.56 induces a surjective map $\alpha : E(\overline{\Bbbk}) \longrightarrow E'(\Bbbk)$. Let $\mathcal{O} \neq Q = (a, b) \in E'(\overline{\Bbbk})$. There exists $(x_0, y_0) \in E(\overline{\Bbbk})$ such that $\alpha(x_0, y_0) = (a, b)$. Since $E'(\overline{\Bbbk})$ is infinite, we can choose (a, b) such that:

1. $a, b \neq 0$.
2. $\deg(p - aq) = \max\{\deg(p), \deg(q)\} = \deg(\alpha)$ – the only way that $\deg(p - aq) < \deg(\alpha)$ is possible is if $\deg(p) = \deg(q)$ and their leading coefficients, λ, δ satisfy $\lambda - a\beta = 0$ and in this case we only need to restrict $a \neq \lambda/\delta$.

Since $\deg(p - aq) = \deg(\alpha)$, it has $\deg(\alpha)$ (possibly indistinct) roots. We claim that the number of distinct roots of $p(x) - aq(x)$ corresponds to the the number of pre-image points $P = (x_0, y_0)$ of Q: since $(a, b) \neq (\infty, \infty)$ we must have $q(x_0) \neq 0$ and since $b \neq 0$ and $y_0 r_2(x_0) = b$ we have $y_0 = b/r_2(x_0)$ ie x_0 determines y_0 and it is thus enough to count only x_0. Since α is a homomorphism, the number of pre-image points Q of any point $Q \in E'(\Bbbk)$ equals $\#\ker(\alpha)$ so it suffices to analyse when $p(x) - aq(x)$ has repeated roots. The polynomial $p(x) - aq(x)$ has a repeated root at x_0 iff $p(x_0) - aq(x_0) = 0$ and $p'(x_0) - aq'(x_0) = 0$. Multiplying the two equations, we get $ap(x_0)q'(x_0) = ap'(x_0)q(x_0)$ and since $a \neq 0$, we get $p(x_0)q'(x_0) - p'(x_0)q(x_0) = 0$. Now, if α is not seperable, then $p(x)q'(x) - p'(x)q(x)$ is identically zero, so we must have a repeated root and hence $\#\ker(\alpha) < \deg(\alpha)$. On the other hand, if α is seperable, $p(x)q'(x) - p'(x)q(x)$ is not identically zero, and hence a finite number of roots S. We may further restrict out choice of a to satisfy $a \notin r_1(S)$. With such a choice of a, $x_0 \notin S$ sp $p(x) - aq(x)$ has no repeated roots and hence $\#\ker(\alpha) = \deg(\alpha)$ as required. □

6.6 The j-invariant of an elliptic curve

Recall that our setup is an elliptic curve in short Weirstrauss form $E/\Bbbk = E_{A,B} : y^2 = x^3 + Ax + B$ where $\Bbbk$ is a field of positive characteristic such that $\operatorname{char}\Bbbk \neq 2, 3$. Since we wanted to deal with smooth curves (ie where tangent line is well-defined and unique at any point), we required that $\Delta(E) = 4A^3 + 27B^2 \neq 0$.

Definition 6.72. The **j-invariant** of an elliptic curve $E = E_{A,B}$ is

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

Remark 6.73. Note that if $A = 0$, $j(E) = 0$ and if $B = 0$, $j(E) = 1728$.

The key property of the j-invariant is that it characterises E up to isomorphism over $\overline{\Bbbk}$. Before proving that, we first note that every element in $\Bbbk$ is the j-invariant of some elliptic curve over $\Bbbk$.

Theorem 6.74. *For every $j_0 \in \Bbbk$ there exists an elliptic curve $E = E_{A,B}/\Bbbk$ such that $j(E) = j_0$.*

Proof. If j_0 is 0 or 1728 we take $E : y^2 = x^3 + 1$ or $E : y^2 = x^3 + x$ respectively. Otherwise, we require A, B such that $j_0(4A^3 + 27B^2) = 1728 \cdot 4A^3$, which can be rewritten as:

$$4A^3(1728 - j_0) = 27j_oB^2$$

Rearranging slightly:

$$A^3 2^2(1728 - j_0) = 3^3 j_o B^2$$

We want each factor of this equation to be a power that is a multiple of 2 or 3. This is achieved by multiplying both sides by $j_0^2(1728 - j_0)^3$. We obtain:

$$A^3 \cdot [2j_0(1728 - j_0)^2]^2 = [3j_o(1728 - j_0)]^3 \cdot B^2$$

It is therefore apparent that $j(A, B) = j_0$ for the following values of A and B:

$$A = 3j_0(1728 - j_0)$$
$$B = 2j_0(1728 - j_0)^2$$

□

Recall from Corollary 6.70 that an isomorphism $\alpha : E \longrightarrow E$ (over $\Bbbk$) has $\deg\alpha = 1$.

Theorem 6.75. *Elliptic curves $E : y^2 = x^3 + Ax + B$ and $E' : y^2 = x^3 + A'x + B'$ defined over $\Bbbk$ are isomorphic (over $\Bbbk$) iff $A' = \mu^4 A$ and $B' = \mu^6 B$ for some $\mu \in \Bbbk^\times$.*

Proof. Let $\alpha : E \longrightarrow E'$ be an isomorphism in standard form $\alpha(x, y) = (R(x), S(x)y)$ with R, S rational functions over $\Bbbk$. Since α is an isomorphism, $\ker\alpha := \ker(\alpha : E(\overline{\Bbbk}) \longrightarrow E'(\overline{\Bbbk})) = \mathcal{O}$ which means that both R and S must be polynomials: if any of them had a non-trivial denominator, it would have a root over $\overline{\Bbbk}$ which would be the x-coordinate of a point in $\ker\alpha$. Furthermore, since $\deg\alpha = 1$ we R must be of degree 1 ie

$$R(x) = ax + b$$

for some $a, b \in \Bbbk$ and $a \neq 0$. Since α is a rational function, we get

$$S^2(x)y^2 = (ax + b)^3 + A'(ax + b) + B'$$
$$S^2(x)(x^3 + Ax + B) = (ax + b)^3 + A'(ax + b) + B'.$$

Comparing degrees, we see that $S(x)$ must be constant, say $S(x) = c$. Considering the coefficient of x^2 we see that $b = 0$ and comparing coefficients of x^3 we see that $c^2 = a^3$ so that $a = (c/a)^2$. If we let $\mu = c/a \in \Bbbk^\times$, so that $a = \mu^2$, we see that

$$\mu^6(x^3 + Ax + B) = \mu^6 x^3 + A'\mu^2 x + B'$$

and it follows that $A' = \mu^4 A$ and $B' = \mu^6 B$ as claimed. Conversely, if $A' = \mu^4 A$ and $B' = \mu^6 B$ for some $\mu \in \Bbbk^\times$, then the map $\alpha : E \longrightarrow E'$ defined by

$$\alpha(x, y) = (\mu^2 x, \mu^3 y)$$

is an isogeny (as it is readily a rational map that preserves $\mathcal{O}$) and it has an inverse $\alpha^{-1} : E' \longrightarrow E$ given by

$$\alpha^{-1}(x, y) = (\frac{1}{\mu^2} x, \frac{1}{\mu^3} y).$$

□

We are ready to prove the main result of this section

Theorem 6.76. *Let E, E' be elliptic curves defined over a field $\Bbbk$. Then E and E' are isomorphic over $\overline{\Bbbk}$ iff $j(E) = j(E')$. If $j(E) = j(E')$ and the characteristic of $\Bbbk$ is not $2, 3$, there exists a field extension $\mathbb{L}/\Bbbk$ of degree at most $6, 4$ or 2, depending on whether $j(E) = 0$, $j(E) = 1728$ or $j(E) \neq 0, 1728$ (respectively) such that E and E' are isomorphic over $\mathbb{L}$.*

Proof. We will assume that $\operatorname{char} \Bbbk \neq 2, 3$. Suppose that $E : y^2 = x^3 + Ax + B$ and $E' : y^2 = x^3 + A'x + B'$ are defined over $\Bbbk$ and isomorphic over $\overline{\Bbbk}$. Then applying Theorem 6.75, there is some $\mu \in \overline{\Bbbk}^\times$ such that $A' = \mu^4 A$ and $B' = \mu^6 B$. Then

$$j(E') = 1728 \frac{4(\mu^4 A)^3}{4(\mu^4 A)^3 + 27(\mu^6 B)^2} = 1728 \frac{4A^3}{4A^3 + 27B^2} = j(E).$$

For the converse, suppose $j(E) = j(E') = j_0$. If $j_0 = 0$ ie $A = A' = 0$ and $B, B' \neq 0$, we want μ such that $B' = \mu^6 B$. We may choose μ to be any root of the polynomial $g(x) = x^6 - B'/B$ but this root may not exist in $\Bbbk$. If $g(x) \in \Bbbk[x]$, Corollary 5.77 asserts that there exists an extension $\mathbb{L}$ of $\Bbbk$ of degree $[\mathbb{L} : \Bbbk] = \deg g = 6$ such that $g(x)$ has a root in $\mathbb{L}$. If $g(x)$ is not prime, choose a prime factor $q(x) | g(x)$. Then Corollary 5.77 asserts that there exists an extension $\mathbb{L}$ of $\Bbbk$ of degree $[\mathbb{L} : \Bbbk] = \deg q < 6$ such that $q(x)$ has a root in $\mathbb{L}$. Similarly, if $j_0 = 1728$ ie $B = B' = 0$ and $A, A' \neq 0$, we may choose $\mu \in \mathbb{L}^\times$ such that $B' = \mu^4 B$ by taking an extension $\mathbb{L}$ of $\Bbbk$ of degree at most 4. In both these case, we may then aply Theorem 6.75 to get an isomorphism of E and E' over an extension $\mathbb{L}$ of $\Bbbk$.

Suppose now that $j_0 \neq 0, 1728$. Because $\mu^4 = A'/A$ and $\mu^6 = B'/B$, we have:

$$\mu^{12} = \frac{B'^2}{B^2} = \frac{A'^3}{A^3} \implies 1 = \frac{B^2}{B'^2} \frac{A'^3}{A^3} = \frac{B}{B'} (\frac{B}{B'} \frac{A'^2}{A^2}) \frac{A'}{A}$$

Define u as follows:

$$u = \frac{B}{B'} \frac{A'^2}{A^2}$$

This implies:

$$1 = u \frac{B}{B'} \frac{A'}{A}$$

Squaring this equation, we obtain:

$$1 = u^2 \frac{B^2}{B'^2} \frac{A'^2}{A^2} = u^2 (\frac{B^2}{B'^2} \frac{A'^3}{A^3}) \frac{A}{A'} = u^2 \frac{A}{A'}$$

Cubing it yields:

$$1 = u^3 \frac{B^3}{B'^3} \frac{A'^3}{A^3} = u^2 \frac{B}{B'} (\frac{B^2}{B'^2} \frac{A'^3}{A^3}) = u^3 \frac{B}{B'}$$

We now choose $\mu \in \mathbb{L}^\times$ such that $\mu^2 = u$ where $\mathbb{L}$ is an extension of degree at most 2. Then $A' = \mu^4 A$ and $B' = \mu^6 B$ and Theorem 6.75 asserts that E and E' are isomorphic over $\mathbb{L}$

□

In light of the last Theorem, let us make the following

Definition 6.77. Let E and E' be elliptic curves over defined over $\mathbb{F}_q$. Then E' is called a **twist** of E if there exists an isomorphism $\phi : E' \longrightarrow E$ over some extension $\mathbb{F}_q \subseteq \mathbb{F}_{q^d}$. the minimal d for which such an isomorphism exists is called the **degree** of the twist.

We then have the following corollary

Corollary 6.78. *Let E be an elliptic curve defined over $\mathbb{F}_q$. Then twists of E must have degree* 2*,* 3*,* 4 *or* 6*.*

6.7 Weil Reciprocity

Our final piece of theory before going to pairings is a result used by Weil to construct what is now known as the Weil Pairing. The statement is as follows.

Theorem 6.79 (Weil reciprocity)**.** *Let $E/\mathbb{k}$ be an elliptic curve defined over an algebraically closed field. If $r, s \in \mathbb{k}(E) \setminus \{0\}$ are rational functions whose divisors have disjoint support, then*

$$r(\mathrm{div}(s)) = s(\mathrm{div}(r))$$

Weil Reciprocity is in fact a general property of "projective" curves (ie ones with an additional point at infinity), not just elliptic curves.

The proof of Weil reciprocity for elliptic curves is carried out in two stages. In the first stage, one proves Weil reciprocity for the projective line $\mathbb{P}^1_{\mathbb{k}}$ (see Definition 6.80 below). In the second stage, one uses a formal argument for projective curves to "transfer" the proof from the projective line to a general elliptic curve.

We will devote the remainder of this section to formulate and prove the first stage of Theorem 6.79, ie in the case of the projective line. We omit the second stage since it requires to develop general theory for curves, which we believe is too big of a digression to take in these notes. We hope that the proof of Weil reciprocity for the projective line will give the reader a feeling why it should be true for elliptic curves as well.

We start with

Definition 6.80. Let $\mathbb{k}$ be a field. The **projective line** over $\mathbb{k}$, denoted $\mathbb{P}^1 = \mathbb{P}^1_{\mathbb{k}}$ is the set

$$\mathbb{P}^1 = \mathbb{k} \cup \{\infty\}.$$

Just like an elliptic curve E is the set of all solutions of a polynomial $f_E(x,y)$ + a point at infinity, the projective line can be viewed as the set of all solutions of the zero polynomial + a point at infinity. Thus, the projective line is another example of a curve.

Warning 6.81. Unlike an elliptic curve, $\mathbb{P}^1_{\mathbb{k}}$ does not have a group structure.

Definition 6.82. Let $\mathbb{k}$ be a field. A **rational function** on $\mathbb{P}^1 = \mathbb{P}^1_{\mathbb{k}}$ is a quotient of polynomials $r(x) = \frac{u(x)}{v(x)}$. We denote the set of all rational functions on $\mathbb{P}^1$ by $\mathbb{k}(\mathbb{P}^1)$.

As before, we can evaluate a rational function at a point:

Definition 6.83. Let $r \in \Bbbk(\mathbb{P}^1)$. The **induced function** from r is the function $r : \mathbb{P}^1 \longrightarrow \mathbb{P}^1$ defined as follows. If $a \neq \infty$, $r(a) = \frac{u(a)}{v(a)}$ except in the case a is a root of $v(x)$, in which case $r(a) : \infty$. If $a = \infty$, $r(a) := \infty$ if $\deg u > \deg v$, $r(a) = 0$ if $\deg u < \deg v$ and if $\deg u = \deg v$, with $u(x) = \sum_{i=1}^n a_i x^i$ and $v(x) = \sum_{i=1}^n b_i x^i$ then $r(a) := \frac{a_n}{b_n}$.

The notion of zeros and poles of a rational function on $\mathbb{P}^1$ is the same as before:

Definition 6.84. Let $r \in \Bbbk(\mathbb{P}^1)$. A point $a \in \mathbb{P}^1$ is called a **zero** of r if $r(a) = 0$ and is called a **pole** of r if $r(a) = \infty$.

We now want to define the multiplicity/order of zeros and poles. Since we are dealing with polynomials in one variable the situation is simpler than with elliptic curves. The only subtlety is the multiplicity of the point at infinity ∞.

Definition 6.85. Let $r \in \Bbbk(\mathbb{P}^1)$ be a rational function and $a \in \mathbb{P}^1$ a point. Suppose $r(x) = \frac{u(x)}{v(x)}$ where u, v do not have common factors (so in particular have a disjoint set of roots).

1. If a is not a zero or a pole of r, we set $\operatorname{ord}_a(r) = 0$.
2. If $a \neq \infty$:
   - If a is a zero of r, its order $\operatorname{ord}_a(r)$ is the largest positive integer n such that $(x-a)^n | u(x)$.
   - If a is a pole of r, its order $\operatorname{ord}_a(r)$ is $-n$ where n is the largest positive integer such that $(x-a)^n | v(x)$.
3. If $a = \infty$: we define $\operatorname{ord}_\infty(r) = \deg v - \deg u$.

Remark 6.86. Note that by definition, we always have $\operatorname{ord}_\infty(r) = \deg v - \deg u$, since when ∞ is not a zero or a pole of r, $\deg u = \deg v$ so that $\operatorname{ord}_\infty(r) = \deg v - \deg u = 0$.

Next, we define divisors on $\mathbb{P}^1$:

Definition 6.87. A **divisor** on $\mathbb{P}^1$ is a formal sum $D = \sum_{a \in \mathbb{P}^1} n_a[a]$ where the n_a's are integers, and only finitely many of them are non-zero.

Similarly to the case of elliptic curves, we can associate a divisor to every rational function:

Definition 6.88. Let $r \in \Bbbk(\mathbb{P}^1)$ be a rational function. The **associated divisor** of r is

$$\operatorname{div}(r) = \sum_{a \in \mathbb{P}^1} \operatorname{ord}_a(r)[a].$$

We now have a simple version of Theorem 6.49:

Proposition 6.89. *Let $r = \frac{u}{v} \neq 0 \in \Bbbk(\mathbb{P}^1)$ be a rational function such that both u and v decompose to linear factors over $\Bbbk$ (eg $\Bbbk$ is algebraically closed). Write* $\operatorname{div}(r) = \sum_{a \in \mathbb{P}^1} \operatorname{ord}_a(r)[a]$. *Then*

$$\sum_{a \in \mathbb{P}^1} n_a = 0.$$

Proof. Since u and v decompose to linear factors, we can write

$$u(x) = \prod_{i=1}^{m} (x - a_i)^{n_{a_i}}$$

and

$$v(x) = \prod_{j=1}^{m'} (x - b_j)^{n_{b_j}}$$

where

$$d_u = \deg u = \sum_{i=1}^{m} n_{a_i}$$

and

$$d_v = \deg v = \sum_{j=1}^{m'} n_{b_j}.$$

Clearly, each a_i is a zero of r with multiplicity/order n_{a_i} and each b_j is a pole of r with multiplicity/order $-n_{b_j}$. Thus, the associated divisor of r can be written as

$$\operatorname{div}(r) = \sum_{i=1}^{m} n_{a_i}[a_i] - \sum_{j=1}^{m'} n_{b_j}[b_j] + \operatorname{ord}_\infty(r)[\infty]$$

and the sum of the coefficients can be written as

$$S = \sum_{i=1}^{m} n_{a_i} - \sum_{j=1}^{m'} n_{b_j} + \operatorname{ord}_\infty(r) = d_u - d_v + \operatorname{ord}_\infty(r).$$

If $d_u = \deg u = \deg v d_v$ then by definition ∞ is neither a zero nor a pole of r so $\operatorname{ord}_\infty(r) = 0$ and in this case $S = d_u - d_v = 0$.

If $d_u > d_v$ then by definition ∞ is a pole and $\operatorname{ord}_\infty(r) = -(d_u - d_v)$ so that again $S = 0$.

Lastly, if $d_u < d_v$, then by definition ∞ is a zero and

$$\operatorname{ord}_\infty(r) = d_v - d_u = -(d_u - d_v)$$

so that $S = 0$ as well.

□

As in the section on divisors that the **support** of a divisor $D = \sum_{a \in \mathbb{P}^1} n_a[a]$ is the set of points $a \in \mathbb{P}^1$ for which $n_a \neq 0$. If $r \in \Bbbk(\mathbb{P}^1)$ is a rational function and $D = \sum_{a \in \mathbb{P}^1} n_a[a]$ a divisor whose support does not contain zeros or poles of r, the **evaluation** of r at D is

$$r(D) = \prod_{a \in \mathbb{P}^1} r(a)^{n_a}.$$

We are now ready to prove Weil Reciprocity for the projective line.

Theorem 6.90 (Weil Reciuprocity for the projective line)**.** *Let $\Bbbk$ be algebraically closed and $r, s \in \Bbbk(\mathbb{P}^1)$ two rational functions with disjoint support. Then*

$$r(\operatorname{div}(s)) = s(\operatorname{div}(r)).$$

Proof. Write $r = u/v$ and $s = u'/v'$. Since $\Bbbk$ is algebraically closed, the polynomials $u, v, u'.v'$ decompose to linear factors. Thus, we can write

$$r(x) = \prod_{i=1}^{m} (x - a_i)^{n_{a_i}}$$

and

$$s(x) = \prod_{j=1}^{m'} (x - b_j)^{n_{b_j}}$$

where $n_{a_i}, n_{b_j} \in \mathbb{Z}$ are the orders of a_i, b_j respectively.

Write

$$\operatorname{div}(r) = \sum_{i=1}^{m} n_{a_i}[a_i] - n_\infty^r[\infty]$$

and

$$\operatorname{div}(s) = \sum_{j=1}^{m'} n_{b_j}[b_j] - n_\infty^s[\infty].$$

By Proposition 6.89, $n_\infty^r = \sum_{i=1}^m n_{a_i}$ and $n_\infty^s = \sum_{j=1}^{m'} n_{b_j}$.

Suppose first that the support of both r and s do not contain the point at infinity ∞, ie $\forall i, j$, $a_i \neq b_j$ and

$$\sum_{i=1}^{m} n_{a_i} = 0 = \sum_{j=1}^{m'} n_{b_j}.$$

Then

$$\begin{aligned} r(\operatorname{div}(s)) &= \prod_{j=1}^{m'} \left[\prod_{i=1}^{m} (b_j - a_i)^{n_{a_i}} \right]^{n_{b_j}} \\ &= \prod_{j=1}^{m'} \prod_{i=1}^{m} (b_j - a_i)^{n_{a_i} n_{b_j}} \\ &= (-1)^{\sum_{i=1}^m \sum_{j=1}^{m'} n_{a_i} n_{b_j}} \prod_{i=1}^{m} \prod_{j=1}^{m'} (a_i - b_j)^{n_{a_i} n_{b_j}} \\ &= s(\operatorname{div}(r)) \end{aligned} \tag{31}$$

where the sign equals 1 since

$$\sum_{i=1}^{m} \sum_{j=1}^{m'} n_{a_i} n_{b_j} = \left(\sum_{i=1}^{m} n_{a_i} \right) \left(\sum_{j=1}^{m'} n_{b_j} \right) = 0 \cdot 0 = 0.$$

If, without loss of generality, ∞ is in the support of $r(x) = \frac{u(x)}{v(x)}$ (so by assumption, ∞ is not in the support of s), then

$$\operatorname{div}(r) = \sum_{i=1}^{m} n_{a_i}[a_i] - n_\infty[\infty]$$

with

$$n_\infty = n_\infty^r = \sum_{i=1}^{m} n_{a_i} = \deg u - \deg v,$$

and

$$\operatorname{div}(s) = \sum_{j=1}^{m'} n_{b_j}[b_j]$$

with

$$\sum_{j=1}^{m'} n_{b_j} = 0$$

as before.

Then,

$$\begin{aligned} r(\mathrm{div}(s)) &= \left[\prod_{j=1}^{m'}\prod_{i=1}^{m}(b_j - a_i)^{n_{a_i}n_{b_j}}\right]\cdot\prod_{j=1}^{m'}(\infty - b_j)^{n_\infty n_{b_j}} = 1 \\ &= \left[(-1)^{\sum_{i=1}^{m}\sum_{j=1}^{m'} n_{a_i}n_{b_j}}\prod_{i=1}^{m}\prod_{j=1}^{m'}(a_i - b_j)^{n_{a_i}n_{b_j}}\right]\cdot\prod_{j=1}^{m'}(\infty - b_j)^{n_\infty n_{b_j}} \\ &= \left[\prod_{i=1}^{m}\prod_{j=1}^{m'}(a_i - b_j)^{n_{a_i}n_{b_j}}\right]\cdot\prod_{j=1}^{m'}(\infty - b_j)^{n_\infty n_{b_j}} \\ &= \prod_{i=1}^{m}\prod_{j=1}^{m'}(a_i - b_j)^{n_{a_i}n_{b_j}} \\ &= s(\mathrm{div}(r)) \end{aligned} \tag{32}$$

where similarly as before, the sign equals 1 since

$$\sum_{i=1}^{m}\sum_{j=1}^{m'} n_{a_i}n_{b_j} = \left(\sum_{i=1}^{m} n_{a_i}\right)\left(\sum_{j=1}^{m'} n_{b_j}\right) = n_\infty\cdot 0 = 0$$

and in addition

$$\prod_{j=1}^{m'}(\infty - b_j)^{n_\infty n_{b_j}} = 1,$$

because

$$\sum_{j=1}^{m'} n_\infty n_{b_j} = n_\infty\sum_{j=1}^{m'} n_{b_j} = n_\infty\cdot 0 = 0$$

and by our convention, for every pair of **monic** polynomials $f(x), g(x)$ of the same degree, $\frac{f(\infty)}{g(\infty)} = 1$ (here, the polynomial $f(x)$ is given by

$$f(x) = \prod_{n_{b_j}>0}(x - b_j)^{n_{b_j}}$$

and the polynomial $g(x)$ is given by

$$g(x) = \prod_{n_{b_j}<0}(x - b_j)^{n_{b_j}}$$

) This completes the proof. □

We finish this section by a generalised form of Weil reciprocity. First, consider the following

Definition 6.91. Let $E/\Bbbk$ be an elliptic curve, $f,g \in \Bbbk(E)$ rational functions and $P \in E(\overline{\Bbbk})$. The **tame symbol** of f and g is

$$\langle f,g\rangle_P = (-1)^{\mathrm{ord}_P(f)\,\mathrm{ord}_P(g)} \left(\frac{f^{\mathrm{ord}_P(g)}}{g^{\mathrm{ord}_P(f)}}\right)(P).$$

Theorem 6.92. *Let $E/\Bbbk$ be an elliptic curve and $f,g \in \Bbbk(E)$ be rational functions. Then*

$$\prod_{P\in E(\overline{\Bbbk})} \langle f,g\rangle_P = 1.$$

Remark 6.93. Note that $\langle f,g\rangle_P = 1$ if P is not a zero nor a pole of both f and g. T Thus, for fixed rational functions $f,g \in \Bbbk(E)$, in case we know f and g do not admit zeros or pole over $\overline{\Bbbk}$ that did not exist in $\Bbbk$, Theorem 6.92 remains valid with the product on the RHS taken over $E(\Bbbk)$.

6.8 Torsion points

Definition 6.94. Let $E : y^2 = x^3 + Ax + B$ be an elliptic curve defined over $\Bbbk$. The n**-torision** points of E are

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

If $\Bbbk \subseteq \Bbbk'$ is a subfield inclusion, we also define

$$E(\Bbbk')[n] = \{P \in E(\Bbbk') \mid nP = \mathcal{O}\} \subseteq E(\overline{\Bbbk})$$

that by observation 6.15 gives an inclusion

$$E(\Bbbk)[n] \subseteq E(\Bbbk')[n].$$

The set $E(\Bbbk)[n]$ has an evident abelian group structure since for $P,Q \in E(\Bbbk)[n]$, $n(P+Q) = nP + nQ$.

Example 6.95. It is easy to determine $E[2]$: over $\overline{\Bbbk}$ we can write E as follows, with distinct α_i:

$$y^2 = (x-\alpha_1)(x-\alpha_2)(x-\alpha_3)$$

A point $P \in E(\overline{\Bbbk})$ satisfies $2P = \mathcal{O}$ iff the tangent line to P is vertical which means that $y = 0$. It follows that $E[2] = \{\mathcal{O}, (\alpha_1,0), (\alpha_2,0), (\alpha_3,0)\}$. As an abelian group, $E[2] \cong \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

Let us examine $E(\overline{\Bbbk})[3]$. $3P = \mathcal{O} \iff 2P = -P$ and this means that the x-coordinate of $2P$ equals to that of P and the y-coordinates of $2P$ and P differ by a sign. In equations:
$m^2 - 2x = x$, where $m = \frac{3x^2+A}{2y}$ (see Equation 16). Substituting that in the equation of E we get

$$\begin{aligned}\frac{(3x^2+A)^2}{4m^2} &= x^3 + Ax + B \iff \\ (3x^2+A)^2 &= 12x(x^3+Ax+B) \iff \\ 3x^4 + 6Ax^2 &+ 12Bx - A^2 = 0.\end{aligned} \tag{33}$$

The discriminant of the resulting polynomial is $-6912(4A^3+27B^2)^2$ (as can be verified by the formula in Example 6.9) which is non-zero since E is smooth. Thus, this polynomial has no multiple roots. There are 4 different roots, ie. values of x, over $\overline{\Bbbk}$ and each of them yields 2 values of y. Together with $\mathcal{O}$ we get $|E[3]| = 9$ where each non-zero element has order 3. It follows that

$$E(\overline{\Bbbk})[3] \cong \mathbb{Z}_3 \times \mathbb{Z}_3.$$

The general situation is given by the following

Theorem 6.96. *Let E be an elliptic curve defined over $\mathbb{F}_q$ and let n be a positive integer. If $p \nmid n$,*

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

In other words, there exists k such that for any $k < N$:

$$E(\mathbb{F}_{p^N})[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n.$$

Proof. We defer the proof to Section 6.13. □

Corollary 6.97. *Let E be an elliptic curve over $\mathbb{F}_{p^k}$. Then*

$$E(\mathbb{F}_{p^k}) \cong \begin{cases} \mathbb{Z}_n \\ \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} & \textit{where } n_1 \mid n_2 \end{cases}$$

Proof. By the structure theorem of finite abelian groups 4.83,

$$E(\mathbb{F}_{p^k}) \cong \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r}$$

where for every $1 \le i < r$, $n_i \mid n_{i+1}$ and r is the "rank". Each component $\mathbb{Z}_{n_i}$ contains n_1 elements of order dividing n_1 so that $E(\mathbb{F}_{p^k})$ contains n_1^r elements of order dividing n_1. However,

$$E(\mathbb{F}_{p^k})[n_1] \subseteq E[n_1] \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_1}$$

so $E(\mathbb{F}_{p^k})$ has at most n_1^2 elements of order dividing n_1. It follows that $r \le 2$. □

As an aside let us state the case where $p \mid n$:

Theorem 6.98. *Let E be an elliptic curve defined over $\mathbb{F}_p$ and let n be a positive integer. If $p \mid n$ where $n = p^r n'$ with $p \nmid n'$ then there exists k such that for any $k < N$:*

$$E(\mathbb{F}_{p^N})[n] = \mathbb{Z}_{n'} \oplus \mathbb{Z}_{n'} \;\; \textit{or} \;\; \mathbb{Z}_n \oplus \mathbb{Z}_{n'} \tag{34}$$

Observation 6.99. For an elliptic curve $E = E_{A,B}$, the group $E(\overline{\mathbb{F}}_p)$ is always infinite: given any point $y_0 \in \overline{\mathbb{F}}_p$, the polynomial $x^3 + Ax + B - y_0$ must have a root x_0 in $\overline{\mathbb{F}}_p$ so that $(x_0, y_0) \in E(\overline{\mathbb{F}_p})$.

6.9 Weil pairing and its properties

Recall our setting: E is an elliptic curve defined over $\mathbb{F}_p$, $1 \le n$ an integer with $p = \operatorname{char} \Bbbk \nmid n$ and $\mathbb{F}_p \subseteq \Bbbk$ a field such that

$$E[n] \coloneqq E(\Bbbk)[n] \cong \mathbb{Z}_n \oplus \mathbb{Z}_n.$$

Let us start with a construction of Weil pairing.

Construction 6.100. Let $Q \in E[n]$ and $f_Q \in \Bbbk(E)$ a rational function such that

$$\operatorname{div}(f_Q) = n[Q] - n[\mathcal{O}].$$

By Proposition 6.56, the map $[n] : E(\overline{\Bbbk}) \longrightarrow E(\overline{\Bbbk})$ given by $P \mapsto nP$ is surjective so let $Q' \in E(\Bbbk)[n^2]$ be such that $nQ' = Q$.

Consider the divisor

$$\sum_{R\in E[n]} ([Q'+R]-[R]).$$

Since $|E[n]| = n^2$,

$$\sum_{R\in E[n]} (Q'+R-R) = n^2 Q' = \mathcal{O}$$

so that there exists a rational function $g = g_Q \in \Bbbk(E)$ such that

$$\operatorname{div}(g_Q) = \sum_{R\in E[n]} ([Q'+R]-[R])$$

Note that g_Q does not depend on the choice of Q': if $Q'' \in E[n^2]$ is such that $nQ'' = Q$, then $Q' - Q'' \in E[n]$ so that

$$\sum_{R\in E[n]} ([Q''+R]-[R])$$

is unchanged.

Consider the rational function $f_Q \circ [n] \in \Bbbk(E)$. The points $R \in E[n]$ are poles of $f_Q \circ [n]$ of order n each. The points $X = Q' + R$ for $R \in E[n]$ are those points X for which $nX = Q$, hence the zeros of $f_Q \circ [n]$, each with order n as before. It follows that

$$\operatorname{div}(f_Q \circ [n]) = n\left(\sum_{R\in E[n]} [Q'+R]\right) - n\left(\sum_{R\in E[n]} [R]\right) = \operatorname{div}(g_Q^n).$$

Thus, $f_Q \circ [n]$ is a constant multiple of g_Q^n and wlog, we may choose f_Q such that $f_Q \circ [n] = g_Q^n$.

Let $P \in E[n]$ and $S \in E(\Bbbk)$. Then

$$g_Q(S+P)^n = f_Q\left(n(S+P)\right) = f_Q(nS) = g_Q(S)^n$$

so that

$$\frac{g_Q(S+P)^n}{g_Q(S)^n} = 1.$$

We define the (abstract) **Weil pairing** to be

$$e_n(P,Q) = \frac{g_Q(S+P)}{g_Q(S)}$$

and thus get a function

$$e_n : E(\Bbbk)[n] \times E(\Bbbk)[n] \longrightarrow \Bbbk$$

where for every $P, Q \in E[n]$, $e_n(P,Q) \in \mu_n(\overline{\mathbb{F}_p})$.

Before going to the main proof of this section, we need an auxiliary result:

Lemma 6.101. *Let $E/\Bbbk$ be an elliptic curve, and $1 \le n$ such that $p = \operatorname{char}(\Bbbk) \nmid n$ and $E(\Bbbk)[n] \cong \mathbb{Z}_n \oplus \mathbb{Z}_n$. Suppose $g \in \Bbbk(E)$ is a rational function such that*

$$g(S+P) = g(S)$$

for all $S \in E(\Bbbk)$ and $P \in E(\Bbbk)[n]$. Then there is a rational function

$$h \in \Bbbk(E)$$

such that for all $S \in E(\mathbb{k})$,

$$g(S) = h(nS).$$

In other words, if we consider the rational map $[n] : E(\mathbb{k}) \longrightarrow E(\mathbb{k})$ given by $S \mapsto nS$ then

$$g = h \circ [n].$$

Proof. The proof involves Galois Theory and thus omitted from these notes. □

We are ready to for the main result of this section.

Theorem 6.102. *The Weil pairing of Construction 6.100 has the following properties:*

1. *$\forall P, Q \in E[n]$, $e_n(P,Q)^n = 1$ ie $e_n(P,Q) \in \mu_n(\overline{\mathbb{F}_p})$.*
2. *$\forall P, Q \in E[n]$, $e_n(P,Q)$ is independent of the choice of function $g = g_Q$ and the point S.*
3. *e_n is **bilinear** in each variable, ie for all $P_1, P_2, Q \in E(\mathbb{k})[n]$*

   $$e_n(P_1 + P_2, Q) = e_n(P_1, Q)e_n(P_2, Q)$$

   and

   $$e_n(Q, P_1 + P_2) = e_n(Q, P_1)e_n(Q, P_2).$$

4. *The Weil pairing is **skew-symmetric** (or: alternating) ie.*

   $$e_n(Q,Q) = 1 \ \forall Q \in E(\mathbb{K})[n]$$

   and

   $$e_n(P,Q) = e_n(Q,P)^{-1} \ \forall P, Q \in E(\mathbb{K})[n].$$

5. *e_n is **non-degenerate** in each variable, ie if $e_n(P,Q) = 1$ for all $Q \in E(\mathbb{k})[n]$ then $P = \mathcal{O}$ and if $e_n(P,Q) = 1$ for all $P \in E(\mathbb{k})[n]$ then $Q = \mathcal{O}$.*
6. *The set of maps $\{e_n\}$ is **compatible** in that for any m, n with $p \nmid n, m$, for any $P \in E[mn]$ and for any $Q \in E[n] \subseteq E[mn]$,*

   $$e_{mn}(P,Q) = e_n(mP, Q).$$

7. *Let $\sigma : \overline{\mathbb{k}} \longrightarrow \overline{\mathbb{k}}$ be an automorphism that fixes the elements in $\mathbb{k}$. Then*

   $$e_n(\sigma P, \sigma Q) = \sigma e_n(P,Q).$$

8. *Let $\alpha : E \longrightarrow E$ be a seperable endomorphism or the Frobenius endomorphism Φ_q. Then for all $P, Q \in E[n]$,*

   $$e_n(\alpha P, \alpha Q) = e_n(P,Q)^{\deg \alpha}.$$

   (In fact, this equality holds for non-seperable endomorphisms as well).

Proof.

1. This was proved in Construction 6.100.

2. Fix $P, Q \in E(\Bbbk)[n]$. Then

$$e_n(P,Q) = \frac{g_Q(P+S)}{g_Q(S)}$$

can be viewed as a function of $S \in E(\Bbbk)$ ie as a rational function $E(\Bbbk) \longrightarrow \Bbbk$. Enlarging $\Bbbk$ if necessary, we invoke Proposition 6.56 that says this function is either constant or surjective. However, since we know $e_n(P,Q) \in \mu_n(\overline{\mathbb{F}_p})$, it cannot be surjective hence constant ie does not depend on S.

3. Let us prove linearity in the first variable. Since e_n is independent of the choice of S, we may replace S by $S+P_1$ to get

$$\begin{aligned} e_n(P_1,Q)e_n(P_2,Q) &= \frac{g(P_1+S)}{g(S)}\frac{g(P_2+P_1+S)}{g(P_1+S)} \\ &= \frac{g(P_1+P_2+S)}{g(S)} = e_n(P_1+P_2,Q). \end{aligned} \tag{35}$$

For linearity in the second variable, suppose $Q_1, Q_2, Q_3 \in E[n]$ such that $Q_1 + Q_2 = Q_3$. For $1 \leq i \leq 3$, let f_{Q_i}, g_{Q_i} be the functions used to define $e_n(P,Q_i)$ in Construction 6.100, and let $h \in \Bbbk(E)$ be a rational function such that

$$\operatorname{div}(h) = [Q_3] - [Q_2] - [Q_1] + [\mathcal{O}].$$

Then we have

$$\operatorname{div}\left(\frac{f_{Q_3}}{f_{Q_1}f_{Q_2}}\right) = n\operatorname{div}(h) = \operatorname{div}(h^n).$$

Thus, there exists a constant $c \in \Bbbk$ such that $f_{Q_3} = c f_{Q_1} f_{Q_2} h^n$ and this means that

$$\begin{aligned} g_{Q_3}^n = f_{Q_3} \circ [n] &= (c \cdot f_{Q_1} \cdot f_{Q_2} \cdot h^n) \circ [n] \\ &= c \cdot (f_{Q_1} \circ [n]) \cdot (f_{Q_2} \circ [n])(h \circ [n])^n. \end{aligned} \tag{36}$$

Taking nth root, we get

$$g_{Q_3} = c^{\frac{1}{n}}(g_{Q_1})(g_{Q_2})(h \circ [n]).$$

The definition of e_n now yields

$$\begin{aligned} e_n(P, Q_1+Q_2) &= \frac{g_{Q_3}(P+S)}{g_{Q_3}(S)} \\ &= \frac{g_{Q_1}(P+S)}{g_{Q_1}(S)}\frac{g_{Q_2}(P+S)}{g_{Q_2}(S)}\frac{h(n(P+S))}{h(nS)} \\ &= e_n(P,Q_1)e_n(P,Q_2) \end{aligned} \tag{37}$$

where the last equality follows since $nP = \mathcal{O}$ so that $h(n(P+S)) = h(nS)$.

4. Let $\tau_{jQ} : E(\Bbbk) \longrightarrow E(\Bbbk)$ be the translation by jQ so $f \circ \tau_{jQ}$ is the function $P \mapsto f(P+jQ)$. A direct inspection shows that

$$\operatorname{div}(f_Q \circ \tau_{jQ}) = n[Q - jQ] - n[-jQ].$$

Thus,

$$\operatorname{div}\left(\prod_{j=0}^{n-1} f_Q \circ \tau_{jQ}\right) = \sum_{j=0}^{n-1}(n[(1-j)Q] - n[-jQ]) = 0,$$

so that $\prod_{j=0}^{n-1} f_Q \circ \tau_{jQ}$ is constant. We now have

$$\begin{aligned} \left(\prod_{j=0}^{n-1} g \circ \tau_{jQ'}\right)^n &= \prod_{j=0}^{n-1} f_Q \circ [n] \circ \tau_{jQ'} \\ &= \prod_{j=0}^{n-1} f_Q \circ \tau_{jQ} \circ [n] \quad (\textit{since } nQ' = Q) \\ &= \left(\prod_{j=0}^{n-1} f_Q \circ \tau_{jQ}\right) \circ [n]. \end{aligned} \tag{38}$$

so that $\left(\prod_{j=0}^{n-1} g \circ \tau_{jQ'}\right)^n$ is constant. We now invoke Proposition 6.56 to deduce that there is some field extension K of $\Bbbk$ for which $\prod_{j=0}^{n-1} g \circ \tau_{jQ'}$ is constant, hence must be constant already over $\Bbbk$. Thus, $\prod_{j=0}^{n-1} g \circ \tau_{jQ'}$ has the same value at S and $S + Q'$ so that

$$\prod_{j=0}^{n-1} g(S + Q' + jQ') = \prod_{j=0}^{n-1} g(S + jQ').$$

Canceling all common terms (we assume S is chosen such that all terms are finite and non-zero), we get

$$g(S + nQ') = g(S).$$

Since $nQ' = Q$ we get

$$e_n(Q, Q) = \frac{g(S + Q)}{g(S)} = 1.$$

As for the second part, bilinearity yields

$$1 = e_n(P + Q, P + Q) = e_n(P, P)e_n(P, Q)e_n(Q, P)e_n(Q, Q)$$

and since we have just showed that $e_n(P, P) = e_n(Q, Q) = 1$ we get

$$e_n(P, Q) = e_n(Q, P)^{-1}.$$

5. Suppose $Q \in E[n]$ is such that $e_n(P, Q) = 1$ for all $P \in E[n]$. By Theorem 6.105, this means that $e_n(P, Q) = 1$ so that $g(S + P) = g(S)$ for all $P \in E[n]$ and $S \in E(\Bbbk)$. By Lemma 6.101, there is a rational function $h \in \Bbbk(E)$ such that $g = h \circ [n]$. Then,

$$(h \circ [n])^n = g^n = f \circ [n].$$

Note that $(h \circ [n])^n = h^n \circ [n]$ so that $h^n \circ [n] = f \circ [n]$. By Proposition 6.56, $[n] : E(\Bbbk) \longrightarrow E(\Bbbk)$ is surjective, so by Lemma 2.17, $h^n = f$. Thus, we have

$$n \operatorname{div}(h) = \operatorname{div}(f) = n[Q] - n[\mathcal{O}],$$

so that

$$\operatorname{div}(h) = [Q] - [\mathcal{O}].$$

Since h is a rational function, $Q = \mathcal{O}$, and this proves half of Theorem 6.102 (5). The second half follows from the first half in conjunction with (4).

6. The proof requires a use of another definition of the Weil pairing. We thus deffer it to the end of section 6.10.

7. Apply σ on all coefficients in the construction of e_n and denote by f_Q^σ, g_Q^σ the resulting rational maps. Then

$$\operatorname{div}(f_Q^\sigma) = n[\sigma Q] - n[\mathcal{O}]$$

and similarly for g_Q^σ. Therefore,

$$\sigma(e_n(P,Q)) = \sigma\left(\frac{g_Q(P+S)}{g_Q(S)}\right) = \frac{g^\sigma(\sigma P + \sigma S)}{g^\sigma(\sigma P)} = e_n(\sigma P, \sigma Q).$$

8. Let $\ker\alpha = \{T_1, .., T_k\}$. Since α is seperable, Proposition 6.71 asserts that, $k = \deg\alpha$. As before, let

$$\operatorname{div}(f_Q) = n[Q] - n[\mathcal{O}], \;\; \operatorname{div}(f_{\alpha(Q)}) = n[\alpha(Q)] - n[\mathcal{O}],$$

$$g_Q^n = f_Q \circ [n], \;\; g_{\alpha(Q)}^n = f_{\alpha(Q)} \circ [n]$$

and $\tau_T : E \longrightarrow E$ denote adding T. Then we have,

$$\operatorname{div}(f_Q \circ \tau_{-T_i}) = n[Q + T_i] - n[T_i].$$

Therefore,

$$\begin{aligned} \operatorname{div}(f_{\alpha(Q)} \circ \alpha) &= n \sum_{\alpha(Q'') = \alpha(Q)} [Q''] - n \sum_{\alpha(T) = \mathcal{O}} [T] \\ &= n \sum_{i=1}^{k} ([Q + T_i] - [T_i]) \\ &= \operatorname{div}(\prod_i (f_Q \circ \tau_{-T_i})). \end{aligned} \tag{39}$$

For each $1 \le i \le k$, choose T_i' such that $nT_i' = T_i$ (recall that $[n]$ is surjective). Then

$$(\star)\; g_Q(S - T_i')^n = f_Q(nS - T_i).$$

Thus,

$$\begin{aligned} \operatorname{div}\left(\prod_i (g_Q \circ \tau_{-T_i'})^n\right) &= \operatorname{div}(\prod_i f_Q \circ \tau_{-T_i} \circ [n]) \\ &= \operatorname{div}(f_{\alpha(Q)} \circ \alpha \circ [n]) \\ &= \operatorname{div}(f_{\alpha(Q)} \circ [n] \circ \alpha) \\ &= \operatorname{div}\left((g_{\alpha(Q)} \circ \alpha)^n\right). \end{aligned} \tag{40}$$

The first equality follows from $\star$. The second equality follows from composing $[n]$ on the right of both sides of Equation 39 and the fact that composing a rational map on the right of a product of maps is the same as composing it on the right of each component of the product. The third equality follows from the fact that $[n]$ commutes with rational maps. The forth equality follows from the fact that $f_{\alpha(Q)} \circ [n] = g_{\alpha(Q)}^n$ and the fact that composing a rational map (α) on the right of a product of maps is the same as composing it on the right of each component of the product.

It follows that $\prod_i g_Q \circ \tau_{-Q'_i}$ and $g_{\alpha(Q)} \circ \alpha$ have the same divisor hence differ by a constant C.

The definition of e_n yields

$$\begin{aligned} e_n(\alpha(P), \alpha(Q)) &= \frac{g_{\alpha(Q)}(\alpha(S+P))}{g_{\alpha(Q)(\alpha(S))}} \\ &= \prod_i \frac{g_Q(S+P-T'_i)}{g_Q(S-T'_i)} \\ &= \prod_i e_n(P,Q) \\ &= e_n(P,Q)^k \\ &= e_n(P,Q)^{\deg(\alpha)} \end{aligned} \tag{41}$$

where the first equality is by definition (choosing $\alpha(S)$ instead of S), the second from the fact that $\prod_i g_Q \circ \tau_{-Q'_i}$ and $g_{\alpha(Q)} \circ \alpha$ have the same divisor (the constant C cancels out) and the last follows from definition by choosing $S - T'_i$ instead of S. Lastly, if $\alpha = \Phi_q$ is the Frobenius endomorphism, then by (7) we get

$$e_n(\Phi_q P, \Phi_q Q) = \Phi_q(e_n(P,Q)) = e_n(P,Q)^q.$$

□

Corollary 6.103. *In the setting of Construction 6.100, $\mu_n(\overline{\mathbb{F}_p}) \subseteq \Bbbk$ so that the Weil pairing can be viewed as a map*

$$e_n : E[n] \times E[n] \longrightarrow \mu_n(\overline{\mathbb{F}_p}).$$

Proof. Let $P_0, Q_0 \in E[n]$ be such that $\langle (P_0, Q_0) \rangle = E[n] = \mathbb{Z}_n \times \mathbb{Z}_n$. We claim that $e_n(P_0, Q_0)$ is a generator of $\mu_n(\overline{\mathbb{F}_p})$ which means that e_n is surjective onto $\mu_n(\overline{\mathbb{F}_p})$. Since by definition $e_n(P,Q) \in \Bbbk$ for all P, Q, we get that $\mu_n(\overline{\mathbb{F}_p}) \subseteq \Bbbk$.

Suppose by contradiction that $e_n(P_0, Q_0)$ is not a generator of μ_n. Then there exists $m < n$ such that $e_n(P_0, Q_0)^m = 1$. By bilinearity we get $e_n(mP_0, Q_0) = 1$. If $X \in E[n]$ then there are $a, b \in \mathbb{Z}$ such that $X = aP_0 + bQ_0$. Then

$$e_n(mP_0, X) = e_n(mP_0, P_0)^a e_n(mP_0, Q_0)^b = e_n(P_0, P_0)^{ma} e_n(P_0, Q_0)^{mb} = 1$$

and thus by non-degeneracy, $mP_0 = \mathcal{O}$ – contradiction. □

6.10 Equivalence of definitions of Weil pairing

The Weil pairing admits a few equivalent definitions in the literature, each useful for different purposes. Our goal in this section is to two other definitions of the Weil pairing and prove their equivalence.

Definition 6.104. Let $E/\Bbbk$ be an elliptic curve. Suppose $P, Q \in E(\Bbbk)[n]$ such that $P \neq Q$ and $P, Q \neq \mathcal{O}$. Let $f_P, f_Q \in \Bbbk(E)$ be monic rational functions such that

$$\operatorname{div}(f_P) = n[P] - n[\mathcal{O}], \quad \operatorname{div}(f_Q) = n[Q] - n[\mathcal{O}].$$

The **efficient Weil pairing** of P, Q is defined to be

$$e_n^{\text{eff}}(P,Q) = (-1)^n \frac{f_P(Q)}{f_Q(P)}.$$

If $P = Q$ or either $P = \mathcal{O}$ or $Q = \mathcal{O}$, we set

$$e_n^{\text{eff}}(P,Q) = 1.$$

The following Theorem is stated here for convenience, but we deffer its proof to a later part in this section.

Theorem 6.105. *Under the setup of Construction 6.100, for every $P, Q \in E[n]$, $e_n(P,Q) = e_n^{\text{eff}}(P,Q)$.*

Remark 6.106. In Definition 6.104, since f_P, f_Q are monic and have a pole of order n at $\mathcal{O}$, it follows that $\frac{f_Q(\mathcal{O})}{f_P(\mathcal{O})} = 1$ so that

$$e_n^{\text{eff}}(P,Q) = (-1)^n \frac{f_P(Q)}{f_Q(P)} \cdot \frac{f_Q(\mathcal{O})}{f_P(\mathcal{O})}.$$

There is yet another definition of the Weil pairing which is convenient for "theoretical" purposes as it is expressed in terms of general divisors. First, let us introduce a bit of terminology: we say that a divisor D on an elliptic curve $E/\Bbbk$ is **principal** if there is a rational function $f \in \Bbbk(E)$ such that $\operatorname{div}(f) = D$. Recall that according to Theorem 6.49, a divisor D is principal if and only if:

$$\operatorname{sum} D = \mathcal{O}, \quad \deg D = 0.$$

Next, let us define

Definition 6.107. Let $E/\Bbbk$ be an elliptic curve and D, D' divisors on E. We say that D is equivalent to D' and write $D \sim D'$ if $D - D'$ is principal.

Exercise 6.108. Prove that equivalence of divisors is an equivalence relation on the set of all divisors.

We are ready to define the generic version of the Weil pairing:

Definition 6.109. Let $E/\Bbbk$ be an elliptic curve. For a degree zero divisor D, such that $nD \sim 0$ we let f_D to denote a monic rational function in $\Bbbk(E)$ such that

$$\operatorname{div}(f_D) = nD.$$

(e.g. $f_{[P]-[\mathcal{O}]} = f_P$ in the notation of Construction 6.100).

Let $P, Q \in E(\Bbbk)[n]$. Choose divisors D_P, D_Q with disjoint support such that

$$D_P \sim [P] - [\mathcal{O}], \quad D_Q \sim [Q] - [\mathcal{O}].$$

The **generic Weil pairing** is defined to be

$$e_n^{\text{gen}}(P,Q) = \frac{f_{D_P}(D_Q)}{f_{D_Q}(D_P)}.$$

Our first order of business is to verify Definition 6.109 is well-defined.

Lemma 6.110. *Definition 6.109 does not depend on the choice of divisors D_P, D_Q or rational functions f_{D_P}, f_{D_Q}.*

Proof. The choice of rational functions with a prescribed divisor is unique up to a constant and the requirement they are monic means they are unique. Let us show independence of the choice of D_Q; independence of the choice of D_P is proven analogously. Suppose D'_Q is another divisor such that $D'_Q \sim [Q] - [\mathcal{O}]$. Then $D'_Q = D_Q + \operatorname{div}(h)$ for some $h \in \Bbbk(E)$ with support disjoint from D_P. Then $f_{D'_Q} = f_{D_Q} h^n$ and thus

$$\begin{aligned}\frac{f_{D_P}(D'_Q)}{f_{D'_Q}(D_P)} &= \frac{f_{D_P}(D_Q) f_{D_P}(\operatorname{div} h)}{f_{D_Q}(D_P) h(D_P)^n} \\ &= \frac{f_{D_P}(D_Q) f_{D_P}(\operatorname{div} h)}{f_{D_Q}(D_P) h(\operatorname{div}(f_{D_P}))} \\ &= \frac{f_{D_P}(D_Q)}{f_{D_Q}(D_P)}.\end{aligned} \tag{42}$$

□

We now wish to show that the Definitions of generic Weil pairing and efficient Weil pairing are equivalent and thus, by Remark 6.106 are equivalent to the (computational) definition of the Weil pairing we originally defined.

Proposition 6.111. *Let $E/\Bbbk$ be an elliptic curve and $1 \leq n$ such that $p = \operatorname{char}\Bbbk \nmid n$ and $E(\Bbbk)[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$. For any $P, Q \in E(\Bbbk)[n]$,*

$$e_n^{\mathrm{eff}}(P,Q) = e_n^{\mathrm{gen}}(P,Q).$$

Proof. Assume first that $P \neq Q$. Let $D_P = [P] - [\mathcal{O}]$ so that $f_{D_P} = f_P$. Let $S \in E(\Bbbk)$ be such that $D_Q := [Q+S] - [S]$ has disjoint support with D_P (i.e. $S \notin \{\mathcal{O}, P, -Q, P-Q\}$).

Then, $D_Q = [Q] - [\mathcal{O}] + \operatorname{div}(h)$ for some monic $h \in \Bbbk(E)$ such that

$$\operatorname{div} h = [Q+S] - [S] - [Q] + [\mathcal{O}]$$

and thus

$$f_{D_Q} = f_Q h^n.$$

By generalised Weil reciprocity 6.92,

$$\begin{aligned} prod_{A\in E(\bar{\Bbbk})}\langle f_P, h\rangle_A &= prod_{A\in E(\Bbbk)}\langle f_P, h\rangle_A \\ &= (-1)^n \frac{f_P(Q+S)f_P(\mathcal{O})}{f_P(S)f_P(Q)h^n(P)h^{-n}(\mathcal{O})} \\ &= \frac{f_P(Q+S)}{f_P(S)f_P(Q)h^n(P)} \cdot (-1)^n (f_P h^n)(\mathcal{O}) \\ &= \frac{f_P(Q+S)}{f_P(S)f_P(Q)h^n(P)} \cdot (-1)^n. \end{aligned} \tag{43}$$

Thus,

$$\begin{aligned} e_n^{\mathrm{gen}}(P,Q) = \frac{f_{D_P}(D_Q)}{f_{D_Q}(D_P)} &= \frac{(f_Q h^n)(\mathcal{O})}{(f_Q h^n)(P)} \cdot \frac{f_P(Q+S)}{f_P(S)} \\ &= \frac{f_P(Q)}{f_Q(P)} \cdot \frac{f_P(Q+S)}{f_P(Q)h^n(P)f_P(S)} \\ &= (-1)^n \frac{f_P(Q)}{f_Q(P)} = e_n^{\mathrm{eff}}(P,Q) \end{aligned} \tag{44}$$

If $P = Q$, a similar calculation shows that

$$e_n^{\mathrm{gen}}(P,Q) = 1.$$

□

In light of Proposition 6.111, the proof of Theorem 6.10 can be reduced to the following Theorem, whose proof will occupy the remain of this section.

Theorem 6.112. *Let $E/\Bbbk$ be an elliptic curve and let $1 \leq n$ be an integer such that $p = \operatorname{char}\Bbbk \nmid n$ and $E(\Bbbk)[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$. Then for any $P, Q \in E(\Bbbk)[n]$,*

$$e_n(P,Q) = e_n^{\mathrm{gen}}(P,Q).$$

Let $V, W \in E(\Bbbk)[n^2]$ and let f_{nV}, g_{nV} be as in Construction 6.100 ie such that

$$\operatorname{div}(f_{nV}) = n[nV] - n[\mathcal{O}],$$

$$g_{nV}^n = f_{nV} \circ [n].$$

Define

$$c(nV, nW) = \frac{f_{nV+nW}(X)}{f_{nV}(X) f_{nW}(X - nV)}$$

and

$$d(V, W) = \frac{g_{nV+nW}(X)}{g_{nV}(X) g_{nW}(X - V)}.$$

where $X \in E(\Bbbk)$. The notation on the left does not include X because

Lemma 6.113. *$c(nV, nW)$ and $d(V, W)$ are constants and*

$$d(V, W)^n = c(nV, nW).$$

Proof. Using $\operatorname{div}(f_{nV}) = n[nV] - n[\mathcal{O}]$ we get that $\operatorname{div} c(nV, nW) = 0$ and thus $c(nV, nW)$ is constant. Since $g_{nV}^n = f_{nV} \circ [n]$ we get that

$$d(V, W)^n = \frac{f_{nV+nW}(X)}{f_{nV}(X) f_{nW}(nX - nV)} = c(nV, nW)$$

where the last equality holds since $c(nV, nW)$ does not depend on X. Thus, $d(V, W)$ is constant as well. □

The next few Lemmas tie c and d to e_n.

Lemma 6.114. *Let $U, V, W \in E[n^2]$. Then*

$$d(V, W + nU) = d(V, W)$$

and

$$d(V + nU, W) = d(V, W) e_n(nU, nW).$$

Proof. Since $n(W + nU) = nW$, we have $g_{nV+n(W+nU)} = g_{nV+nW}$. Thus,

$$d(V, W + nU) = \frac{g_{nV+nW}(X)}{g_{nV}(X) g_{nW}(X - V)} = d(V, W).$$

Similarly,

$$\begin{aligned} d(V + nU, W) &= \frac{g_{nV+nW}(X)}{g_{nV}(X) g_{nW}(X - V - nU)} \\ &= \frac{g_{nV+nW}(X)}{g_{nV}(X) g_{nW}(X - V)} \frac{g_{nW}(X - V)}{g_{nW}(X - V - nU)} \\ &= d(V, W) \frac{g_{nW}\left((X - V - nU) + nU\right)}{g_{nW}(X - V - nU)} \\ &= d(V, W) e_n(nU, nW) \end{aligned} \tag{45}$$

where the last equality follows from taking $S = X - V - nU$ in the definition of e_n. □

Lemma 6.115. *For $U, V, W \in E[n^2]$,*

$$\frac{d(U,V)}{d(V,U)} = \frac{d(V,W)d(U+W,V)}{d(V,U+W)d(W,V)}.$$

Proof. From the definition of d we get:

$$\begin{aligned} g_{nU+(nV+nW)}(X) &= d(U,V+W)g_{nU}(X)g_{nV+nW}(X-U) \\ &= d(U,V+W)g_{nU}(X)d(V,W)g_{nV}(X-U)g_{nW}(X-U-V). \end{aligned} \tag{46}$$

Similarly,

$$\begin{aligned} g_{(nU+nV)+nW}(X) &= d(U+V,W)g_{nU+nV}(X)g_{nW}(X-U-V) \\ &= d(U+V,W)d(U,V)g_{nU}(X)g_{nV}(X-U)g_{nW}(X-U-V). \end{aligned} \tag{47}$$

Since

$$g_{nU+(nV+nW)} = g_{(nU+nV)+nW},$$

we can cancel common terms and obtain

$$d(U,V+W)d(V,W) = d(U+V,W)d(U,V). \tag{48}$$

Interchange the roles of U, V in Equation 48 and divide to obtain

$$\frac{d(U,V)}{d(V,U)} = \frac{d(U,V+W)d(V,W)}{d(V,U+W)d(U,W)}. \tag{49}$$

Now, swap the roles of V, W in Equation 48, solve for $d(U,W)$ and substitute in 49 to obtain the result. □

Lemma 6.116. *Let $P, Q \in E[n]$. Then*

$$e_n(P,Q) = \frac{c(P,Q)}{c(Q,P)}.$$

Proof. By Proposition 6.56, the map $[n] : E(\Bbbk) \longrightarrow E(\Bbbk)$ is surjective so we may choose $U, V \in E[n^2]$ such that $nU = P$ and $nV = Q$. The left-hand side of the formula of Lemma 6.115 does not depend on W so we may substitute $W = jU$ for $0 \le j < n$ and multiply the results to obtain:

$$\frac{c(P,Q)}{c(Q,P)} = \left(\frac{d(U,V)}{d(V,U)}\right)^n = \prod_{j=0}^{n-1} \frac{d(V,jU)d(U+jU,V)}{d(V,U+jU)d(jU,V)}. \tag{50}$$

Most terms on the right-hand side of Equation 50 cancel except those for $j = 0$ and $j = n-1$ so that

$$\frac{c(P,Q)}{c(Q,P)} = \frac{d(V,\mathcal{O})d(nU,V)}{d(V,nU)d(\mathcal{O},V)}.$$

In the first equation of Lemma 6.114 we substitute $W = \mathcal{O}$ to obtain $d(V,nU) = d(V,\mathcal{O})$. In the second equation of Lemma 6.114 we substitute $V = \mathcal{O}$ and $W = V$ to get

$$d(nU,V) = d(\mathcal{O},V)e_n(nU,nV) = d(\mathcal{O},V)e_n(P,Q),$$

and the result follows. □

We are ready for the

Proof of Theorem 6.112. Lemma 6.116 and the definition of c shows that

$$e(P,Q) = \frac{c(P,Q)}{c(Q,P)} = \frac{f_Q(X)f_P(X-Q)}{f_P(X)f_Q(X-P)},$$

which is independent of X.

Let $D_P = [P] - [\mathcal{O}]$ and $D_Q = [S] - [S-Q]$ where S is chosen such that $\operatorname{supp}(D_P) \cap \operatorname{supp}(D_Q) = \varnothing$ ie $S \notin \{P, \mathcal{O}, Q, P+Q\}$.

Let $F_P(X) = f_P(X)$ and $F_Q(X) = \frac{1}{f_Q(S-X)}$.

Then

$$\operatorname{div}(F_P) = n[P] - n[\mathcal{O}] = nD_P$$

and

$$\operatorname{div}(F_Q) = n[S] - n[S-Q] = nD_Q.$$

We therefore have

$$e_n(P,Q) = \frac{F_Q(D_P)}{F_P(D_Q)} = e_n^{\mathrm{gen}}(P,Q)$$

and this completes the proof.

□

Corollary 6.117. *In the setup of Construction 6.100, for any $P,Q \in E[n]$,*

$$e_n(P,Q) = e_n^{\mathrm{eff}}(P,Q) = e_n^{\mathrm{gen}}(P,Q).$$

Let us finish this section with the proof of the last property of Weil pairing.

Proof of Theorem 6.102(6) . In light of Corollary 6.117, let us write e_n for e_n^{gen} in this proof. We have $mnP = \mathcal{O}$ and $nQ = \mathcal{O}$. Let

$$\begin{aligned} f_1: &\ \operatorname{div}(f_1) = mn([P] - [\mathcal{O}]), \\ f_2: &\ \operatorname{div}(f_2) = n([Q+T] - [T]), \\ f_3: &\ \operatorname{div}(f_3) = n([mP] - [\mathcal{O}]), \end{aligned} \tag{51}$$

where $T \notin \{P, -Q, P-Q, \mathcal{O}, mP, mP-Q\}$. Then, for $i \neq j$, $\operatorname{div}(f_i), \operatorname{div}(f_j)$ have disjoint support. Let $D_P = [P] - [\mathcal{O}]$ and $D_Q = [Q+T] - [T]$. Then D_P, D_Q have disjoint support and

$$\operatorname{div} f_1 = (mn)D_P$$

and

$$\operatorname{div}(f_2^m) = (mn)D_Q.$$

Thus,

$$e_{mn}(P,Q) = \frac{f_1([Q+T] - [T])}{f_2^m([P] - [\mathcal{O}])}.$$

Similarly, let $D_P = [mP] - [\mathcal{O}]$ and $D_Q = [Q+T] - [T]$. Then, by the choice of T, D_P, D_Q have disjoint support. Moreover

$$\operatorname{div} f_3 = nD_P$$

and

$$\operatorname{div} f_2 = nD_Q.$$

Thus,

$$e_n(mP, Q) = \frac{f_3([Q+T]-[T])}{f_2([mP]-[\mathcal{O}])}.$$

Observe that

$$\begin{aligned} \operatorname{div}(f_3) &= n([mP]-[\mathcal{O}]) \\ &= n([mP]+(m-1)[\mathcal{O}]-m[P]) + mn([P]-[\mathcal{O}]) \\ &= \operatorname{div}(f_4^n \cdot f_1) \end{aligned} \tag{52}$$

where $\operatorname{div}(f_4) = [mP] + (m-1)[\mathcal{O}] - m[P]$. Thus,

$$\begin{aligned} e_{mn}(P,Q) &= \frac{f_3 f_4^{-n}([Q+T]-[T])}{f_2^m([P]-[\mathcal{O}])} \\ &= \frac{f_3([Q+T]-[T])f_4(-\operatorname{div}(f_2))}{f_2^m([P]-[\mathcal{O}])} \\ &= \frac{f_3([Q+T]-[T])}{f_2(\operatorname{div}(f_4)+m([P]-[\mathcal{O}]))} \\ &= \frac{f_3([Q+T]-[T])}{f_2([mP]-[\mathcal{O}])} \\ &= e_n(mP,Q). \end{aligned} \tag{53}$$

□

6.11 Miller's algorithm

In order to use the Weil pairing e_n in practice, we need an efficient algorithm that given a point $P \in E(\Bbbk)[n]$, allows us to evaluate a rational function f_P such that $\operatorname{div}(f_P) \sim n[P] - n[\mathcal{O}]$ at various points. The main algorithm for such a calculation is called **Miller's algorithm** and originally appeared in [Mil].

Henceforth, we work in the setup of the Weil pairing appearing in Construction 6.100: $E/\mathbb{F}_p$ an elliptic curve, $1 \le n$ s.th. $p \nmid n$ and $\mathbb{F}_p \subseteq \Bbbk$ a finite field extension such that

$$E(\Bbbk)[n] = E[n] = \mathbb{Z}_n \times \mathbb{Z}_n.$$

Suppose $U, V \in E(\Bbbk)$.

Let $\mathcal{L}_{U,V}$ be the **(horizontal) line** through U, V. For $U \neq V$ and $U, V \neq \mathcal{O}$,

$$\operatorname{div} \mathcal{L}_{U,V} = [U] + [V] + [-(U+V)] - 3[\mathcal{O}].$$

For $U \neq V$ and $V = \mathcal{O}$,

$$\operatorname{div} \mathcal{L}_{U,V} = [U] + [-U] - 2[U].$$

Let $\mathcal{T}_U$ be the **tangent line** through U. Then

$$\operatorname{div}\mathcal{T}_U = 2[U] + [-2U] - 3[\mathcal{O}].$$

Let $\mathcal{V}_U$ be the **vertical line** through U. Then

$$\operatorname{div}\mathcal{V}_U = [U] + [-U] - 2[\mathcal{O}].$$

Observation 6.118. For any $U \in E(\Bbbk)$, $\mathcal{T}_U = \mathcal{L}_{U,U}$ and $\mathcal{V}_U = \mathcal{L}_{U,-U}$.

Let $P, Q \in E[n]$ and $R, S \in E(\Bbbk)$ such that

$$S \notin \{R, P+R, P+R-Q, R-Q\}.$$

Consider the divisors

$$D_P = [P+R] - [R]$$

and

$$D_Q = [Q+S] - [S]$$

that have disjoint support by the choice of R, S. Clearly, $D_P \sim [P] - [\mathcal{O}]$ and $D_Q \sim [Q] - [\mathcal{O}]$. Let

$$f_P \text{ monic} : \operatorname{div} f_P = n[P+R] - n[R],$$

$$f_Q \text{ monic} : \operatorname{div} f_Q = n[Q+S] - [S]$$

Then $\operatorname{div} f_P = nD_P$ and $\operatorname{div} f_Q = nD_Q$ so by Definition 6.109 and Theorem 6.105 we can write

$$e_n(P,Q) = \frac{f_P(D_Q)}{f_Q(D_P)} = \frac{f_P(Q+S)}{f_P(S)} \Bigg/ \frac{f_Q(P+R)}{f_Q(R)}.$$

Henceforth, we aim to describe a way to evaluate $f_P(X)$ for various $X \in E(\Bbbk)$. One can then use this to evaluate $f_Q(Y)$ for various $Y \in E(\Bbbk)$ in a similar way.

For $k \in \mathbb{Z}$, let

$$f_k \text{ monic} : \operatorname{div} f_k = k[P+R] - k[R] - [kP] + [\mathcal{O}]$$

so that $f_n = f_P$.

We now have the following

Lemma 6.119. *For $a, b \in \mathbb{Z}$,*

$$\operatorname{div}\left(f_a \cdot f_b \cdot \frac{\mathcal{L}_{aP,bP}}{\mathcal{V}_{(a+b)P}}\right) = \operatorname{div}(f_{a+b}).$$

Proof. WLOG, $P \neq \mathcal{O}$. Then

$$\begin{aligned} \text{LHS} &= a[P+R] - a[R] - [aP] + [\mathcal{O}] \\ &\quad + b[P+R] - b[R] - [bP] + [\mathcal{O}] \\ &\quad + [aP] + [bP] + [-(a+b)P] - 3[\mathcal{O}] \\ &\quad - ([(a+b)P] + [-(a+b)P] - 2[\mathcal{O}]) \\ &= (a+b)[P+R] - a[R] - b[R] - [(a+b)P] + [\mathcal{O}] \\ &= (a+b)[P+R] - (a+b)[R] - [(a+b)P] + [\mathcal{O}] \\ &= \text{RHS}. \end{aligned} \tag{54}$$

□

Corollary 6.120.

$$\operatorname{div}\left(f_k^2 \cdot \frac{\mathcal{T}_{kP}}{\mathcal{V}_{2kP}}\right) = \operatorname{div}(f_{2k}).$$

Furthermore,

Lemma 6.121. *We have*

$$\operatorname{div}(f_1) = \operatorname{div}\left(\frac{\mathcal{V}_{P+R}}{\mathcal{L}_{P,R}}\right).$$

Proof.

$$\text{LHS} = [P+R] - [R] - [P] + [\mathcal{O}] \tag{55}$$

whereas

$$\begin{aligned}\text{RHS} = [P+R] + [-(P+R)] - 2[\mathcal{O}] - ([P] + [R] + [-(P+R)] - 3[\mathcal{O}]) \\ = \text{LHS}.\end{aligned} \tag{56}$$

□

To describe Miller's Algorithm, recall that

$$f_P : \operatorname{div}(f_P) = n[P+R] - n[R]$$

and let us write

$$n = \sum_{i=0}^{t} 2^i \cdot n_i$$

where $n_i \in \{0, 1\}$.

Algorithm 2 Miller's algorithm to compute $f_P(Q)$. Output $x = f_P(Q)$.

```
$x_1 := \frac{\mathcal{V}_{P+R}(Q)}{\mathcal{L}_{P,R}(Q)}$.
$x := x_1$.
$Z := P$.
for $i := t-1, ..., 0$ do
    $x := x^2 \cdot \frac{\mathcal{T}_Z(Q)}{\mathcal{V}_{2Z}(Q)}$.
    $Z := 2Z$.
    if $n_i = 1$ then
        $x := x \cdot x_1 \cdot \frac{\mathcal{L}_{Z,P}(Q)}{\mathcal{V}_{Z+P}(Q)}$.
        $Z := Z + P$.
    end if
end for
```

Proposition 6.122. *Miller's algorithm has computational complexity of $\mathcal{O}(\log n)$ points addition. On termination, we have $Z = nP = \mathcal{O}$ and $x = f_P(Q)$.*

Proof. The complexity statement follows from the fact that we run a loop of order $\log n$, in which a constant number of points addition needs to be calculated. The rest of the computation admits closed formulas hence can be done in constant time. It is trivial to check the Algorithm works when $t = 0$, so suppose $t \geq 1$. By Lemma 6.121, the pre-for loop setup is:

$$\begin{aligned} x_1 &= f_1(Q), \\ x &= f_1(Q), \\ Z &= P. \end{aligned} \tag{57}$$

For $i = t - 1$, we have

$$x = f_1^2(Q) \cdot \frac{\mathcal{T}_P(Q)}{\mathcal{V}_{2P}(Q)} = f_2(Q)$$

and we then set $Z = 2P$. If $n_{t-1} = 1$ we get by Step 8 & 9:

$$x = f_1(Q) \cdot f_2(Q) \cdot \frac{\mathcal{L}_{2P,P}(Q)}{\mathcal{V}_{3P}(Q)} = f_3(Q)$$

(where the last equality holds by Lemma 6.119) and $Z = 3P$. The proof proceeds by a straightforward induction on t. □

6.12 Hasse-Weil Theorem

In this section we use the Weil pairing and basic properties of isogenies to prove Hasse-Weil theorem, which allows on to bound the number of points of an elliptic curve over a finite field.

The following lemma is our starting point as it identifies the number of points of an elliptic curve as something more amenable to calculations.

Lemma 6.123. *Let E be an elliptic curve defined over $\mathbb{F}_q$. Then $\#E(\mathbb{F}_{q^n}) = \ker(\Phi_q^n - \mathrm{id})$*

Proof. Since $\Phi_q^n = \Phi_{q^n}$ is the Frobenius endomorphism on $E(\mathbb{F}_{q^n})$, $\Phi_q^n(P) = P$ iff $P \in E(\mathbb{F}_{q^n})$ as desired. □

In light of Lemma 6.123, we would like to show that $\Phi_q^n - \mathrm{id}$ is seperable so that its degree will equal to the cardinality of its kernel by Proposition 6.71.

It will be convenient to have a criterion for seperability. If (x, y) is a variable point on $E: y^2 = x^3 + Ax + B$, then we can differentiate y wrt x:

$$2yy' = 3x^2 + A.$$

Similarly, we can differentiate a rational function $f(x, y)$ wrt x:

$$\frac{d}{dx} f(x,y) = f_x(x,y) + f_y(x,y)y'$$

where f_x denotes the partial derivative wrt x ie deriving $f(x, y)$ while "treating y as a constant" and f_y denotes the partial derivative wrt y ie deriving $f(x, y)$ while "treating x as a constant".

Lemma 6.124. *Let $E/\mathbb{k} = E_{A,B} : y^2 = x^3 + Ax + B$ be an elliptic curve and $(u, v) \in E(\mathbb{k})$ a point. Write*

$$(x, y) + (u, v) = (f(x, y), g(x, y)),$$

where $f(x, y), g(x, y)$ are rational functions of x, y with coefficients depending on u, v, and y is regarded as a function of x with $dy/dx = (3x^2 + A)/2y$. Then

$$\frac{\frac{d}{dx} f(x,y)}{g(x,y)} = \frac{1}{y}.$$

Proof. By the formulas for point addition, we have

$$\begin{aligned} f(x,y) &= \left(\frac{y-v}{x-u}\right)^2 - x - u \\ g(x,y) &= \frac{-(y-v)^3 + x(y-v)(x-u)^2 + 2u(y-v)(x-u)^2 - v(x-u)^3}{(x-u)^3} \\ \frac{d}{dx} f(x,y) &= \frac{2y'(y-v)(x-u) - 2(y-v)^2 - (x-u)^3}{(x-u)^3} \end{aligned} \tag{58}$$

A straightforward but lengthy calculation using the fact that $2yy' = 3x^2 + A$ yields

$$\begin{aligned} &(x-u)^3 (y\frac{d}{dx} f(x,y) - g(x,y)) \\ &= v(Au + u^3 - v^2 - Ax - x^3 + y^2) + y(-Au - u^3 + v^2 + Ax + x^3 - y^2) \end{aligned} \tag{59}$$

Since $(x,y),(u,v)$ are points on $E(\Bbbk)$, we have $y^2 = x^3 + Ax + B$ and $v^2 = u^3 + Au + B$ so the last expression becomes

$$v(-B+B) + y(B-B) = 0.$$

Thus,

$$y\frac{d}{dx} f(x,y) = g(x,y)$$

as polynomials.

□

Lemma 6.125. *Let $\alpha_1, \alpha_2, \alpha_3$ be three endomorphisms of an elliptic curve E with $\alpha_1 + \alpha_2 = \alpha_3$ and write*

$$\alpha_i(x,y) = (R_i(x), yS_i(x))$$

for $i = 1,2,3$. Suppose there are constants c_1, c_2 such that

$$\frac{R_1'(x)}{S_1(x)} = c_1$$

and

$$\frac{R_2'(x)}{S_2(x)} = c_2.$$

Then

$$\frac{R_3'(x)}{S_3(x)} = c_3.$$

Proof. Let $(x_1,y_1),(x_2,y_2)$ be two variable points on E. Write

$$(x_3,y_3) = (x_1,y_1) + (x_2,y_2)$$

where $(x_1,y_1) = \alpha_1(x,y)$ and $(x_2,y_2) = \alpha_2(x,y)$. Then x_3 and y_3 are rational functions of x_1, y_1, x_2, y_2 which in turn are rational functions of x, y. By Lemma 6.124 with $(u,v) = (x_2,y_2)$,

$$\frac{\partial x_3}{\partial x_1} + \frac{\partial x_3}{\partial y_1}\frac{dy_1}{dx_1} = \frac{y_3}{y_1}.$$

Similarly,

$$\frac{\partial x_3}{\partial x_2} + \frac{\partial x_3}{\partial y_2}\frac{dy_2}{dx_2} = \frac{y_3}{y_2}.$$

By Assumption,

$$\frac{dx_i}{dx} = c_i S_i(x) = c_i \frac{y_i}{y}, \quad i = 1, 2.$$

Thus, by the chain rule,

$$\begin{aligned} \frac{dx_3}{dx} &= \frac{\partial x_3}{\partial x_1}\frac{dx_1}{dx} + \frac{\partial x_3}{\partial y_1}\frac{dy_1}{dx_1}\frac{dx_1}{dx} + \frac{\partial x_3}{\partial x_2}\frac{dx_2}{dx} + \frac{\partial x_3}{\partial x_2}\frac{dx_2}{dx} + \frac{\partial x_3}{\partial y_2}\frac{dy_2}{dx_2}\frac{dx_2}{dx} \\ &= \frac{y_3}{y_1}\frac{y_1}{y}c_1 + \frac{y_3}{y_2}\frac{y_2}{y}c_2 = (c_1 + c_2)\frac{y_3}{y} = (c_1 + c_2)S_3(x) \end{aligned} \tag{60}$$

and dividing by $S_3(x)$ yields the desired result. □

The next proposition characterises seperability of multiplication by n.

Proposition 6.126. *Let $E/\Bbbk$ be an elliptic curve and $1 \le n$. Suppose that multiplication by n on E is given by*

$$[n](x, y) = (R_n(x), yS_n(x))$$

for all $(x, y) \in E(\overline{\Bbbk})$ where R_n, S_n are rational functions. Then

$$\frac{R_n'(x)}{S_n(x)} = n.$$

Therefore, multiplication by n is seperable iff it is not a multiple of char $\Bbbk$.

Proof. Observe first that $R_{-n} = -R_n$ and $S_{-n} = -S_n$. Deriving the first equation and dividing it by the second equation yields

$$R_{-n}'/S_{-n} = -R_n'/S_n.$$

The first part of the proposition is trivially true for $n = 1$, so assume by induction it is true for n. Then Lemma 6.125 implies it is true for $n+1$ since $[n+1] = [n] + [1]$. Therefore,

$$\frac{R_n'(x)}{S_n(x)} = n$$

for all n so $R_n'(x) \neq 0$ iff $n = R_n'(x)/S_n(x) \neq 0$ which is equivalent to n not a multiple of char $\Bbbk$.

□

Using the previous proposition, we now derive a criterion for seperability

Proposition 6.127. *Let $E/\mathbb{F}_q$ be an elliptic curve where $q = p^n$ for a prime p. Let $r, s \neq 0$ be integers. Then the endomorphism $r\Phi_q + s$ is seperable iff $p \nmid s$.*

Proof. Write multiplication by r as

$$[r](x, y) = (R_r(x), yS_r(x)).$$

Then

$$\begin{aligned} &\left(R_{r\Phi_q}(x), yS_{r\Phi_q}(x)\right) \\ &= (\Phi_q r)(x,y) = (R_r^q(x), y^q S_r^q(x)) \\ &= \left(R_r^q(x), y(x^3+Ax+B)^{(q-1)/2} S_r^q(x)\right). \end{aligned} \tag{61}$$

Therefore,

$$c_{r\Phi_q} = R'_{r\Phi_q}/S_{r\Phi_q} = qR_r^{q-1}R'_r/S_{r\Phi_q} = 0.$$

By Proposition 6.126, $c_s = R'_s(x)/S_s(x) = s$ so by Lemme 6.125

$$R'_{r\Phi_q+s}(x)/S_{r\Phi_q+s}(x) = c_{r\Phi_q} + c_s = 0 + s = s.$$

Therefore, $R'_{r\Phi_q+s} \neq 0$ iff $p \nmid s$. □

Corollary 6.128. *Let $E/\mathbb{F}_q$ be an elliptic curve. Then the endomorphism $\Phi_q^n - \mathrm{id}$ is seperable. Thus $\#\ker(\Phi_q^n - \mathrm{id}) = \deg(\Phi_q^n - \mathrm{id}) = q^n$.*

Proof. Since Φ_q^n is the Frobenius map $\Phi_{q^n} : \mathbb{F}_{q^n}$ and E is defined over $\mathbb{F}_{q^n}$ (as it is defined over $\mathbb{F}_q$), the first part now follows from Proposition 6.127 with $s = 1$. The second part is an immediate application of Proposition 6.71 and the fact that $\Phi_q^n - \mathrm{id}$ has degree q^n as it is defined by polynomials of degree q^n. □

Let $E/\Bbbk$ be an elliptic curve and $\alpha : E \longrightarrow E$ an endomorphism. Since $E[n] \subseteq E(\overline{\Bbbk})$, α induces a group homomorphism $\alpha : E[n] \longrightarrow E[n]$. Let $\{T_1, T_2\}$ be generators of $E[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$. Then we can represent $\alpha|_{E[n]}$ as a matrix $\alpha_n = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ where $a,b,c,d \in \mathbb{Z}_n$, describing the action of α.

Proposition 6.129. *Let $E/\mathbb{F}_q$ be an elliptic curve and $\alpha : E \longrightarrow E$ a seperable endomorphism or the Frobenius map Φ_q. Suppose $0 < n$ such that $\operatorname{char}\Bbbk \nmid n$. Then*

$$\det(\alpha_n) = \deg(\alpha)(mod\, n)$$

Proof. By a corollary to Theorem 6.102, $\zeta = e_n(T_1, T_2)$ is a primitive nth root of unity. By Theorem 6.102 (8),

$$\begin{aligned} \zeta^{\deg(\alpha)} &= e_n(\alpha(T_1), \alpha(T_2)) = e_n(aT_1 + cT_2, bT_1 + dT_2) \\ &= e_n(T_1, T_T 1)^{ab} e_n(T_1, T_2)^{ad} e_n(T_2, T_1)^{cb} e_n(T_2, T_2)^{cd} \\ &= \zeta^{ad-bc} = \zeta^{\det(\alpha)}. \end{aligned} \tag{62}$$

and the desired result follows since ζ is a generator of a group of order n, μ_n. □

Let $E/\Bbbk$ be an elliptic curve, $E \underset{\beta}{\overset{\alpha}{\rightrightarrows}} E$ be two endomorphisms and $a, b \in \mathbb{Z}$ integers. We define the endomorphism $a\alpha + b\beta$ by

$$(a\alpha + b\beta)(P) = a\alpha(P) + b\beta(P).$$

Proposition 6.130. *If $a\alpha + b\beta$ is seperable, then*

$$\deg(a\alpha + b\beta) = a^2 \deg\alpha + b^2 \deg\beta + ab(\deg(\alpha+\beta) - \deg\alpha - \deg\beta).$$

Proof. Let n be an integer not divisible by $\operatorname{char} \Bbbk$ and represent $E[n] \overset{\alpha}{\underset{\beta}{\rightrightarrows}} E[n]$ by matrices α_n, β_n with respect to $\{T_1, T_2\}$. Then the matrix $a\alpha_n + b\beta_n$ represents $a\alpha + b\beta : E[n] \longrightarrow E[n]$. By standard linear algebra

$$\det(a\alpha_n + b\beta_n) = a^2 \det(\alpha_n) + b^2 \det(\beta_n) + ab(\det(\alpha_n + \beta_n) - \det \alpha_n - \det \beta_n)$$

for any square matrices α_n, β_n (we only need the 2×2 case, which can be verified with a direct calculation). Therefore, by Proposition 6.129

$$\begin{aligned} &\deg(a\alpha + b\beta) \\ &= a^2 \deg(\alpha_n) + b^2 \deg(\beta_n) + ab(\deg(\alpha_n + \beta_n) - \deg \alpha_n - \deg \beta_n) \ \ (mod\, n). \end{aligned} \tag{63}$$

Since this holds for infinitely many n, it must be an equality.

□

Lemma 6.131. *Let $E/\mathbb{F}_q$ be an elliptic curve and let r, s be integers with* $\gcd(s, q) = 1$. *Let Let* $t = q + 1 - \deg(\Phi_q - \mathrm{id})$. *Then* $\deg(r\Phi_q - s) = r^2 q + s^2 - rst$

Proof. By Proposition 6.130 implies that

$$\deg(r\Phi_q - s) = r^2 \deg(\Phi_q) + s^2 \deg(-\mathrm{id}) + rs\left(\deg(\Phi_q - \mathrm{id}) - \deg(\Phi_q) - \deg(-1)\right).$$

Since $\deg(-\mathrm{id}) = 1$ and $\deg(\Phi_q) = q$, the result follows from the definition of t. □

We are now ready to prove

Theorem 6.132 (Hasse-Weil)**.** *Let $E/\mathbb{F}_q$ be an elliptic curve. Then the order of $E(\mathbb{F}_q)$ satisfies*

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

Proof. By Corollary 6.128, $\deg(\Phi_q - \mathrm{id}) = \#\ker(\Phi_q - \mathrm{id}) = \#E(\mathbb{F}_q)$. Let $t = q + 1 - \deg(\Phi_q - \mathrm{id})$. We want to show that $|t| = 2\sqrt{q}$. For any r, s with $\gcd(s, q) = 1$, $0 \leq \deg(r\Phi_q - s)$ and Lemma 6.131, implies that

$$q(\frac{r}{s})^2 - t(\frac{r}{s}) + 1 \geq 0.$$

The graph of the function $y = qx^2 - tx + 1$ over $\mathbb{R}$ is a parabola that takes only non-negative y values for rational numbers $x = r/s$ with $\gcd(s, p) = 1$. By Lemma 6.133, this means that for every interval $[a, b] \in \mathbb{R}$, y takes some non-negative value. It follows that $y = qx^2 - tx + 1$ is a parabola that does not reach below the x-axis ie

$$qx^2 - tx + 1 \geq 0 \tag{64}$$

for all real numbers x. Thus the discriminant of the polynomial of Equation 64 is non-positive, ie

$$t^2 - 4q \leq 0 \iff |t| \leq 2\sqrt{q}.$$

This completes the proof. □

Lemma 6.133. *Let p be a prime and $a < b \in \mathbb{R}$ be real numbers. Then the interval $[a, b] \subseteq \mathbb{R}$ contains a rational point of the form r/s with* $\gcd(s, p) = 1$.

Proof. Recall that Archemedes axiom asserts that for any positive real numbers $x, y \in \mathbb{R}_{\geq 0}$, there is $n \in \mathbb{N}$ such that $nx > y$. Choose a prime $p' > p$ such that $p'(b-a) > 1 \iff b - a > \frac{1}{p'}$. Let $j = \lfloor p'a \rfloor$. Then $j \leq p'a \leq j+1$ so

$$\frac{j}{p'} \leq a \leq \frac{j+1}{p'} = \frac{j}{p'} + \frac{1}{p'} \leq a + (b-a) = b.$$

Thus, $\frac{r}{s} := \frac{j+1}{p'} \in [a,b]$ with $\gcd(s,p) = 1$ □

We finish this section with a re-statement of Hasse-Weil theorem that gives a more conceptual view.

Corollary 6.134. *Let $E/\mathbb{F}_q$ be an elliptic curve. Then in End(E) we have*

$$\Phi_q^2 - [t]\Phi_q + [q] = 0.$$

Proof. Let $\alpha = \Phi_q^2 - [t]\Phi_q + [q]$. If $\alpha \neq 0$ (as an endomorphism $E \longrightarrow E$) then $\# \ker \alpha \leq \deg \alpha$ (6.71) so in particular $\ker \alpha$ is a finite set. We'll show that $\ker \alpha$ is infinite hence $\alpha = 0$. For any integer n such that $p \nmid n$ we represent the restriction of Φ_q to $E[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$ by a matrix

$$A_n = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

whose entries are in $\mathbb{Z}_n$. As can be checked by direct inspection, for any 2×2 matrix A_n,

$$A_n^2 - \operatorname{tr}(A_n)A_n + \det(A_n)I = 0$$

where $\operatorname{tr}(A_n) = a + d$. We have shown that $\det(A_n) = \deg(\Phi_q)(mod\, n)$ and another direct calculation shows that

$$\operatorname{tr}(A_n) = 1 + \det(A_n) - \det(I - A_n).$$

Thus

$$\operatorname{tr}(A_n) = 1 + \deg(\Phi_q) + \deg(\operatorname{id} - \Phi_q)(mod\, n).$$

Recall that $\deg(\Phi_q) = q$ and in our notation $\deg(\operatorname{id} - \Phi_q) = \# E(\mathbb{F}_q) = q + 1 - t$ so we get

$$A_n - [1 + q - (q+1) - t]A_n + qI = 0 \; (mod\, n)$$

$$\iff$$

$$A_n - [t]A_n + qI = 0 \; (mod\, n)$$

Since the last equation is true for any n that is prime to p, $\alpha(P) = 0$ for any $P \in \bigcup_{p \nmid n} E[n]$ so $\ker \alpha$ is infinite, as desired. □

Remark 6.135. t above is called the **trace** of Frobeniuos.

6.13 Identifying torsion points via division polynomials

In this section we pay on old debt and prove Theorem 6.96 that asserts that for an elliptic curve $E_{A,B} = E/\mathbb{F}_q$ $(q = p^n)$ and n such that $\gcd(n,p) = 1$, we have an isomorphism

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

Proposition 6.126 tells us that for n with $p \nmid p$, multiplication by n, $[n] : E \longrightarrow E$ is seperable and Proposition 6.71 tells us that $\ker[n] = \deg[n]$. Observe that $\ker[n] := \ker([n] : E(\overline{\mathbb{F}}_q) \longrightarrow E(\overline{\mathbb{F}}_q)) = E[n]$ so we may hope to calculate $\#E[n]$ by calculating $\deg[n]$. In fact, once we know that $\#E[n] = n^2$, a simple group-theoretic argument shows that $E[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$.

Thus, our main focus is to find a formula for the canonical form

$$[n](x, y) = (R_n(x), S_n(x)y)$$

in order to calculate $\deg[n]$. For this, we will use

Construction 6.136 (Division polynomials)**.** We define polynomials $\psi_n, \phi_n, \omega_n \in \mathbb{Z}[x, y, A, B]$ inductively as follows.

1. $\psi_0 = 0$, $\psi_1 = 1$
2. $\psi_2 = 2y$
3. $\psi_3 = 3x^4 + 6Ax^2 + 12Bx - A^2$
4. $\psi_4 = 4y(x^6 + 5Ax^4 + 20Bx^3 - 5A^2x^2 - 4ABx - A^3 - 8B^2)$
5. $\psi_{2n+1} = \psi_{n+2}\psi_{3n} - \psi_{n-1}\psi_{3n+1}$
6. $\psi_{2n} = \frac{1}{2y}\psi_n(\psi_{n+2}\psi_{2n-1} - \psi_n - 2\psi_{2n+1})$

where we reduce the result modulo the curve equation so that ψ_n is at most linear in y. It is not difficult to show that $\psi_n(\psi_{n+2}\psi_{2n-1} - \psi_n - 2\psi_{2n+1})$ is always divisible by $2y$, so that ψ_{2n} is in fact a polynomial. If we define $\psi_{-n} := -\psi_n$, one can check that these recurrences hold for all integers n.

We then define ϕ_n and ω_n via

1. $\phi_n := x\psi_{2n} - \psi_{n+1}\psi_{n-1}$
2. $\omega_n := 14y(\psi_{n+2}\psi_{2n-1} - \psi_{n-2}\psi_{2n+1})$.

These equations hold for all integers n, and one finds that $\phi_n = \phi_{-n}$ and $\omega_n = \omega_{-n}$. As above, we reduce ϕ_n and ω_n modulo the curve equation to make them at most linear in y.

We then have

Theorem 6.137. *For any elliptic curve $E/\Bbbk = E_{A,B}$, any $n \in \mathbb{Z}$, and any $P = (x, y) \in E(\overline{\Bbbk})$,*

$$[n]P = (\frac{\phi_n(x, y)}{\psi_n^2(x, y)}, \frac{\omega_n(x, y)}{\psi_n^3(x, y)})$$

Proof. The proof is a straightforward but very tedious calculation that is feasible with a symbolic manipulation program like Matlab. □

The terms in the formulas of Theorem 6.137 satisfy the following properties

Lemma 6.138. *For every integer n,*

- *ψ_n lies in* $\begin{cases} \mathbb{Z}[x, A, B] & n \textit{ odd} \\ 2y\mathbb{Z}[x, A, B] & n \textit{ even.} \end{cases}$

- *ϕ_n lies in $\mathbb{Z}[x, A, B]$ for all n.*
- *ω_n lies in* $\begin{cases} \mathbb{Z}[x,A,B] & n \text{ even} \\ y\mathbb{Z}[x,A,B] & n \text{ odd.} \end{cases}$

Proof. Standard induction using the elliptic curve equation to reduce terms y^2 to $x^3 + Ax + B$. □

It follows from Lemma 6.138 that, after replacing y^2 with $x^3 + Ax + B$ if needed, ψ_n^2 lies in $\mathbb{Z}[x, A, B]$ for all n so we think of ϕ_n and ψ_n^2 as polynomials in x alone. Moreover, for all n, exactly one of ω_n and ψ_n^3 depends on y. In the later case we can multiply the numerator and denominator by y and replace the resulting y^2 in the denominator by $x^3 + Ax + B$ to so that we may view $\frac{\omega_n}{\psi_n^3}$ as y times a quotient of two polynomials in $\mathbb{Z}[x, A, B]$.

In this way, we may view $(\frac{\phi_n}{\psi_n^2}, \frac{\omega_n}{\psi_n^3})$ as a canonical from for the isogeny

$$[n] : E \longrightarrow E.$$

To compute the degree of $[n] : E \longrightarrow E$, we need to know the degrees of the polynomials $\phi_n(x)$ and $\psi_n^2(x)$, and we need to verify that they are relatively prime.

Lemma 6.139. *For every positive integer n the polynomials ϕ_n and ψ_n satisfy*

$$\phi_n(x) = x^{n^2} + ...,$$

$$\psi_n(x) = \begin{cases} nx^{\frac{n^2-1}{2}} + ... & \text{if } n \text{ odd} \\ ynx^{\frac{n^2-4}{2}} + ... & \text{if } n \text{ even} \end{cases}$$

where ... hides terms of lower x degree.

Proof. Tedious calculation though feasible by human. □

Corollary 6.140. *For all positive integers n, we have $\psi_n^2(x) = n^2 x^{n-1} + ...$, where ... hides terms of degree less than $n - 1$.*

Lemma 6.141. *For an elliptic curve $E_{A,B}/\Bbbk$, the polynomials $\phi_n(x)$ and $\psi_n^2(x)$ are relatively prime over $\overline{\Bbbk}$*

Proof. Suppose not and let $x_0 \in \overline{\Bbbk}$ a common root. Let $P = (x_0, y_0)$ be a point on $E(\overline{\Bbbk})$ that is different than $\mathcal{O}$. Then $nP = \mathcal{O}$ since $\psi_n^2(x_0) = 0$ and we also have

$$\phi_n(x_0) = x_0\psi_{2n}(x_0) - \psi_{n+1}(x_0, y_0)\psi_{n-1}(x_0, y_0)$$

$$\Longleftrightarrow$$

$$0 = 0 - \psi_{n+1}(x_0, y_0)\psi_{n-1}(x_0, y_0),$$

so at least one of $\psi_{n+1}(x_0, y_0)$ and $\psi_{n-1}(x_0, y_0)$ is zero. But then either $(n-1)P = \mathcal{O}$ or $(n+1)P = \mathcal{O}$ and after substracting $nP = \mathcal{O}$ we get either $-P = \mathcal{O}$ or $P = \mathcal{O}$ – contradiction. □

Corollary 6.142. *Let $E/\Bbbk$ be an elliptic curve. Then $[n] : E \longrightarrow E$ has degree n^2 and is seperable iff* $\operatorname{char} \Bbbk \nmid n$.

We are ready to prove

Theorem 6.143 (Theorem 6.96)**.** *Let $E/\Bbbk$ be an elliptic curve and n an integer with* $\operatorname{char}\Bbbk \nmid n$*. Then*

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

Proof. By Corollary 6.142 we have $\#E[n] = n^2$. By Theorem 4.83 we have

$$E[n] \cong \mathbb{Z}_{n_1} \times ... \times \mathbb{Z}_{n_{k-1}} \times \mathbb{Z}_{n_k}$$

where

$$n_1 \mid ... \mid n_{k-1} \mid n_k \mid n.$$

By the last decomposition, for any $P \in E[n]$ we have $n_k P = \mathcal{O}$ ie

$$E[n] \subseteq E[n_k]$$

so that $E[n] = E[n_k]$ and thus $n_k^2 = n^2 \iff n_k = n$.

Continuing, since $E[n_{k-1}] \subseteq E[n]$, the elements of $E[n_{k-1}]$ are precisely those elements of $E[n]$ that vanish under multiplication by n_{k-1}. In $\mathbb{Z}_n$ there are exactly n_{k-1} elements whose order divides n_{k-1} and clearly every element in $\mathbb{Z}_{n_1} \times ...\mathbb{Z}_{n_{k-1}}$ vanishes after multiplication by n_{k-1}. Therefor, we have

$$n_{k-1}^2 = \#E[n_{k-1}] = n_1 \cdot ... \cdot n_{k-1} \cdot n_{k-1}$$

and thus

$$n_1 = ... = n_{k-2} = 1.$$

But then

$$E[n] \cong \mathbb{Z}_{n_{k-1}} \times \mathbb{Z}_n$$

and since $\#E[n] = n^2$ we get $n_{k-1} = n$. □

6.14 The Tate Pairing

In this section we define the Tate pairing and prove it's basic properties. Tate pairing is an example of an **asymmetric** bilinear pairing and its implementation is considered to have lower computational complexity.

As we will see below, the Tate pairing can be expressed in terms of the Weil pairing. Thus, the effort we spent on proving the properties of the Weil pairing helps in proving the properties of the Tate pairing.

Our setup throughout this section is identical to that of the Weil pairing, but we repeat it for convenience. Let $E/\mathbb{F}_p$ be an elliptic curve defined over $\mathbb{F}_p$ and $1 \leq n$ an integer such that $p \nmid n$. We let $\mathbb{F}_p \subseteq \Bbbk$ be finite field extension such that $E(\Bbbk)[n] = E(\overline{\Bbbk})[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$. Recall from Theorem 6.102 that this setup implies that the group of nth roots of unity μ_n satisfies $\mu_n \subseteq \Bbbk = \mathbb{F}_{p^k}$.

Our setup implies, in particular, $\Bbbk = \mathbb{F}_{p^k}$ for some k. We start with a general result that expresses μ_n as a quotient of $\Bbbk^\times$.

Lemma 6.144. *Let $1 \leq n$ be an integer and suppose k is such that $\mu_n \subseteq \mathbb{F}_{p^k}$. Then*

$$\mathbb{F}_{p^k}^\times / \left(\mathbb{F}_{p^k}^\times\right)^n \cong \mu_n.$$

Proof. Define a map

$$\Phi : \mathbb{F}_{p^k}^\times / \left(\mathbb{F}_{p^k}^\times\right)^n \longrightarrow \mu_n$$

by $\Phi(\gamma) = \gamma^{\frac{p^k-1}{n}}$. The map Φ has range μ_n since $\Phi(\gamma)^n = \gamma^{p^k-1} = 1$ (given that $\mathbb{F}_{p^k}^\times$ is a cyclic group of order $p^k - 1$) and one can easily check it is a group homomorphism. Recall that $\left(\mathbb{F}_{p^k}^\times\right)^n = \{\alpha^n | \alpha \in \mathbb{F}_{p^k}^\times\}$ is a subgroup of $\mathbb{F}_{p^k}^\times$ since $\alpha^n\beta^n = (\alpha\beta)^n$.

Let us evaluate the order of $\left(\mathbb{F}_{p^k}^\times\right)^n$. If $\alpha, \beta \in \mathbb{F}_{p^k}^\times$, then

$$\alpha^n = \beta^n \iff \left(\frac{\alpha}{\beta}\right)^n = 1 \iff \frac{\alpha}{\beta} \in \mu_n.$$

Since $\mu_n \subseteq \mathbb{F}_{p^k}^\times$, it follows that $|\left(\mathbb{F}_{p^k}^\times\right)^n = \frac{p^k-1}{n}| = \frac{p^k-1}{n}$. By the first Isomorphism theorem (Theorem 4.47), it follows that

$$|\mathrm{Im}\,(\Phi)| = |\mathbb{F}_{p^k}^\times|/|\left(\mathbb{F}_{p^k}^\times\right)^n| = p^k - 1/\frac{p^k-1}{n} = n = |\mu_n|$$

and thus Φ is an isomorphism □

We are ready to phrase

Construction 6.145. Let $E/\mathbb{F}_p$ be an elliptic curve, n a prime with $n \mid \#E(\mathbb{F}_p)$ and k the minimal integer such that $n \mid p^k - 1$. Note that by Theorem 7.5, our assumptions mean that $E[n] \subseteq E(\mathbb{F}_{p^k})$. Let $P \in E[n]$ and $Q \in E(\mathbb{F}_{p^k})$.

Let D_P, D_Q be divisors with disjoint support such that

$$D_P \sim [P] - [\mathcal{O}]$$
$$D_Q \sim [Q] - [\mathcal{O}],$$

and let f_{nD_P} be a function such that

$$\mathrm{div}(f_{nD_P}) = nD_P.$$

We define the (reduced) **Tate pairing** to be

$$\tau_n(P, Q) = (f_{nD_P}(D_Q))^{\frac{p^k-1}{n}}.$$

Theorem 6.146. *The Tate pairing of Construction 6.145 defines a bilinear non-degenerate pairing*

$$\tau_n : E[n] \times E(\mathbb{F}_{p^k})/nE(\mathbb{F}_{p^k}) \longrightarrow \mu_n.$$

Proof. We first show that $\tau_n(P, Q)$ does not depend on the choices of D_P and D_Q (since we are not in a symmetric setting, both cases require a proof).

Let $D'_P = D_P + \mathrm{div}(g)$. If f_{nD_P} is a rational function corresponding to nD_P then

$$\mathrm{div}(f_{nD_P} g^n) = nD_P + n\,\mathrm{div}(g) = nD'_P$$

and thus:

$$\begin{aligned} f_{nD'_P}(D_Q)^{\frac{p^k-1}{n}} &= f_{nD_P}(D_Q)^{\frac{p^k-1}{n}} \cdot g(D_Q)^{p^k-1} \\ &= (f_{nD_P}(D_Q))^{\frac{p^k-1}{n}}. \end{aligned}$$

Similarly, if $D'_Q = D_Q + \operatorname{div}(h)$ s.th. D_P and $\operatorname{div}(h)$ have disjoint support, then

$$f_{nD_P}(\operatorname{div} h) = h(\operatorname{div} f_{nD_P}) = h([P] - [\mathcal{O}])^n$$

and thus

$$\begin{aligned} f_{nD_P}(D'_Q)^{\frac{p^k-1}{n}} &= f_{nD_P}(D_Q)^{\frac{p^k-1}{n}} f_{nD_P}(\operatorname{div} h)^{\frac{p^k-1}{n}} \\ = f_{nD_P}(D_Q)^{\frac{p^k-1}{n}} \cdot h([P] - [\mathcal{O}])^{p^k-1} &= f_{nD_P}(D_Q)^{\frac{p^k-1}{n}}. \end{aligned}$$

Note that a-priori, Construction 6.145 only defines a map

$$\tau_n : E[n] \times E(\mathbb{F}_{p^k}) \longrightarrow \mu_n.$$

To prove that we have a map

$$\tau_n : E[n] \times E(\mathbb{F}_{p^k})/nE(\mathbb{F}_{p^k}) \longrightarrow \mu_n,$$

fix $P \in E[n]$ and consider the map

$$\tau_n(P, -) : E(\mathbb{F}_{p^k}) \longrightarrow \mu_n$$

evaluated at $Q + nR$ for $Q, R \in E(\mathbb{F}_{p^k})$. Choose a divisor $D_P \sim [P] - [\mathcal{O}]$ whose support is disjoint from $\{\mathcal{O}, R, nR, Q\}$. Let $D_Q = [Q] - [\mathcal{O}]$ and $D_{Q+nR} = [Q + nR] - [\mathcal{O}]$. Then

$$D_{Q+nR} \sim D_Q + n[R] - n[\mathcal{O}].$$

Since τ_n is independent of the divisor class of D_{Q+nR} we get:

$$\begin{aligned} \tau_n(P, Q + nR) &= f_{nD_P}(D_{Q+nR})^{\frac{p^k-1}{n}} \\ &= f_{nD_P}(D_Q + n[R] - n[\mathcal{O}])^{\frac{p^k-1}{n}} \\ &= f_{nD_P}(D_Q)^{\frac{p^k-1}{n}} \cdot f_{nD_P}(n[R] - n[\mathcal{O}])^{\frac{p^k-1}{n}} \\ &= f_{nD_P}(D_Q)^{\frac{p^k-1}{n}} \cdot f_{nD_P}([R] - [\mathcal{O}])^{p^k-1} \\ &= f_{nD_P}(D_Q)^{\frac{p^k-1}{n}} = \tau_n(P, Q) \end{aligned}$$

as required.

To prove that τ_n is a non-degenerate bilinear pairing, let us give an alternative definition of τ_n.

Let $P \in E[n]$ and $Q \in E(\mathbb{F}_{p^k})$. Choose $Q_0 \in E(\overline{\mathbb{F}}_{p^k})$ s.th. $Q = nQ_0$. Let

$$\Phi : E(\overline{\mathbb{F}}_p) \longrightarrow E(\overline{\mathbb{F}}_p)$$

be the Frobenius map over $\mathbb{F}_p$ (given by $\Phi(x, y) = (x^p, y^p)$). Then

$$\Phi^k : E(\overline{\mathbb{F}}_{p^k}) \longrightarrow E(\overline{\mathbb{F}}_{p^k})$$

is the Frobenius map over $\mathbb{F}_{p^k}$. Denote $Q_1 = (\Phi^k - 1)(Q_0) = \Phi^k Q_0 - Q_0$. Then

$$nQ_1 = \Phi^k(nQ_0) - nQ_0 = \Phi^k Q - Q = 0$$

where the last equality follows since Φ^k acts as identity on points $Q \in E(\mathbb{F}_{p^k})$. Thus $Q_1 \in E[n]$.

Observe – Q_1 does not depend on the choice of Q_0: if $T \in E[n]$ then $n(Q_0 + T) = Q_1$ and

$$(\Phi^k - 1)(Q_0 + T) = Q_1 + (\Phi^k - 1)(T) = Q_1,$$

where the last equality holds since $T \in E[n] \subseteq E(\mathbb{F}_{p^k})$.

We thus get a well-defined map

$$\frac{\Phi^k - 1}{n} : E(\mathbb{F}_{p^k}) \longrightarrow E[n] \subseteq E(\mathbb{F}_{p^k})$$

given by

$$Q \mapsto (\frac{\Phi^k - 1}{n})(Q) = (\frac{\Phi^k - 1}{n})(nQ_0) = Q_1.$$

Furthermore, it is easy to check that $\frac{\Phi^k-1}{n}$ is an endomorphism of $E(\mathbb{F}_{p^k})$.

Recall from Construction 6.100 that

$$e_n(P, Q_1) = \frac{g_P(S + Q_1)}{g_P(S)}$$

for an arbitrary S. Taking $S = Q_0$, we get

$$e_n(P, Q_1) = e_n(P, \Phi^k Q_0 - Q_0) = \frac{g_P(\Phi^k Q_0)}{g_P(Q_0)}.$$

Now, $P \in E[n] \subseteq E(\mathbb{F}_{p^k})$ so that $g_P \in \mathbb{F}_{p^k}(E)$.

By Corollary 6.59, we get:

$$\begin{aligned} e_n(P, Q_1) &= \frac{g_P(\Phi^k Q_0)}{g_P(Q_0)} \\ &= \frac{\Phi^k(g_P(Q_0))}{g_P(Q_0)} = g_P(Q_0)^{p^k-1} \\ &= (g_P^n(Q_0))^{\frac{p^k-1}{n}} = \left(f_{n[P]-n[\mathcal{O}]}(nQ_0)\right)^{\frac{p^k-1}{n}} \\ &= f_{nD_P}(Q)^{\frac{p^k-1}{n}} = \tau_n(P, Q) \end{aligned} \tag{65}$$

where $D_P = [P] - [\mathcal{O}]$.

Using equation 65 we deduce bilinearity of the Tate pairing from bilinearity of the Weil pairing.

For non-degeneracy, it is enough to show that

$$\frac{\Phi^k - 1}{n} : E(\mathbb{F}_{p^k}) \longrightarrow E[n]$$

is surjective since we could then use non-degeneracy of the Weil pairing. We have

$$\ker\left(\frac{\Phi^k - 1}{n}\right) = nE(\mathbb{F}_{p^k})$$

and so by the first isomorphism theorem 4.47,

$$\operatorname{Im}\left(\frac{\Phi^k - 1}{n} : E(\mathbb{F}_{p^k}) \longrightarrow E[n]\right) \cong E(\mathbb{F}_{p^k})/nE(\mathbb{F}_{p^k}).$$

Now, by Corollary 6.97 , $E(\mathbb{F}_{p^k}) \cong \mathbb{Z}_a \times \mathbb{Z}_b$ for some $a \mid b$ and since $\mathbb{Z}_n \times \mathbb{Z}_n \leq E(\mathbb{F}_{p^k}) \cong \mathbb{Z}_a \times \mathbb{Z}_b$, we deduce from [Toth, Theorem 4.5] that $n \mid a$ and $n \mid b$.

Write $a = \alpha n$ and $b = \beta n$. Then the map

$$\varphi : \mathbb{Z}_a \times \mathbb{Z}_b \longrightarrow \mathbb{Z}_n \times \mathbb{Z}_n$$

given by

$$\varphi(x, y) = (x mod\, n, y mod\, n)$$

is a homomorphism whose kernel satisfies

$$\ker \varphi = \{(x, y) | n \mid x \wedge n \mid y\} = n(\mathbb{Z}_a \times \mathbb{Z}_b).$$

Since φ is clearly surjective, we deduce from the first isomorphism theorem that

$$E(\mathbb{F}_{p^k})/nE(\mathbb{F}_{p^k}) \cong \mathbb{Z}_a \times \mathbb{Z}_b/n(\mathbb{Z}_a \times \mathbb{Z}_b) \cong \mathbb{Z}_n \times \mathbb{Z}_n.$$

Since $E[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$, we deduce that the image of $\frac{\Phi^k - 1}{n}$ is $E[n]$, ie that it is surjective.

□

7 Pairing-friendly elliptic curves

7.1 Ordinary and supersingular elliptic curves

Definition 7.1. Let $q = p^n$ be a prime power and $E/\mathbb{F}_q$ an elliptic curve with t being the trace of Frobenius endomorphism $\Phi_q : E(\mathbb{F}_q) \longrightarrow E(\mathbb{F}_q)$. The curve E is called **ordinary** if $t \neq 0\ (mod\, p)$ and **supersingular** otherwise.

Theorem 7.2. *Let p be a prime and $E/\mathbb{F}_p$ an elliptic curve. Then E is supersingular iff $t = 0$ (or equivalently iff $\#E(\mathbb{F}_p) = p + 1$) and ordinary otherwise.*

7.2 Embedding degree

Definition 7.3. Let p be a prime and $q = p^n$. Let $E/\mathbb{F}_q$ be an elliptic curve and suppose that r is a prime with $r \mid \#E(\mathbb{F}_q)$. The **embedding degree** of E wrt r is the smallest integer k such that $r \mid p^k - 1$.

In the definition above, since $E(\mathbb{F}_q)$ is a finite abelian group, it has the converse Lagrange property, meaning that if $r \mid \#E(\mathbb{F}_q)$, there is a subgroup $\mathbb{G}_1 \leq E(\mathbb{F}_q)$ of order r. Since r is prime, G_1 is cyclic, say $\mathbb{G}_1 = \langle g_1 \rangle$. Moreover, since $\mathbb{G}_1$ is of prime order, every element has order r, which means that $\mathbb{G}_1 \leq E[r]$. In order to transform the "theoretical" Tate pairing 6.146 into a type II pairing, we need to find an integer k such that $E[r] \subseteq E(\mathbb{F}_{q^k})$ and for efficiency, we'd like it to have the minimal such k. As we will see, the embedding degree is the integer we're after.

Remark 7.4. When using the embedding degree, we need to add two more assumptions. First, that $\gcd(r, q - 1) = 1$ and second that $\gcd(r, k) = 1$.Both assumptions are satisfied in all practical cases.

Our goal in this section is to prove a light version of the following

Theorem 7.5 (Balasubramanian-Koblitz)**.** *[BK] Let $E/\mathbb{F}_q$ be an elliptic curve and suppose that $E(\mathbb{F}_q)$ has a subgroup $G = \langle P \rangle$ of order r with $\gcd(r, p - 1) = 1$. Then $E[r] \subseteq E(\mathbb{F}_{q^k})$ iff $r \mid q^k - 1$.*

Let $E/\mathbb{F}_q$ be an elliptic curve and $\Phi = \Phi_q : E(\overline{\mathbb{F}}_q) \longrightarrow E(\overline{\mathbb{F}}_q)$ be the Frobenius endomorphism. Recall from 6.134 that the trace of Frobenius $t := q + 1 - \#E(\mathbb{F}_q)$ satisfies

$$\Phi^2 - [t]\Phi + [q] = 0.$$

Lemma 7.6. *Let r be a prime such that $r \mid \#E(\mathbb{F}_q)$. Then in $End(E)$ we have*

$$(\Phi - [1])(\Phi - [q]) = 0(mod\, r)$$

Proof. Denote $hr = \#E(\mathbb{F}_q)$ and $p(x) = x^2 - tx + q$ the characteristic polynomial of Φ. Since the trace of Frobenius has the form $t = q + 1 - \#E(\mathbb{F}_q)$, we get

$$p(x) = x^2 - (q + 1 - hr)x + q \cong x^2 - (q+1)x + q \ (mod\, r)$$

and thus

$$p(x) \cong (x-1)(x-q) \ (mod\, r)$$

and we conclude that

$$0 = \Phi^2 - [t]\Phi + [q] \cong (\Phi - [1])(\Phi - [q]) \ (mod\, r).$$

□

Let $\mathrm{Eig}_\ell(\Phi) = \{P \in E(\overline{\mathbb{F}}_q) | \Phi(P) = \ell P\}$ be the ℓ eigenspace of Φ. We denote $H_1 := \mathrm{Eig}_1(\Phi) \cap E[r]$ and $H_q := \mathrm{Eig}_q(\Phi) \cap E[r]$.

Corollary 7.7. *We have*

$$E[r] = \{aP + bQ | P \in H_1, \ Q \in H_q, a, b \in \mathbb{Z}\}.$$

Proof. Clearly, H_1 is a subgroup of $E[r]$ since if $P, P' \in H_1$, $\Phi(P+P') = \Phi(P) + \Phi(P') = P+P'$ and $\Phi(-P) = -\Phi(P) = -P$. Similarly, $H_q \leq E[r]$. Now, $H_1 \cap H_q = \{\mathcal{O}\}$ and from Lemma 7.6 it follows that

$$E[r] \subseteq \{R \in E(\overline{\mathbb{F}}_q) | (\Phi - [1])(\Phi - [q])(R) = 0\}$$

so $E[r] = \langle H_1, H_q \rangle$ as desired. □

Definition 7.8. Let $E/\mathbb{F}_q$ be an elliptic curve, $r \mid \#E(\mathbb{F}_q)$ a prime with $\gcd(r, p-1) = 1$ and k the embedding degree of E wrt r, ie the minimal integer such that $r \mid q^k - 1$. The **trace map** (not to be confused with the trace of Frobenius) is the map

$$\mathrm{Tr} : E(\mathbb{F}_{q^k}) \longrightarrow E(\mathbb{F}_q)$$

defined by

$$\mathrm{Tr}(P) = P + \Phi(P) + ... + \Phi^{k-1}(P)$$

.

Remark 7.9. Note that Φ^k is the Forbenius map for q^k hence the identity on $\mathbb{F}_{q^k}$. Thus

$$\Phi(\mathrm{Tr}(P)) = \mathrm{Tr}(P)$$

so Φ fixes all points of the form $\mathrm{Tr}(P)$, which means that $\mathrm{Im}\,(\mathrm{Tr}) \subseteq E(\mathbb{F}_q)$.

Lemma 7.10. *The k-eigenspace of* Tr *is* $E(\mathbb{F}_q)[r]$.

Proof. If $R \in E(\mathbb{F}_q)[r]$ then $\Phi(R) = R$ which means $\mathrm{Tr}(R) = R + ... + R = kR$. Conversely, if $R \in E[r]$ such that $\mathrm{Tr}(R) = kR$, we want to show that R is defined over $\mathbb{F}_q$. Indeed, $\Phi(\mathrm{Tr}(R)) = \Phi(kR) = k\Phi(R)$ but since $\Phi(\mathrm{Tr}(R)) = \mathrm{Tr}(R)$ we get $k\Phi(R) = \Phi(\mathrm{Tr}(R)) = \mathrm{Tr}(R) = kR$ so that

$$k(\Phi(R) - R) = 0.$$

Since $\gcd(k, r) = 1$, $\Phi(R) - R = \mathcal{O}$ ie $\Phi(R) = R$ and since Φ fixes R we conclude that $R \in E(\mathbb{F}_q)$. □

Proposition 7.11. *Let $E/\mathbb{F}_q$ be an elliptic curve, $r \mid \#E(\mathbb{F}_q)$ prime with $\gcd(r, q-1) = 1$ and k the embedding degree of E wrt r. For H_1, H_q as above, we have:*

1. $H_1 = E(\mathbb{F}_q)[r]$
2. $H_q = \{R \in E[r] | \mathrm{Tr}(R) = \mathcal{O}\}$.

Proof.

1. Follows from the fact that the fixed points of Φ are precisely $E(\mathbb{F}_q)$.
2. Recall from Corollary 7.7 that we can write

$$E[r] = \{aP + bQ | P \in H_1, Q \in H_q, a, b \in \mathbb{Z}\}$$

Let $R \in E[r]$ with $\mathrm{Tr}(R) = \mathcal{O}$ and write $R = aP + bQ$ with $P \in H_1$ and $Q \in H_q$.
Then, $\Phi(R) = \Phi(aP + bQ) = aP + bqQ$ and

$$\Phi^2(R) = \Phi(aP + bqQ) = aP + bq^2Q.$$

We thus get

$$\begin{aligned}\mathcal{O} = \mathrm{Tr}(R) &= (1 + \Phi + ... + \Phi^{k-1})(R)\\ &= kaP + (1 + q + ... + q^{k-1})bQ\\ &= kaP + \left(\frac{q^k - 1}{q - 1}\right)bQ\end{aligned}$$

so $ka \cong 0\ (mod\ r)$. Since by assumption $\gcd(k, r) = 1$, we get $a \cong 0\ (mod\ r)$ so that $R = bQ \in H_q$. Conversely, if $R \in H_q$, then

$$\mathrm{Tr}(R) = \frac{q^k - 1}{q - 1}Q$$

and since $r \mid q^k - 1$ and $r \nmid q - 1$, we conclude that $r \mid \frac{1-q^k}{1-q}$ so $\mathrm{Tr}(R) = \mathcal{O}$.

□

Corollary 7.12 (Balasubramanian & Koblitz, [BK])**.** *Let $E/\mathbb{F}_q$ be an elliptic curve and r a prime such that $r \mid \#E(\mathbb{F}_q)$,* $\gcd(r, q-1) = 1$ *and k the embedding degree of E wrt to r (so $r \mid q^k - 1$). Then*

$$E[r] \subseteq E(\mathbb{F}_q^k).$$

Proof. Let $R \in E[r]$ and write $R = aP + bQ$ with $P \in H_1$ and $Q \in H_q$. Then $\mathrm{Tr}(Q) = \mathcal{O}$ by Proposition 7.11 and $\Phi^k(Q) = q^k Q = Q$ since $r \mid q^k - 1$. Furthermore, $\Phi^k(P) = P$ since $P \in E(\mathbb{F}_q)$.

Thus, $\Phi^k(aP + bQ) = aP + bQ$ so that Φ^k fixes $E[r]$. Since Φ^k is the Frobenius map for q^k, (that fixes precisely $E(\mathbb{F}_{q^k})$), we conclude that

$$E[r] \subseteq E(\mathbb{F}_{q^k}).$$

□

Although there is no standard notion of a pairing-friendly curve, the following definition seems to capture the most relevant properties in practice.

Definition 7.13. Let q be a prime. We say that an elliptic curve $E/\mathbb{F}_q$ is **pairing friendly** if

1. There exists a prime $r > \sqrt(q)$ with $r \mid \#E(\mathbb{F}_q)$.
2. The embedding degree k of E wrt r satisfies $k \leq \log_2(r)/8$.

We are interested in extracting a type II pairing

$$\mathbb{G}_1 \times \mathbb{G}_2 \longrightarrow \mathbb{G}_T$$

Let $E/\mathbb{F}_q$ be a pairing-friendly elliptic curve with a prime r with $\gcd(r, q-1) = 1$ and such that $r \mid \#E(\mathbb{F}_q)$ with an embedding degree k. For $\Phi : E(\overline{\mathbb{F}}_q) \longrightarrow E(\overline{\mathbb{F}}_q)$ the Frobenius endomorphism, Proposition 7.11 shows, $|H_1| = |H_q| = r$ so a canonical choice for the groups is

$$\begin{aligned} \mathbb{G}_1 = H_1 = E[r] \cap \mathrm{Eig}_1(\Phi) &= E[r] \cap E(\mathbb{F}_q) \\ \mathbb{G}_2 = H_q = E[r] \cap \mathrm{Eig}_q(\Phi) &= E[r] \cap \ker(\Phi - [q]) \subseteq E(\mathbb{F}_{q^k}) \\ \mathbb{G}_T &= \mu_r \subseteq \mathbb{F}_{q^k} \end{aligned}$$

Thus, we can use the (reduced) Tate pairing restricted to the groups above to obtain a type II pairing:

$$e : \mathbb{G}_1 \times \mathbb{G}_2 \longrightarrow \mathbb{G}_T$$

To find a generator of $\mathbb{G}_1$, write $\#E(\mathbb{F}_q) = hr$ (h is called the "cofactor"). Then given any point $P \in E(\mathbb{F}_q)$, such that $hP \neq \mathcal{O}$, we have $r(hP) = \mathcal{O}$ so $hP \in \mathbb{G}_1$ is a generator.

To represent the group $\mathbb{G}_2$, we use the following

Theorem 7.14 ([HSV])**.** *Let $E/\mathbb{F}_q$ be an ordinary elliptic curve and $r > d$ a prime such that $r \mid \#E(\mathbb{F}_q)$ and $r^2 \mid \#E(\mathbb{F}_{q^d})$ with d minimal. Then there is a unique degree d twist E' of E such that $r \mid E'(\mathbb{F}_{q^d})$ which we denote as*

$$\varphi_d : E'(\mathbb{F}_q) \longrightarrow E(\mathbb{F}_{q^d}).$$

Furthermore, if we denote by $\mathbb{G}_2'$ the (unique) subgroup of $E'(\mathbb{F}_{q^d})$ of order r, the twist $\varphi_m : E'(\mathbb{F}_{q^2}) \longrightarrow E(\mathbb{F}_{q^k})$ is a monomorphism that maps $\mathbb{G}_2'$ isomorphically to $\mathbb{G}_2$.

Construction 7.15. Let $E/\mathbb{F}_q$ be an ordinary (pairing-friendly) elliptic curve with a subgroup of order r. Assume that E admits a twist of degree d and let $m = \gcd(k, d)$ and $e = k/m$. Then there is a unique degree m twist E' of E over $\mathbb{F}_{q^e}$ such that $r \mid E'(\mathbb{F}_{q^e})$ which we denote as

$$\varphi_m : E'(\mathbb{F}_{q^e}) \longrightarrow E(\mathbb{F}_{q^{em}}) = E(\mathbb{F}_{q^k}).$$

Thus, we obtain a modified type II pairing:

$$\hat{e} : \mathbb{G}_1 \times \mathbb{G}_2' \longrightarrow \mathbb{G}_T$$

given by

$$\hat{e}(P, Q') = e(P, \varphi_m(Q')).$$

7.3 BLS12-381

In this section we discuss a specific pairing-friendly curve which is optimised for zk-snarks and DFT. BLS12 is a parametrised family of pairing-friendly elliptic curves whose construction can be found in [FST, Construction 6.6]. The basic equation of the curve is

$$E : y^2 = x^3 + 4$$

defined over a prime field $\mathbb{F}_q$. The embedding degree is set to 12 and the key parameters are set using a single parameter x (different the the x of the curve's equation):

1. Subgroup order $r(\mathrm{x}) = \mathrm{x}^4 - \mathrm{x}^2 + 1$.
2. Frobenius trace $t(\mathrm{x}) = \mathrm{x} + 1$.
3. Field size $q(\mathrm{x}) = (\mathrm{x} - 1)^2/3 \cdot (\mathrm{x}^4 - \mathrm{x}^2 + 1) + \mathrm{x}$.

Specific design goals for BLS12-381 are:
x has "low hamming weight", meaning that it has very few bits set to 1. This is particularly important for the efficiency of the algorithm that calculates pairings (the Miller loop). The field modulus q mentioned above is prime and has 383 bits or fewer, which makes 64-bit or 32-bit arithmetic on it more efficient. The order r of the subgroups we use is prime and has 255 bits or fewer, which is good for the same reason as above. The security target is 128 bits - see below. To support zkSnark schemes, we want to have a large power of two root of unity in the field $\mathbb{F}_r$. This means we want $2n$ to be a factor of $r - 1$, for some biggish n. (Making x a multiple of $2^{\frac{n}{2}}$ will achieve this.) This property is key to being able to use fast Fourier transforms for interesting things like polynomial multiplication.

The value $\mathrm{x} = -0xd201000000010000$ (hexadecimal, note that it is negative) gives the largest q and the lowest Hamming weight meeting these criteria. With this x value we have

$$\begin{aligned} q = 0x1a0111ea397fe69a4b1ba7b6434bacd764774b84& \\ &f38512bf6730d2a0f6b0f6241eabfffeb153ffffb9feffffffffaaab \end{aligned}$$

$$r = 0x73eda753299d7d483339d80809a1d80553bda402fffe5bfeffffffff00000001$$

Since $j(E) = 0$, we can expect there will be a degree 6 twist by Theorem 6.76 which is indeed the case. By Construction 7.15, we are guaranteed to have a degree $m = \gcd(k, d) = 6$ twist E' of E over $\mathbb{F}_{q^2}$ such that $r \mid E'(\mathbb{F}_{q^2})$ and the twist $\varphi = \varphi_m : E'(\mathbb{F}_{q^2}) \longrightarrow E(\mathbb{F}_{q^12})$ that maps the group $\mathbb{G}_2' \subseteq E(\mathbb{F}_{q^2})$ isomorphically to $\mathbb{G}_2 \subseteq E(\mathbb{F}_{q^k})$.

To represent $\mathbb{F}_{q^2}$, lets first note the following

Lemma 7.16. *Let q be a prime. The polynomial $g(x) = x^2 + 1$ is prime over $\mathbb{F}_q$ iff $q \neq 1 \ (mod\, 4)$.*

Proof. Otherwise, let α be a root of g. Then $\alpha^2 = -1 \Rightarrow \alpha^4 = 1$. But $|\mathbb{F}_q^x| = q - 1$ so $4 \mid q - 1 \iff q = 1 \ (mod\, 4)$. □

As can be checked, in our case $q = 3 \ (mod\, 4)$ so we can view $\mathbb{F}_{q^2}$ as the field $\mathbb{F}_q/\langle x^2 + 1\rangle$ which can be conveniently be written as $\mathbb{F}_q[i]$ where i satisfies $i^2 = -1$ (cf. Corollary 5.77). Then, the equation of E' is given by

$$E' : y^2 = x^3 + 4(1 + i)$$

and the monomorphism φ is given by

$$(x, y) \mapsto (1/(1 + i)^{1/3}x, 1/(1 + i)^{1/2}y)$$

8 Other topics

8.1 Lattice theory

8.1.1 Matrices and determinant

Let V, W be finitely generated vector spaces over a (common) field $\mathbb{F}$, of dimension n, m respectively, and let $f : V \longrightarrow W$ be a linear map. If we choose a basis $\mathcal{B} = (v_1, ..., v_n)$ for V, then since $\operatorname{Span}(\mathcal{B}) = V$ and f commutes with linear combinations, the values $f(v_1), ..., f(v_n)$ completely determine f: if $v \in V$, then there are scalars $a_1, ..., a_n$ such that $v = a_1 v_1 + ... + a_n v_n$ and thus $f(v) = f(a_1 v_1 + ... + a_n v_n) = a_1 f(v_1) + ... + a_n f(v_n)$ (the last equality used linearity of f n-times). If in addition we are given a basis $\mathcal{C} = (w_1, ..., w_m)$ for W, then for every $1 \leq k \leq n$, there are scalars $a_{1,k}, ..., a_{m,k}$ such that

$$f(v_k) = a_{1,k} w_1 + ... + a_{m,k} w_m.$$

In light of the previous discussion we can say that the scalars $\{a_{i,j}\}_{i\in[n],j\in[m]}$ determine f and we assemble them into a matrix of size $m \times n$ (note the swap of orders) $A = A(f) = (a_{i,j})_{i\in[m],j\in[n]}$ The picture is as follows

$$[A(f)]_{\mathcal{C}}^{\mathcal{B}} = \begin{bmatrix} f(\mathbf{v}_1)_{\mathcal{C}} & f(\mathbf{v}_2)_{\mathcal{C}} & \cdots & f(\mathbf{v}_n)_{\mathcal{C}} \end{bmatrix}$$

Here, $f(\mathbf{v}_k)_{\mathcal{C}} = a_{1,\bullet}^T$ represents the coordinates of $f(\mathbf{v}_k)$ with respect to the basis $\mathcal{C}$, as a column vector.

Example 8.1. Let V be an n-dimensional vector space with a chosen basis $\mathcal{B}$. Then for the identity map $\mathrm{id}_V : V \longrightarrow V$ we have

$$[\mathrm{id}_V]_{\mathcal{B}}^{\mathcal{B}} = I_n$$

where $I \equiv I_n$ is the matrix whose diagonal entries are all 1 and all other entries are 0

Consider two matrices $A \in \mathrm{Mat}_{m\times n}(\mathbb{F})$ and $B \in \mathrm{Mat}_{n\times k}(\mathbb{F})$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nk} \end{bmatrix}$$

The product AB is computed as follows:

$$AB = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} & \cdots & a_{11}b_{1k} + a_{12}b_{2k} + \cdots + a_{1n}b_{nk} \\ a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1} & \cdots & a_{21}b_{1k} + a_{22}b_{2k} + \cdots + a_{2n}b_{nk} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{11} + a_{m2}b_{21} + \cdots + a_{mn}b_{n1} & \cdots & a_{m1}b_{1k} + a_{m2}b_{2k} + \cdots + a_{mn}b_{nk} \end{bmatrix} \in \mathrm{Mat}_{m\times k}(\mathbb{F})$$

There is a compact way to view matrix multiplication. For vectors in $a, b \in \mathbb{F}^n$ with coordinates

$$a = (a_1, ..., a_n),\ b = (b_1, ..., b_n)$$

let $a \cdot b := \sum_{i=1}^n a_i b_i$. For A, B as above, denote by A^i the ith row vector of A and by B_j the jth column vector of B. If we view A^i and B_j as vectors in $\mathbb{R}^n$ so that

A =

$$\begin{bmatrix} — & A^1 & — \\ — & A^2 & — \\ & \vdots & \\ — & A^m & — \end{bmatrix}$$

B =

$$\begin{bmatrix} | & | & & | \\ B_1 & B_2 & \cdots & B_k \\ | & | & & | \end{bmatrix}$$

Then the ijth entry of AB is $A^i \cdot B_j$ ie
AB =

$$\begin{bmatrix} A^1 \cdot B_1 & A^1 \cdot B_2 & \cdots & A^1 \cdot B_k \\ A^2 \cdot B_1 & A^2 \cdot B_2 & \cdots & A^2 \cdot B_k \\ \vdots & \vdots & \ddots & \vdots \\ A^m \cdot B_1 & A^m \cdot B_2 & \cdots & A^m \cdot B_k \end{bmatrix}$$

For square matrices $A, B \in \mathrm{Mat}_{n\times n}$ both AB and BA are defined but are generally not equal so the binary operation $\cdot$ on $\mathrm{Mat}_{n\times n}$ is not commutative.

In contrast to the complicated definition of matrix multiplication (which will become clear in 8.5), addition and multiplication by scalar of matrices is defined entry-wise. That is, if $A = (a_{ij})$ and $B = (b_{ij})$ are two $n \times m$ matrices then $A + B := (a_{ij} + b_{ij})$ is the obvious $n \times m$ matrix and for $\lambda \in \mathbb{F}$, $\lambda \cdot A := (\lambda \cdot a_{ij})$.

Exercise 8.2.

1. Show that the collection of all matrices $\mathrm{Mat}_{n\times m}(\mathbb{F})$ is a vector space over $\mathbb{F}$. What is its dimension? describe a basis for it.

2. Given square matrices A, B, C show that $A(B + C) = AB + AC$ and that $(B + C)A = BA + CA$.

However, since we have a neutral element wrt to multiplication, we may define

Definition 8.3. A square matrix $A \in \mathrm{Mat}_{n\times n}$ is **invertible** if there exists a square matrix $A^{-1} \in \mathrm{Mat}_{n\times n}$ such that $A \cdot A^{-1} = I = A^{-1} \cdot A$.

Exercise 8.4.

1. Show that for any n and $A \in \mathrm{Mat}_{n\times n}(\mathbb{F})$, $A \cdot I = A = I \cdot A$.

2. Let $A \in \mathrm{Mat}_{n\times n}$ be a square matrix. Show that if there is $B \in \mathrm{Mat}_{n\times n}$ such that $AB = I$ then $B = A^{-1}$. In other words a one-sided inverse for A is automatically a two-sided inverse.

We then have the following theorem

Theorem 8.5. *Let U, V, W be vector spaces over a field $\mathbb{F}$ and $f : U \longrightarrow V$, $g : V \longrightarrow W$ linear maps. If we are given bases $\mathcal{B}, \mathcal{C}, \mathcal{D}$ for U, V, W respectively, then*

$$[A(g \circ f)]^{\mathcal{B}}_{\mathcal{D}} = [A(g)]^{\mathcal{C}}_{\mathcal{D}}[A(f)]^{\mathcal{B}}_{\mathcal{C}}.$$

In other words, matrix multiplication corresponds to composition of linear maps.

Proof. Direct (though tedious) computation. □

Corollary 8.6. *Let V, W be vector spaces with chosen bases $\mathcal{B} = (v_1, ..., v_n)$ and $\mathcal{C} = (w_1, ..., w_m)$ respectively. Then a linear map $f : V \longrightarrow W$ is an isomorphism iff $n = m$ and $A = A(f)$ is an invertible $n \times n$ matrix.*

Proof. First, since an isomorphism maps a basis to a basis, f is an isomorphism iff $n = m$. Second, f is an isomorphism iff $\exists g : W \longrightarrow V$ such that $g \circ f = \mathrm{id}_V$ and $f \circ g = \mathrm{id}_W$. If we denote $A = A(f)^{\mathcal{B}}_{\mathcal{C}}$ then by Theorem 8.5, this is equivalent to the existence of $B = B(g)^{\mathcal{C}}_{\mathcal{B}}$ such that $AB = I = BA$. □

8.1.2 Determinant

In order to prepare the ground for lattices, we need an axiomatic treatment of the notion of determinant, which is the content of this section.

Definition 8.7. A **determinant** is a function $\det : \mathrm{Mat}_{n\times n}(\mathbb{F}) \cong (\mathbb{F}^n)^n \longrightarrow \mathbb{F}$ such that, for a matrix C with column vectors $\{C_1, ..., C_n\} \subseteq \mathbb{F}^n$ we have:

1. (multi-linearity) For each $1 \le i \le n$ the map $\mathbb{F}^n \longrightarrow \mathbb{F}$ given by $C_i \mapsto \det(C)$ is linear. That is, for a scalar b and two column vectors C_i and C'_i we have:
    - $\det(..., bC_i, ...) = b \det(..., C_i, ...)$.
    - $\det(..., C_i + C'_i, ...) = \det(..., C_i, ...) + \det(..., C'_i, ...)$.
2. (alternating) if two columns of a matrix are equal, the determinant is 0.
3. (normalisation) the determinant of the identity matrix is 1: $\det(I) = 1$.

The remainder of this section is devoted to show that there is precisely one function satisfying Definition 8.7. We will do so by showing, via small claims, what is implied by the definition (that is, if such a function exists, what are its additional properties) and then derive an explicit definition.

Let us start with

Lemma 8.8. *If two columns of a matrix are interchanged, the value of the determinant is multiplied by -1. That is, writing the determinant as a function of columns*

$$\det(C_1, ..., C_n),$$

we have for $i < j$:

$$\det(..., C_i, ..., C_j, ...) = (-1) \cdot \det(..., C_j, ..., C_i, ...)$$

Proof. We have

$$\begin{aligned} 0 &= \det(..., C_i + C_j, ..., C_i + C_j, ...) \\ &= \det(..., C_i, ...C_i) + \det(..., C_i, ..., C_j, ...) + \det(..., C_j, ..., C_i, ...) + \det(..., C_j, ..., C_j, ...) \\ &= \det(..., C_i, ..., C_j, ...) + \det(..., C_j, ..., C_i, ...) \end{aligned} \tag{66}$$

where the first equality is from alternating , the second equality is from multilinearity and the third equality is again from alternating. □

Remark 8.9. The property of Lemma 8.8 is called skew-symmetry. If $\mathrm{char}\,\mathbb{F} \neq 2$, the alternating property is equivalent to skew-symmetry.

For the next result, recall that S_n is the group of permutations on letters $\{1,...,n\}$ and that any permutation $\pi \in S_n$ can be decomposed as a product of transpositions $(i,j) \in S_n$ (although not uniquely). The **sign** of a permutation π is the $\sigma(\pi) = (-1)^m$ where m is the number of transpositions in some decomposition of π.

Corollary 8.10. *For any permutation $\pi \in S_n$ we have*

$$\det(C_{\pi(1)},...,C_{\pi(n)}) = \sigma(\pi)\det(C_1,...,C_n).$$

Proof. Decompose π into a product of transpositions and use Definition 4.92 with application of 8.8. □

We also have

Lemma 8.11. *The value of* det *is unchanged if a multiple of one column is added to another. That is, for a scalar $b \in \mathbb{F}$:*

$$\det(...,C_i,...,C_j,...) = \det(...,C_i,...,C_j + bC_i,...)$$

Proof. Using linearity in the jth column, we have

$$\begin{aligned} &\det(...,C_i,...,C_j+bC_i,...)\\ &= \det(...,C_i,...,C_j,...) + b\det(...,C_i,...,C_i,...)\\ &= \det(...,C_i,...,C_j,...). \end{aligned} \tag{67}$$

□

A key property is as follows:

Proposition 8.12. *Let $C_j = \sum_i b_{ij}A_i$ where $b_{ij} \in \mathbb{F}$ and A_i is column vector in $\mathbb{F}^n$. Let C be a matrix with ith column C_i and A the matrix with ith column A_i. Then*

$$\begin{aligned} \det(C) &= \left(\sum_{\pi\in S_n} \operatorname{sign}(\pi) b_{\pi(1),1}\cdot ...\cdot b_{\pi(n),n}\right)\det(A)\\ &= \left(\sum_{\pi\in S_n} \operatorname{sign}(\pi) b_{1,\pi^{-1}(1)}\cdot ...\cdot b_{n,\pi^{-1}(n)}\right)\det(A) \end{aligned} \tag{68}$$

Proof. Expanding, using multilinearity, we have:

$$\det(...,C_j,...) = \det(...,\sum_i b_{ij}A_i,...) = \sum_{i_1,...,i_n} b_{i_1,1}\cdot ...\cdot b_{i_n,n}\det(A_{i_1},...,A_{i_n})$$

where $i_1,...,i_n$ range over all n-tuples with entries from $\{1,...,n\}$. If any two of them, say i_p and i_q for $p \neq q$ satisfy $i_p = i_q$ then the determinant is 0 by alternating property so we may as well sum over all permutations π of the ordered tuple $(1,...,n)$. Letting π be the permutation that takes ℓ to i_ℓ, we have

$$\det(A_{i_1},...,A_{i_n}) = \operatorname{sign}(\pi)\det(A_1,...,A_n).$$

We thus get

$$\det(C) = \left(\sum_{\pi\in S_n} \operatorname{sign}(\pi) b_{\pi(1),1}\cdot ...\cdot b_{\pi(n),n}\right)\det(A)$$

as desired. For the last equality, observe that since multiplication in $\mathbb{F}$ is commutative and any permutation $\pi \in S_n$ is invertible,

$$b_{\pi(1),1} \cdot ... \cdot b_{\pi(n),n} = b_{1,\pi^{-1}(1)} \cdot ... \cdot b_{n,\pi^{-1}(n)}$$

so

$$\sum_{\pi \in S_n} \operatorname{sign}(\pi) b_{\pi(1),1} \cdot ... \cdot b_{\pi(n),n} = \sum_{\pi \in S_n} \operatorname{sign}(\pi) b_{1,\pi^{-1}(1)} \cdot ... \cdot b_{n,\pi^{-1}(n)}$$

□

Corollary 8.13. *If a determinant function of Definition 8.7, it is unique.*

Proof. Proposition 8.12 gives an explicit formula for a determinant function so any two such functions must coincide. □

Definition 8.14. For $A = (a_{ij}) \in \operatorname{Mat}_{n\times n}(\mathbb{F})$, the **transpose** of A is the matrix $A^T \in \operatorname{Mat}_{n\times n}(\mathbb{F})$ whose (i,j)th entry is a_{ji}. In other words, ith column of A^T is the ith row of A.

Corollary 8.15. *for any square matrix C,*

$$\det(C^T) = \det(C).$$

Proof. This follows immediately form the equality

$$\sum_{\pi \in S_n} \operatorname{sign}(\pi) c_{\pi(1),1} \cdot ... \cdot c_{\pi(n),n} = \sum_{\pi \in S_n} \operatorname{sign}(\pi) c_{1,\pi(1)} \cdot ... \cdot c_{n,\pi(n)}$$

□

Corollary 8.16. *For any two square matrices A, B, the product $C = AB$ satisfies*

$$\det(AB) = \det(A)\det(B).$$

Proof. The jth column C_j of C is the linear combination

$$C_j = A_1 b_{1j} + ... + A_n b_{nj}$$

of the columns $A_1, ..., A_n$ of A. Thus, by Proposition 8.12 we get

$$\det(AB) = \det(C) = \left(\sum_{\pi \in S_n} \operatorname{sign}(\pi) b_{\pi(1),1} \cdot ... \cdot b_{\pi(n),n} \right) \det(A) \tag{69}$$

and we know that the sum is $\det(B)$. □

The formula of proposition 8.12 does not give an immediate proof for existence of the determinant function since its not clear that formula satisfies the axioms of Definition 8.7. In fact, to show that directly would take a considerable computation. Instead, we will circumvent this by giving an alternative, more convenient, definition of determinant and show that it too satisfies the axioms.

Definition 8.17. Let $A \in \operatorname{Mat}_{n\times n}(\mathbb{F})$ be a square matrix and $1 \le k \le n$. For $i, j \in [n] \times [n]$, denote by $A_{ij} \in \operatorname{Mat}_{(n-1)\times(n-1)}(\mathbb{F})$ the matrix obtained from A by removing the ith row and the jth column. When $n = 1$, ie $A = (a)$ we set $\det A = a$. For a fixed i, the ith row **expansion by minors** of A (or Laplace expansion) is defined inductively as

$$\det(A) = (-1)^{i+1} a_{i1} \det(A_{i1}) - a_{i2} \det(A_{i2}) + ... + (-1)^{i+n} a_{in} \det(A_{in}) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij}.$$

Theorem 8.18. *The function* $\det : \mathrm{Mat}_{n\times n}(\mathbb{F}) \longrightarrow \mathbb{F}$ *satisfies the axioms of definition 8.7.*

Proof. We prove the case $i = 1$. The other cases are similar. Let us denote the right hand side of the above equation by $f(A_1, A_2, ..., A_n)$. We show that $f(A_1, A_2, ..., A_n)$ is a determinant function by induction on n. It is easily checked for $n = 1$ and $n = 2$. Suppose that the columns A_j and A_{j+1} of A are equal. Then A_{i1} have equal columns except when $i = j$ or $i = j + 1$. By induction, $\det A_{i1} = 0$ for $i \neq j, j + 1$. Thus,

$$\det(A) = a_{1j}[(-1)^{1+j}\det(A_{1j})] + a_{1,j+1}[(-1)^{2+j}\det(A_{1,j+1})].$$

Since we have equality of columns $A_j = A_{j+1}$, we clearly have $a_{1j} = a_{1j+1}$ and $A_{1j} = A_{1j+1}$ so that $\det(A) = 0$. Thus det is alternating. If $e_1, ..., e_n$ is the standard basis for $\mathbb{F}^n$ then by induction

$$\det(A) = 1 \cdot \det(A_{11}) = \det(e_1, ..., e_{n-1}) = 1.$$

For multilinearity, suppose we have a column form $A = [A_1 \cdots A_i + A_i' \cdots A_n]$ and argue by induction on n.

$$\det(A) = (-1)^{1+1}a_{11}\det(A_{11}) + ... + (-1)^{1+i}a_{1i}\det(A_{1i}) + ... + (-1)^{1+n}a_{1n}\det(A_{1n})$$

The ith summand equals $[(A_i)_1 + (A_i')_1]\det(A_{1i})$ and the other summands are determinants of $(n-1)\times(n-1)$ matrices whose ith column is a sum of two columns (obtained from the columns A_i, A_i' by removing an element). By the induction hypothesis these determinants are multilinear hence are equal to a sum of the respective $(n-1)\times(n-1)$ determinants. We thus obtain

$$\det(..., A_i + A_i', ...) = \det(..., A_i, ...) + \det(..., A_i', ...).$$

as desired. We leave the case of multiplication by scalar for the reader. □

Corollary 8.19. *There exists a unique function* $\det : \mathrm{Mat}_{n\times n}(\mathbb{F}) \longrightarrow \mathbb{F}$ *satisfying the properties of Definition 8.7.*

Proof. We have shown uniqueness before. The existence follows from Theorem 8.18 since the expansion by minors is an explicit formula that satisfies the axioms of Definition 8.7. □

Notation 8.20. To shorten notation, we will sometime use $|A|$ to denote $\det A$.

Corollary 8.21. *For* $A \in \mathrm{Mat}_n(\mathbb{F})$ *and* $1 \leq j \leq n$ *we have*

$$\det A = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}$$

Proof. This formula is the ith expansion by minors for A^T so by Corollary 8.15 it equals $\det A$. □

Remark 8.22. The formula above is called the jth column expansion by minors of A.

When minor expansions are taken along the wrong row or column they nullify:

Proposition 8.23. *Let* $A \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ *and fix* $1 \leq i \neq k \leq n$. *Then*

$$\sum_{j=1}^{n} a_{ij}(-1)^{k+j}\det(A_{kj}) = 0$$

Proof. Let B the matrix obtained from A by replacing the kth row with the ith row. Then $\det B = 0$ as it has two equal rows. Now,

$$\sum_{j=1}^{n} a_{ij}(-1)^{k+j}\det(A_{kj}) = \sum_{j=1}^{n} b_{ij}(-1)^{k+j}\det(A_{kj}) = \sum_{j=1}^{n} b_{kj}(-1)^{k+j}\det(B_{kj}) = \det B = 0$$

where the first equality is because the ith row of A equals the ith row of B and the second equality is because in B the ith row equals to the kth row. □

The historical origins of matrices in linear algebra is as a mean to solve a system of linear equations. Suppose we have a system of n linear equations in n variables over a field $\mathbb{F}$.

$$\begin{aligned} a_{11}x_1 + \dots + a_{1n}x_n &= b_1 \\ \vdots \qquad\qquad &\quad \vdots \\ a_{n1}x_1 + \dots + a_{nn}x_n &= b_n \end{aligned} \tag{70}$$

We may write it in matrix form as

$$Ax = b$$

where $A = (a_{ij}) \in \mathrm{Mat}_{n,n}$, $b = (b_1, \dots, b_n)$ and $x = (x_1, \dots, x_n)$ are column vectors.

Theorem 8.24 (Cramer's rule)**.** *If $\det(A) \neq 0$, there is a unique solution $(\overline{x}_1, \dots, \overline{x}_n)$ to 70 given by:*

$$\overline{x}_i = (\det A)^{-1}\det(A \overset{i}{\leftsquigarrow} b)$$

where

$$A \overset{i}{\leftsquigarrow} b$$

is the matrix obtained from A by replacing the ith column with b.

Proof. Using expansion by minors we get

$$x_i = \det(I \overset{i}{\leftsquigarrow} x)$$

(since the determinant of a triangular matrix is the product of its diagonal entries.) Observe that $A(I \overset{i}{\leftsquigarrow} x) = A \overset{i}{\leftsquigarrow} b$. Thus,

$$(\det A)\det(I \overset{i}{\leftsquigarrow} x) = \det(A \overset{i}{\leftsquigarrow} b)$$

ie

$$(\det A)x_i = \det(A \overset{i}{\leftsquigarrow} b)$$

and since $\det A \neq 0$ we get the desired. □

Corollary 8.25. *A square matrix A is invertible iff $\det A \neq 0$.*

Proof. If A is invertible then $A \cdot A^{-1} = I$ so $\det(A)\det(A^{-1}) = 1$ hence $\det A \neq 0$. Conversely, consider the linear map $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by $f(X) = AX$ (where X is viewed as a column vector) – note that A is the matrix representing f wrt the standard basis $e_1, \dots, e_n$. By Theorem 8.24 for any $B \in \mathbb{R}^n$ (viewed as a column vector) the system $AX = B$ has a unique solution. This means that f is bijective and hence an isomorphism. We can thus take the matrix B representing f^{-1} wrt the standard basis and by Theorem 8.5 $AB = I$ so A is invertible with $B = A^{-1}$. □

Next, we would like to get a formula for the inverse of A (when it exists), at least for theoretical use.

Definition 8.26. Let $A \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ be a square matrix. Let $M(A) \in \mathbb{R}^{n\times n}$ be the matrix whose ij-entry is the minor $(-1)^{i+j}\det A_{ij} \equiv (-1)^{i+j}|A_{ij}|$ (that is, the determinant of A when removing the ith row and jth column multiplied by a sign $(-1)^{i+j}$). The **adjoint** matrix of A is the matrix $\mathrm{adj}\, A := M(A)^T$

Proposition 8.27. *Let $A \in \mathrm{Mat}_{n\times n}$ be an invertible matrix. Then $A^{-1} = \frac{1}{\det A}\mathrm{adj}\, A$.*

Proof. Write

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

and

$$M(A) = \begin{bmatrix} (-1)^{1+1}|A_{11}| & (-1)^{1+2}|A_{12}| & \cdots & (-1)^{1+n}|A_{1n}| \\ (-1)^{2+1}|A_{21}| & (-1)^{2+2}|A_{22}| & \cdots & (-1)^{2+n}|A_{2n}| \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{n+1}|A_{n1}| & (-1)^{n+2}|A_{n2}| & \cdots & (-1)^{n+n}|A_{nn}| \end{bmatrix}$$

The product of A with the transpose of $M(A)$ (denoted as $M(A)^T$) is $A \cdot M^T =$

$$\begin{bmatrix} a_{11}(-1)^{1+1}|A_{11}| + a_{12}(-1)^{1+2}|A_{21}| + \cdots + a_{1n}(-1)^{1+n}|A_{n1}| & a_{11}(-1)^{2+1}|A_{11}| + a_{12}(-1)^{2+2}|A_{21}| + \cdots + a_{1n}(-1)^{2+n}|A_{n2}| & \cdots & a_{11}(-1)^{n+1}|A_{11}| + a_{12}(-1)^{n+2}|A_{21}| + \cdots + a_{1n}(-1)^{n+n}|A_{nn}| \\ a_{21}(-1)^{1+1}|A_{11}| + a_{22}(-1)^{1+2}|A_{21}| + \cdots + a_{2n}(-1)^{1+n}|A_{n1}| & a_{21}(-1)^{2+1}|A_{11}| + a_{22}(-1)^{2+2}|A_{21}| + \cdots + a_{2n}(-1)^{2+n}|A_{n2}| & \cdots & a_{21}(-1)^{n+1}|A_{11}| + a_{22}(-1)^{n+2}|A_{21}| + \cdots + a_{2n}(-1)^{n+n}|A_{nn}| \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}(-1)^{1+1}|A_{11}| + a_{n2}(-1)^{1+2}|A_{21}| + \cdots + a_{nn}(-1)^{1+n}|A_{n1}| & a_{n1}(-1)^{2+1}|A_{11}| + a_{n2}(-1)^{2+2}|A_{21}| + \cdots + a_{nn}(-1)^{2+n}|A_{n2}| & \cdots & a_{n1}(-1)^{n+1}|A_{11}| + a_{n2}(-1)^{n+2}|A_{21}| + \cdots + a_{nn}(-1)^{n+n}|A_{nn}| \end{bmatrix}$$

The ij entry in $A \cdot M(A)^T$ is an expansion by minors of A (wrt to a column i) if $i = j$, hence is equal to $\det(A)$. If $i \neq j$, it is the expansion by minors of A along the "wrong" row as in 8.23 (convince yourself for $ij = 12$ for example), hence equals 0. We thus get $AM(A)^T = \mathrm{diag}(\det(A))$ so $A \cdot \frac{1}{\det A}\mathrm{adj}\, A = I$ as desired. □

8.1.3 Inner products

Throughout this section, let $\mathbb{F} = \mathbb{R}$ and $V = \mathbb{R}^n$. If $a = (a_1, ..., a_n), b = (b_1, ..., b_n) \in V$, their **inner product** is defined as $\langle a, b\rangle \equiv a \cdot b = \sum_i a_i b_i \in \mathbb{R}$.

Exercise 8.28.

1. $\sqrt{\langle a, a\rangle} = \|a\| := \sqrt{\sum_i a_i^2}$
2. $\langle a, b\rangle = \langle b, a\rangle$
3. for $\lambda \in \mathbb{R}$ and $a \in \mathbb{R}^n$, $\|\lambda a\| = |\lambda|\|a\|$
4. The function $\langle \cdot, \cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ is bilnear.

Proposition 8.29 (Cauchy-Schwarz inequality)**.** *For $v, w \in \mathbb{R}^n$, $|v \cdot w| \leq \|v\|\|w\|$*

Proof. If either v or w is 0_V the claim is trivial so assume otherwise. Set $a = -\langle v, w\rangle$ and $b = \langle v, v\rangle = \|v\|^2$. Then

$$\begin{aligned} 0 &\leq \langle av + bw, av + bw\rangle \\ &= \langle av, av\rangle + \langle av, bw\rangle + \langle bw, av\rangle + \langle bw, bw\rangle \\ &= a^2\|v\|^2 + ab\langle v, w\rangle + ab\langle v, w\rangle + b^2\|w\|^2 \\ &= a^2 b - a^2 b - a^2 b + b^2\|w\|^2 = b[-a^2 + b\|w\|^2] \end{aligned} \tag{71}$$

Since $v \neq 0_V$, we know that $b > 0$ so that $0 \leq -a^2 + b\|w\|^2$ which means $|\langle v, w\rangle| \leq \|v\|\|w\|$ as desired. □

By Cauchy-Schwarz, for any $v, w \in \mathbb{R}^n$, $0 \le \frac{|v \cdot w|}{\|v\|\|w\|} \le 1$ and thus there exists $\theta \in [0, \pi]$ such that $|v \cdot w| = \|v\|\|w\| \cos\theta$ we call θ the **angle** between v and w.

Exercise 8.30. Take two vectors $v, w \in \mathbb{R}^2$ and show that the angle between them as defined above is the same as the geometric angle between them. Do the same for $\mathbb{R}^3$

Corollary 8.31 (Triangle inequality)**.** *For any $v, w \in \mathbb{R}^n$,*

$$\|v + w\| \le \|v\| + \|w\|.$$

Proof.

$$\begin{aligned} \|v + w\|^2 &= \langle v + w, v + w\rangle \\ &= \langle v, v\rangle + \langle v, w\rangle + \langle w, v\rangle + \langle w, w\rangle \\ &= \|v\|^2 + 2 \cdot \langle v, w\rangle + \|w\|^2 \\ &\le \|v\|^2 + 2 \cdot \|v\|\|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2. \end{aligned} \tag{72}$$

using the Cauchy-Schwartz inequality in the last step. Taking square root gives the desired. □

Observe that the standard basis $e_1, ..., e_n \in \mathbb{R}^n$ satisfies $\langle e_i, e_j\rangle = 0$ for any $i \ne j$. That is, every pair of distinct vectors in it is **orthogonal**, ie have an angle of 90 degrees. Since the standard basis has proven to be useful, we are interested in finding other bases of $\mathbb{R}^n$ whose vectors are mutually orthogonal

Definition 8.32. A basis $\mathcal{B} \subseteq \mathbb{R}^n$ is called **orthogonal** if any two vectors $v, w \in \mathcal{B}$ satisfy $\langle v, w\rangle = 0$.

Theorem 8.33 (Gram-Schmidt)**.** *Every subspace $V \subseteq \mathbb{R}^d$ admits an orthogonal basis.*

Proof. We argue by induction on $\dim V \le d$. If $\dim V = 1$, there is nothing to prove. Assume the result is true for all subspaces of dimension k and let V be a subspace of dimension $k + 1$. Then V is spanned by $k + 1$ linearly independent vectors hence admits a subspace W of dimension k (take the span of, eg, the first k vectors in that basis). By the induction hypothesis, W admits an orthogonal basis $v_1, ..., v_k$. Choose $v \in V \setminus W$. For each $1 \le i \le k$ let $c_i := \langle v, v_i\rangle / \langle v_i, v_i\rangle \in \mathbb{R}$ and set $v_{k+1} = v - \sum_{i=1}^k c_i v_i$. Then $v_{k+1} \notin W$ since $v \notin W$. Moreover, for each $1 \le j \le k$, we have

$$\begin{aligned} \langle v_{k+1}, v_j\rangle &= \langle v, v_j\rangle - \sum_{i=1}^k c_i \langle v_i, v_j\rangle \\ &= \langle v, v_j\rangle - c_j \langle v_j, v_j\rangle \\ &= \langle v, v_j\rangle - \langle v, v_j\rangle / \langle v_j, v_j\rangle \langle v_j, v_j\rangle \\ &= 0. \end{aligned} \tag{73}$$

Thus, $\{v_1, ..., v_{k+1}\} \subseteq V$ is a set of mutually orthogonal vectors and it is enough to show it is linearly independent. Suppose there are scalars $a_1, ..., a_{k+1}$ such that

$$\sum_{i=1}^{k+1} a_i v_i = 0$$

. Then for each $1 \le \ell \le k + 1$ we have

$$\begin{aligned} c_\ell \langle v_\ell, v_\ell\rangle &= \sum_{i=1}^{k+1} a_i \langle v_i, v_\ell\rangle \\ &= \langle \sum_{i=1}^{k+1} c_i v_i, v_\ell\rangle = \langle 0, v_\ell\rangle = 0 \end{aligned} \tag{74}$$

and since $\langle v_\ell, v_\ell\rangle \neq 0$ it follows that $c_\ell = 0$. Thus, $\{v_1, ..., v_{k+1}\}$ are also linearly independent, hence form an orthogonal basis. □

The standard basis $e_1, ..., e_n \in \mathbb{R}^n$ admits another useful property: each vector e_i is **normal** in that $\|e_i\| = 1$. Such a basis is called **orthonormal**.

Observation 8.34. The proof of Gram-Schmidt theorem yields an algorithm that takes as input any basis $\mathcal{B} = \{v_1, ..., v_n\}$ for $V \subseteq \mathbb{R}^d$ and outputs an orthogoanl (resp. orthonormal) basis $\mathcal{B}^* = \{v_1^*, ..., v_n^*\}$ (resp $\mathcal{B}_n^*$): For $1 \leq k \leq n$ do:

1. $$v_k^* := v_k - \sum_{j=1}^{k-1} \langle v_k, v_j^*\rangle v_j^*.$$

2. (for orthonormality) $$v_k^* := v_k^* / \|v_k^*\|.$$

Note that for any $i < k$, $\langle v_k^*, v_i\rangle = 0$ as the proof of Gram-Schmidt theorem shows.

We now want to extend the notion of linear independence. Suppose that W_1 and W_2 are subspaces of a vector space V over a field $\mathbb{F}$. Denote $W_1 + W_2 = \{w_1 + w_2 | w_1 \in W_1 \wedge w_2 \in W_2\}$ and note that this is another subspace of V. We know that any vector $v \in W_1 + W_2$ can be written in the form $w_1 + w_2$, where $w_1 \in W_1$ and $w_2 \in W_2$. This representation of v is not unique in general. It will be unique if $\mathcal{W}_1 \cap W_2 = \{0_V\}$ for then, if $w_1 + w_2 = w_1' + w_2'$for $w_1, w_1' \in W_1$ and $w_2, w_2' \in W_2$ we have $w_1 - w_1' = w_2' - w_2 \in W_1 \cap W_2 = \{0_V\}$ and so $w_1 = w_1'$ and $w_2 = w_2'$. This consideration suggests the need to emphasize those situations in which the intersection of two subspaces of a given vector space is the trivial subspace. Indeed, if W_1 and W_2 satisfy $W_1 \cap W_2 = \{0_V\}$ we will write $W_1 \oplus W_2$ instead of $W_1 + W_2$ and say that the set $\{W_1, W_2\}$ is **independent**. The subspace $W_1 \oplus W_2$ of V is called the **direct sum** of W_1 and W_2.

Definition 8.35. Let $U \subseteq \mathbb{R}^n$ be a subspace. The **orthogonal complement** of U is defined as

$$U^\perp = \{v \in \mathbb{R}^n | v \perp u, \ \forall u \in U\}.$$

Proposition 8.36. *Let $U \subset V = \mathbb{R}^n$ be any subspace. Then there exists a direct sum decomposition $V = U \oplus U^\perp$*

Proof. Clearly $U \cap U^\perp = \{0_V\}$ so it remains to show that $U + U^\perp = V$. Use Observation 8.34 to get an orthonormal basis $\{u_1, ..., u_k\}$ for U. For $v \in V$, let

$$P(v) \equiv P_U(v) := \sum_{j=1}^{k} \langle v, u_j\rangle u_j.$$

Write $v = P(v) + (v - P(v))$. Clearly, $P(v) \in U$ so it remains to show that $v - P(v) \in U^\perp$. If $1 \leq j \leq k$ then

$$\begin{aligned}
&\langle v - P(v), u_j\rangle \\
&= \langle v - \sum_{\ell=1}^{k} \langle v, u_\ell\rangle u_\ell, u_j\rangle \\
&= \langle v, u_j\rangle - \sum_{\ell=1}^{k} \langle\langle v, u_\ell\rangle u_\ell, u_j\rangle \\
&= \langle v, u_j\rangle - \sum_{\ell=1}^{k} (\langle v, u_\ell\rangle\langle u_\ell, u_j\rangle) \\
&= \langle v, u_j\rangle - \langle v, u_j\rangle = 0.
\end{aligned} \tag{75}$$

by noting that $\langle u_\ell, u_j \rangle = 0, \forall \ell \neq j$ and 1 otherwise by the orthogonality of $\{u_1, \ldots, u_k\}$.

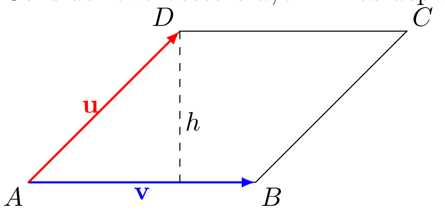
Since $v - P(v)$ is orthogonal to every basis vector of U, $v - P(v) \in U^\perp$. □

Definition 8.37. The map $P_U : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is called the **orthogonal projection** on U.

8.1.4 Determinant as a volume function

We now turn to a geometric interpretation of det that is important both from a conceptual reason and for the material on lattices.

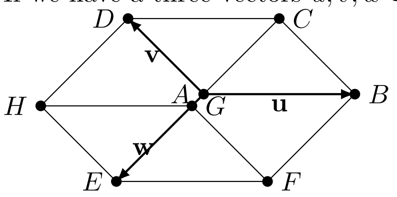
Consider two vectors $u, v \in \mathbb{R}^2$ as depicted below.



To find the area of the parallelogram $\mathcal{P}$ spanned by u and v we may proceed as follows: first calculate the length of v ie $\|v\|$. Then write $u = P_{\mathrm{Span}(v)}(u) + h$ where $h \perp v$. Since $\mathbb{R}^2 = \mathrm{Span}(v) \oplus \mathrm{Span}(v)^\perp$, we know that this h is unique. Then

$$\mathrm{Vol}_2(P) = \|v\| \|h\|.$$

If we have a three vectors $u, v, w \in \mathbb{R}^3$ they span a parallelepiped $\mathcal{P}'$



whose volume can be computed as follows: we first compute the $2D$ volume (ie area) A of the parallelogram spanned by v, w then write $u = P_{\mathrm{Span}(v,w)}(u) + h'$ where h' is a vector orthogonal to $\mathrm{Span}(v, w)$, which is uniquely determined since $\mathbb{R}^3 = \mathrm{Span}(v, w) \oplus \mathrm{Span}(v, w)^\perp$ and then set $\mathrm{Vol}_3(\mathcal{P}') = A \cdot \|h'\|$.

How do we proceed to the n-dimensional case? Note first that in 2-dimensions, the parallelogram $\mathcal{P}$ spanned by $u, v \in \mathbb{R}^2$ can be defined as

$$\mathcal{P} = \{su + tv | s, t \in [0, 1]\}$$

Similarly, the parallelepiped $\mathcal{P}'$ spanned by $u, v, w \in \mathbb{R}^3$ is

$$\mathcal{P}' = \{ru + sv + tw | r, s, t \in [0, 1]\}$$

In the n-dimensional case we make the following

Definition 8.38. Let $v_1, ..., v_n \in \mathbb{R}^n$ be vectors. The n-dimensional **parallelepiped** spanned by the v_i's is

$$\mathcal{P}(v_1, ..., v_n) = \{\sum_{i=1}^{n} t_i v_i | \forall i : \ t_i \in [0, 1]\}.$$

To define the volume, let us make the following

Construction 8.39. For an n-dimensional parallelepiped $\mathcal{P} = \mathcal{P}(v_1, ..., v_n)$, set

1. $u_1 = v_1$
2. $u_2 = v_2 - P_{\text{Span}(u_1)}(v_2)$
3. $u_3 = v_3 - P_{\text{Span}(u_1,u_2)}(v_3)$
4. $u_n = v_n - P_{\text{Span}(u_1,...,u_{n-1})}(v_n)$

Observe that each u_i is orthogonal to all $u_1, ..., u_{i-1}$ so that u_i's are mutually orthogonal.

The following definition is then a straightforward generalisation of our discussion on areas and volumes:

Definition 8.40. The **volume** of an n-dimensional parallelepiped $\mathcal{P} = \mathcal{P}(v_1, ..., v_n)$ is defined by

$$\text{Vol}_n(\mathcal{P}) := \|u_1\| \cdot ... \cdot \|u_n\|.$$

Theorem 8.41. *For an n-dimensional parallelepiped* $\mathcal{P} = \mathcal{P}(v_1, ..., v_n)$, $\text{Vol}(\mathcal{P}) = |\det(v_1, ..., v_n)|$.

Proof. By construction, we have

$$\begin{aligned} \det(v_1, ..., v_n) = \\ \det(u_1, u_2 + P_{\text{Span}(u_1)}(v_2), ..., u_n + P_{\text{Span}(u_1,...,u_{n-1})}(v_n)) \\ = \det(u_1, ..., u_n) \end{aligned} \tag{76}$$

where the last equality follows from multilinearity and the fact that $P_{\text{Span}(u_1,...,u_{i-1})}(v_i)$ is a linear combination of $u_1, ..., u_{i-1}$ so that putting it in the ith column makes the determinant vanish. Let U be the matrix whos columns are the u_i's. Since the u_i's are mutually orthogonal,

$$U^T U = \text{diag}(\|u_1\|^2, ..., \|u_n\|^2)$$

is a diagonal matrix whose diagonal is $\|u_1\|^2, ..., \|u_n\|^2$. We can see this by observing $(U^T U)_{ij} = \langle u_i, u_j \rangle = 0$ if $i \neq j$ by their orthonormality, otherwise it is $\|u_i\|^2$. Thus,

$$\det(U)^2 = \det(U^T U) = \|u_1\|^2 \cdot ... \cdot \|u_n\|^2$$

and it follows that

$$|\det(v_1, ..., v_n)| = |\det(U)| = \|u_1\| \cdot ... \cdot \|u_n\| = \text{Vol}(\mathcal{P})$$

□

8.1.5 Lattices

Intuitively, a lattice is similar to a vector space except that it consists of discrete vectors only, that is, elements in lattice vectors have discrete values as opposed to real-valued vectors in a vector space.

Definition 8.42. Let $\mathcal{B} = \{v_1, ..., v_n\} \in \mathbb{R}^d$ be a set of linearly independent vectors. The **lattice generated** by $\mathcal{B}$ is the set of $\mathbb{Z}$-linear combinations of $v_1, ..., v_n$. That is,

$$L \equiv L(\mathcal{B}) = \text{Span}_{\mathbb{Z}}(\mathcal{B}) = \{a_1 v_1 + \; + a_n v_n | a_1, ..., a_n \in \mathbb{Z}\}.$$

The integers d and n are the **dimension** and **rank** of the lattice respectively. If $d = n$, then L is a **full-rank** lattice. A **basis** for the lattice $L = L(\mathcal{B})$ is a set $\mathcal{B}' \subseteq \mathbb{R}^d$ of linearly independent vectors such that $\text{Span}_{\mathbb{Z}}(\mathcal{B}') = L(\mathcal{B})$. In particular, $\mathfrak{B}$ is a basis for $L(\mathfrak{B})$. We will denote by B the matrix obtained from $\mathfrak{B}$ viewed as an array of column vectors.

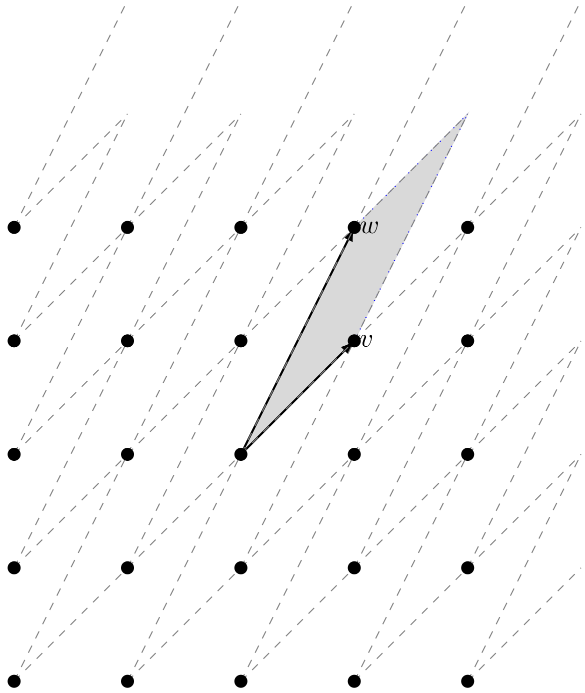


Figure 3: The lattice in $\mathbb{R}^2$ spanned by vectors v and w. The dashed parallelograms are called **tiles**. The area in gray is called the **fundamental domain**.

Remark 8.43.

1. Henceforth, by default all lattices are assumed to be full-rank lattices, ie $d = n$ unless stated otherwise.
2. Observe that since the vectors in $\mathcal{B}$ are linearly independent in $\mathbb{R}^n$, for a lattice point $v \in L(\mathcal{B})$ with $v = \sum_i a_i v_i$, the integers $a_1, ..., a_n$ are unique.
3. A basis $\mathcal{B}$ for a lattice is always a basis for $\mathbb{R}^n$ but the converse is not true: $2\mathcal{B}$ is a basis for $\mathbb{R}^n$ but not for $L(\mathcal{B})$ since all $\mathbb{Z}$-linear combinations of $\mathcal{B}$ do not contain all vectors in L.

Exercise 8.44. Show that a lattice is an abelian group wrt the addition operation (and zero vector) in $\mathbb{R}^n$ and that it is always isomorphic to $\mathbb{Z}^n$ (as an abelian group).

Example 8.45. The lattice spanned by v, w is shown in Figure 8.45

As one can expect, a basis for a lattice L is not unique, so we would like to know how to transition from one basis to another.

Definition 8.46. A matrix $U \in \mathbb{Z}^{n\times n}$ is **unimodular** if it has an inverse in $\mathbb{Z}^{n\times n}$ ie there exists $V \in \mathbb{Z}^{n\times n}$ such that

$$UV = VU = I.$$

Proposition 8.47.

1. *If U is unimodular, so is U^{-1}.*

2. *If U, V are unimodular, so is UV.*

3. *$U \in \mathbb{Z}^{n\times n}$ is unimodular iff $\det U = \pm 1$.*

Proof.

1. trivial.

2. $(UV)^{-1} = V^{-1}U^{-1}$ and the latter is in $\mathbb{Z}^{n\times n}$ since $U^{-1}, V^{-1} \in \mathbb{Z}^{n\times n}$.

3. if U is unimodular, then $\det(UU^{-1}) = \det(I) = 1$ and since $\det(U) \in \mathbb{Z}$ we must have $\det U = \pm 1$. Conversely, if $U \in \mathbb{Z}^{n\times n}$ with $\det U = \pm 1$ then $\det U \neq 0$ so U is invertible and $\det(U^{-1}) = \det(U)^{-1} = \pm 1$. By Proposition 8.27, $U^{-1} = \frac{1}{\det U}\operatorname{adj}(U)$. Since $U \in \mathbb{Z}^{n\times n}$, so is $M(U)$ and since $\det U = \pm 1$, $U^{-1} \in \mathbb{Z}^{n\times n}$

□

Theorem 8.48. *Let $\mathcal{B}, \mathcal{C} \subseteq \mathbb{R}^n$ be two bases with corresponding matrices B, C. Then $L(\mathcal{B}) = L(\mathcal{C})$ iff there exists a unimodular matrix U such that $B = CU$.*

Proof. Suppose $B = CU$ for some unimodular U. Then U is invertible so $BU^{-1} = C$ and since $U, U^{-1} \in \mathbb{Z}^{n\times n}$ we get $L(\mathcal{B}) \subseteq L(C)$ and $L(\mathcal{C}) \subseteq L(\mathcal{B})$ ie $L(\mathcal{B}) = L(\mathcal{C})$. Conversely, is $L(\mathcal{B}) = L(\mathcal{C})$ then by definition of a lattice, there are two square matrices $U, V \in \mathbb{Z}^{n\times n}$ such that $B = CU$ and $C = BV$. Combining these equations we get $C(I - UV) = 0$. Since C is invertible we get $I - UV = 0$ so that $V = U^{-1}$. □

A simple way to obtain a basis of a lattice from another is to apply (a sequence of) elementary column operations, as defined below.

Definition 8.49. Elementary (integer) column operations on a matrix $B \in \mathbb{R}^{d\times n}$ are:

1. Swap (i,j): $(bi, bj) \to (bj, bi)$. (Exchange two basis vectors)

2. Invert (i): $b_i \to -b_i$. (Change the sign of a basis vector)

3. Add (i,c,j) : $b_i \to (b_i + c\cdot b_j)$ where $i \neq j$ and $c \in \mathbb{Z}$. (Add an integer multiple of a basis vector to another).

It follows from Theorem 8.48 that elementary column operations do not change the lattice generated by the basis because they can be expressed as right multiplication by a unimodular matrix.

Exercise 8.50. Give unimodular matrices corresponding to the elementary column operations swap (i,j), invert (i) and add (c,i,j) for $c \in \mathbb{Z}$ and $i, j \in 1, ..., n, i \neq j$. Hint: what do you get when applying these operations on I? For each operation, prove that your matrix is indeed unimodular by giving the inverse matrix and showing that it has integer entries. Give also an English description of the operation specified by the inverse matrix.

Definition 8.51. Let $L(\mathcal{B}) \subseteq \mathbb{R}^n$ be a lattice spanned by $\mathcal{B} = \{v_1, ..., v_n\}$. The **fundamental domain** of L is

$$\mathcal{P}(\mathcal{B}) = \{\sum_{i=1}^{n} t_i v_i | \forall i,\ t_i \in [0,1]\}.$$

Although the fundamental domain depends on a choice of basis, its volume does not:

Proposition 8.52. *For two bases $\mathcal{B}, \mathcal{C}$ for a lattice $L \subseteq \mathbb{R}^n$, we have:*

$$\operatorname{Vol} \mathcal{P}(\mathcal{B}) = \operatorname{Vol} \mathcal{P}(\mathcal{C}).$$

Proof. By Theorem 8.41, $\operatorname{Vol} \mathcal{P}(\mathcal{B}) = |\det B|$ and $\operatorname{Vol} \mathcal{P}(\mathcal{C}) = |\det C|$. By Theorem 8.48, there exists a unimodular matrix U such that $B = CU$ so $\det B = \det C \cdot \det U$. Since $\det U = \pm 1$ we get $|\det B| = |\det C|$ as desired. □

Recall that the Gram-Schmidt theorem allowed us to take any basis $\mathcal{B}$ of $\mathbb{R}^n$ and turn it into orthogonal (or even orthonormal) basis. This does not work for bases of lattices as the following example shows

Example 8.53. The Gram-Schmidt orthogonalisation of the basis $B = [(2,0)^T, (1,2)^T] \subseteq \mathbb{R}^2$ is $B^* = [(2,0)^T, (0,2)^T]$. However this is not a lattice basis for $L(\mathcal{B})$ because the vector $(0,2)^T$ does not belong to the lattice. $L(\mathcal{B})$ contains a sublattice generated by a pair of orthogonal vectors $(2,0)^T$ and $(0,4)^T$, but no pair of orthogonal vectors generate the entire lattice $L(\mathcal{B})$.

The next result shows that although not every lattice has an orthogonal basis, every integer lattice contains an orthogonal sublattice.

Theorem 8.54. *For any invertible $B \in \mathbb{Z}^{n \times n}$, with $d = |\det B|$ we have $d \cdot \mathbb{Z}^n \subseteq L(\mathcal{B})$.*

Proof. Let v be a vector in $d \cdot \mathbb{Z}^n$ so that $v = dy$ for some integer vector $y \in \mathbb{Z}^n$. We want to prove that $v \in L(\mathcal{B})$. Since B is invertible, there is a unique solution to the equation

$$Bx = dy$$

for some $x \in \mathbb{R}^n$. It is enough to show that $x \in \mathbb{Z}^n$ since then $dy \in L(\mathcal{B})$. By Cramer's rule (8.24),

$$\begin{aligned} x_i &= \frac{\det(B \overset{i}{\leftsquigarrow} dy)}{\det B} \\ &= \frac{d \det(B \overset{i}{\leftsquigarrow} y)}{\det B} \\ &= \pm \det(B \overset{i}{\leftsquigarrow} y) \in \mathbb{Z} \end{aligned} \tag{77}$$

Since $B \in \mathbb{Z}^{n \times n}$ and $y \in \mathbb{Z}^n$ we get $\det(B \overset{i}{\leftsquigarrow} y) \in \mathbb{Z}$ so $x_i \in \mathbb{Z}$ as desired. □

8.1.6 Minimum distance

Recall that for a set $S \subseteq \mathbb{R}$, the **infimum** of S, $m = \inf S$ is the maximal number m such that $m \leq s$ for any $s \in S$ (we set $\inf S = -\infty$ if no such $m \in \mathbb{R}$ exists). If S is bounded from below, $\inf S$ exists and is finite.

Definition 8.55. For any lattice $\Lambda = L(\mathcal{B})$, the minimum distance of Λ is the smallest distance between any two lattice points:

$$\lambda(\Lambda) = \inf\{\|x - y\| | x, y \in \Lambda, x \neq y\}$$

We observe that the minimum distance can be equivalently defined as the length of the shortest nonzero lattice vector:

$$\lambda(\Lambda) = \inf\{\|v\| | v \in \Lambda \setminus \{0\}\}.$$

This follows from the fact that lattices are additive subgroups of $\mathbb{R}^n$, i.e., they are closed under addition and subtraction. So, if x and y are distinct lattice points, then $x - y$ is a nonzero lattice point.

The first thing we want to prove about the minimum distance is that it is always achieved by some lattice vector, i.e., there is a lattice vector $x \in \Lambda$ of length exactly $\|x\| = \lambda(\Lambda)$. To prove this, we need first to establish a lower bound on $\lambda(\Lambda)$.

8.1.7 Lebesgue measure on $\mathbb{R}^n$

Our goal is to construct a notion of the volume, or Lebesgue measure, of rather general subsets of $\mathbb{R}^n$ that reduces to the usual volume of elementary geometrical sets such as cubes or rectangles. If $\mathcal{L}(\mathbb{R}^n)$ denotes the collection of Lebesgue measurable sets and

$$\mu : \mathcal{L}(\mathbb{R}^n) \longrightarrow [0,\infty]$$

denotes Lebesgue measure, then we want $\mathcal{L}(\mathbb{R}^n)$ to contain all n-dimensional rect- angles and $\mu(R)$ should be the usual volume of a rectangle R. Moreover, we want μ to be countably additive. That is, if

$$\{A_i \in \mathcal{L}(\mathbb{R}^n) | i \in \mathbb{N}\}$$

is a countable collection of disjoint measurable sets, then their union should be measurable and

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

The reason for requiring countable additivity is that finite additivity is too weak a property to allow the justification of any limiting processes, while uncountable additivity is too strong; for example, it would imply that if the measure of a set consisting of a single point is zero, then the measure of every subset of $\mathbb{R}^n$ would be zero.

Definition 8.56. An n-dimensional, closed rectangle with sides oriented parallel to the coordinate axes, or **rectangle** for short, is a subset $R \subseteq \mathbb{R}^n$ of the form

$$R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n]$$

where $-\infty < a_i < b_i < \infty$ for $i = 1, ..., n$. The volume $\mu(R)$ of R is

$$\mu(R) = (b_1 - a_1)(b_2 - a_2)...(b_n - a_n).$$

We denote the collection of all n-dimensional rectangles (including the empty rectangle) by $\mathcal{R}(\mathbb{R}^n)$ or $\mathcal{R}$ for short.

We thus get a function

$$\mu : \mathcal{R} \longrightarrow [0,\infty)$$

which we would like to extend.

To get started, for a set $E \subseteq \mathbb{R}^n$, we say that a collection of rectangles $\{R_i\}_{i\in\mathbb{N}}$ is a **cover** of E if $E \subseteq \bigcup_i R_i$. If $\mathcal{C} = \{R_i\}_{i\in\mathbb{N}}$ is a rectangle cover of E, we get a sequence of non-negative numbers $\{\mu^*(R_i)\}_{i\in\mathbb{N}}$ and hence a monotonically increasing sequence of partial sums $\{s_n\}_{n\in\mathbb{N}}$ where $s_n := \sum_{i=1}^n \mu^*(R_i)$. We set $\sum_{i=1}^\infty \mu^*(R_i) := \infty$ if $\{s_n\}_n$ is not bounded. If $\{s_n\}_n$ is bounded, with supremum $\sigma = \sup\{s_n\}_n$, we set $\sum_{i=1}^\infty \mu^*(R_i) := \sigma$. An easy fact from calculus (supremum is the limit of a monotonically increasing sequence that is bounded from above) is that $\sum_{i=1}^\infty \mu^*(R_i) = \lim_{n\to\infty} s_n \equiv \lim_{n\to\infty} \sum_{i=1}^n \mu^*(R_i)$.

Example 8.57. Let $R_i = [0,1] \times [0, \frac{1}{2^i}]$ so that $\mu(R_i) = 1/2^i$. Then

$$\sum_{i=1}^{\infty} \mu(R_i) = \sum_{i=1}^{\infty} 1/2^i = \lim_{n\to\infty} \sum_{i=1}^{n} 1/2^i = \lim_{n\to\infty} (1 - 1/2^n) = 1.$$

Definition 8.58. The **outer Lebesgue measure** $\mu^*(E)$ of a subset $E \subseteq \mathbb{R}^n$ is

$$\mu^*(E) = \inf_{\mathcal{C}=\{R_i\}} \left\{ \sum_{i=1}^{\infty} \mu(R_i) | E \subseteq \bigcup_{i=1}^{\infty} R_i,\ R_i \in \mathcal{R} \right\}$$

where the infimum is taken over all covers $\mathcal{C}$ of E by rectangles.

The map $\mu^* : \mathcal{P}(\mathbb{R}^n) \longrightarrow [0,\infty]$ is the called the (Lebesgue) outer measure.

Example 8.59. Let $E = \mathbb{Q} \cap [0,1] \subseteq \mathbb{R}$ be the set of rational numbers between 0 and 1. Then E has outer measure zero. To prove this, let $\{q_i | i \in \mathbb{N}\}$ be an enumeration of the points in E. Given $\epsilon > 0$, let R_i be an interval of length $\epsilon/2^i$ which contains q_i. Then

$$E \subseteq \bigcup_{i=1}^{\infty} \mu(R_i)$$

so

$$0 \le \mu^*(E) \le \sum_{i=1}^{\infty} \mu(R_i) = \epsilon$$

(here we used that $\sum_{i=1}^{\infty} 1/2^i = 1$: this is because $\sum_{i=1}^{n} 1/2^i = 1 - 1/2^n$ so the second term vanishes when n tends to infinity).

Hence $\mu^*(E) = 0$ since $\epsilon > 0$ is arbitrary. The same argument shows that any countable set has outer measure zero. Note that if we cover E by a finite collection of intervals, then the union of the intervals would have to contain $[0,1]$ (we say that E is dense in $[0,1]$) so their lengths sum to at least one.

Example 8.60. For $n = 2$, consider E as the non-negative x-axis $E = \mathbb{R}_{\ge 0} \times \{0\} \subseteq \mathbb{R}^2$ in the plane. For a fixed $\epsilon > 0$, construct a cover $\{R_i\}_{i=1}^{\infty}$ of E by setting

$$R_i = [i-1, i] \times [-\frac{\epsilon}{2^i}, \frac{\epsilon}{2^i}].$$

Then

$$\sum_{i=1}^{\infty} = \mu^*(R_i) = \sum_{i=1}^{\infty} \frac{2\epsilon}{2^i} = 2\epsilon$$

and since μ^* is an infimum over all such sums, $\mu^*(E) \le 2\epsilon$. Since ϵ can be arbitrarily small, $\mu^*(E) = 0$.

Our next goal is to show that the outer measure of a rectangle coincides with its volume as defined above. We begin with some combinatorial facts about finite covers of rectangles. We denote the interior of a rectangle R by R°, and we say that rectangles R, S are almost disjoint if $R^\circ \cap S^\circ = \varnothing$, meaning that they intersect at most along their boundaries. The proofs of the following results are cumbersome to write out in detail (it's easier to draw a picture) but we briefly explain the argument.

Lemma 8.61. *Suppose that*

$$R = I_1 \times \ldots \times I_n$$

is an n-dimensional rectangle, and each closed, bounded interval $I_i \subseteq \mathbb{R}$ is an almost disjoint union of closed, bounded intervals

$$\{I_{i,j} \subseteq \mathbb{R} | j = 1, \ldots, N_i\}, \quad I_i = \bigcup_{j=1}^{N_i} I_{i,j}.$$

Define the rectangles

$$S_{j_1 j_2 \ldots j_n} = I_{1,j1} \times I_{2,j2} \times \ldots \times I_{n,jn}.$$

Then

$$\mu(R) = \sum_{j_1=1}^{N_1} \ldots \sum_{j_n=1}^{N_n} \mu(S_{j_1 j_2 \ldots j_n})$$

Proof. Denoting the lengeth of an interval I by $|I|$ and using the fact that

$$|I_i| = \sum_{j=1}^{N_i} |I_{i,j}|$$

we get

$$\mu(R) = |I_1|...|I_n| = \left(\sum_{j_1=1}^{N_1} |I_{1,j_1}|\right)...\left(\sum_{j_n=1}^{N_n} |I_{n,j_n}|\right) = \sum_{j_1=1}^{N_1}...\sum_{j_n=1}^{N_n} |I_{1,j_1}|...|I_{n,j_n}| = \sum_{j_1=1}^{N_1}...\sum_{j_n=1}^{N_n} \mu(S_{j_1j_2...j_n}). \tag{78}$$

□

Proposition 8.62. *If a rectangle R is an almost disjoint, finite union of rectangles $\{R_1, R_2, ..., R_N\}$, then*

$$\mu(R) = \sum_{i=1}^{N} \mu(R_i).$$

If R is covered by rectangles $\{R_1, R_2, ..., R_N\}$, which need not be disjoint, then

$$\mu(R) \le \sum_{i=1}^{N} \mu(R_i).$$

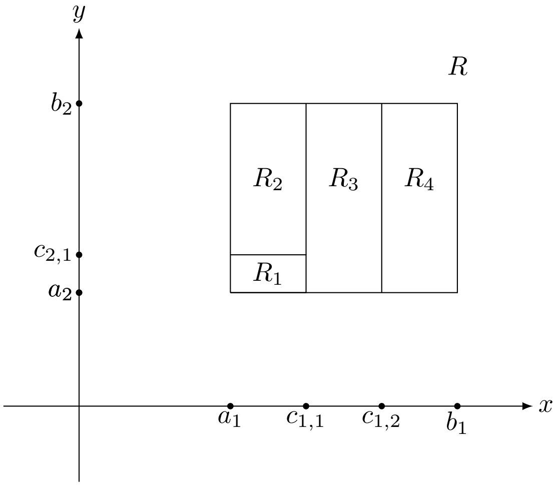
Proof. Suppose that

$$R = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n].$$

we partition $[a_i, b_i]$ into

$$a_i = c_{i,0} \le c_{i,1} \le ... \le c_{i,N_i} = b_i, \quad I_{i,j} = [c_{i,j-1}, c_{i,j}].$$

where the $c_{i,j}$ are obtained by ordering the left and right ith coordinates of all intervals that appear in some rectangle in the collection $\{R_1, R_2, ..., R_N\}$, and define rectangles $S_{j_1j_2...j_n} = \prod_{i=1}^{n} I_{i,j_i}$ as Lemma 8.61. The drawing below is an illustration of an $n = 2$ case.



Clearly, $\mu(R) = \sum_{j_1=1}^{N_1} ... \sum_{j_n=1}^{N_n} \mu(S_{j_1j_2...j_n})$. On the other hand, each R_i is an almost disjoint union of some of the $S_{j_1j_2...j_n}$'s, each appearing only once in a unique R_i (since the R_i's are almost disjoint). Thus, by Lemma 8.61 also

$$\sum_{i=1}^{N} \mu(R_i) = \sum_{j_1=1}^{N_1} ... \sum_{j_n=1}^{N_n} \mu(S_{j_1j_2...j_n})$$

as desired.

Suppose $\{R_1, ..., R_n\}$ is a collection of rectangles that covers R. Then there is an almost disjoint, finite collection of rectangles $\{S_1, S_2, ..., S_M\}$ such that

$$R = \bigcup_{i=1}^{M} S_i, \quad \sum_{i=1}^{M} \mu(S_i) \leq \sum_{i=1}^{N} \mu(R_i).$$

To obtain the S_i, we replace R_i by the rectangle $R \cap R_i$, and then decompose these possibly non-disjoint rectangles into an almost disjoint, finite collection of sub-rectangles with the same union; we discard 'overlaps' which can only reduce the sum of the volumes. Then, we get

$$\mu(R) = \sum_{i=1}^{M} \mu(S_i) \leq \sum_{i=1}^{N} \mu(R_i)$$

as desired. □

The outer measure of a rectangle is defined in terms of countable covers. We want to reduce these to finite covers by using the topological properties of $\mathbb{R}^n$

Fact 8.63. *If a countable collection of open rectangles $\{S_i^\circ\}_{i \in \mathbb{N}}$ covers a rectangle $R \subseteq \mathbb{R}^n$ then there exists a finite subcollection $\{S_{j_1}^\circ, ..., S_{j_n}^\circ\}$ that also covers R.*

With this in mind, we are ready for

Proposition 8.64. *Let R be an n-dimensional rectangle. Then $\mu^*(R) = \mu(R)$.*

Proof. Since $\{R\}$ covers R, we have $\mu^*(R) \leq \mu(R)$, so we only need to prove the reverse inequality.

Suppose that $\{R_i | i \in \mathbb{N}\}$ is a countably infinite collection of rectangles that covers R. By enlarging R_i slightly we may obtain a rectangle S_i whose interior S_i° contains R_i such that

$$\mu(S_i) \leq \mu(R_i) + \epsilon/2^i.$$

Then $\{S_i^\circ | i \in \mathbb{N}\}$ is a countable open cover of R so there is a finite subcollection that covers R. By relabaling, we may denote this subcollection as $\{S_1^\circ, ..., S_N^\circ\}$. It follows that also $\{S_1, .., S_N\}$ covers R so we may use the second part of Proposition 8.61 to get

$$\mu(R) \leq \sum_{i=1}^{N} \mu(S_i) \leq \sum_{i=1}^{N} (\mu(R_i) + \epsilon/2^i) \leq \sum_{i=1}^{\infty} \mu(R_i) + \epsilon.$$

Since ϵ is arbitrary we get

$$\mu(R) \leq \sum_{i=1}^{\infty} \mu(R_i)$$

so that

$$\mu(R) \leq \mu^*(R).$$

□

Definition 8.65. A subset $A \subseteq \mathbb{R}^n$ is **Lebesgue measurable** if for any $E \subseteq \mathbb{R}^n$,

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

We denote the collection of Lebesgue measurable sets in $\mathbb{R}^n$ by $\mathcal{L}(\mathbb{R}^n)$.

Thus, a measurable set A splits any set E into disjoint pieces whose outer measures add up to the outer measure of E.

Remark 8.66. Since μ^* is subadditive, we always have

$$\mu^*(E) \le \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Thus, to prove that $A \subseteq \mathbb{R}^n$ is measurable, it is sufficient to show that for every $E \subseteq \mathbb{R}^n$,

$$\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

Exercise 8.67. Show that the complement of a measurable set is again measurable and the intersection of two measurable sets is again measurable

Theorem 8.68. *The restriction of Lebesgue outer measure μ^* to $\mathcal{L}(\mathbb{R}^n)$ is a measure on $\mathcal{L}(\mathbb{R}^n)$.*

Proof. It follows immediately from Definition 8.65 that $\varnothing$ is measurable.

It remains to prove that μ^* is countably additive on $\mathcal{L}(\mathbb{R}^n)$.

First, we prove that the union of measurable sets is measurable. Suppose that $A, B \in \mathcal{L}(\mathbb{R}^n)$ and $E \subseteq \mathbb{R}^n$. The measurability of A and B implies that

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c).$$

Since $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$ and μ^* is subadditive, we have

$$\mu^*(E \cap (A \cup B)) \le \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B).$$

The use of this inequality and the relation $A^c \cap B^c = (A \cup B)^c$ in the first inequality implies that

$$\mu^*(E) \ge \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$$

so $A \cup B$ is measurable.

Moreover, if A is measurable and $A \cap B = \varnothing$, then by taking $E = A \cup B$ in Definition 8.65, we see that

$$\mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

Thus, the outer measure of the union of disjoint, measurable sets is the sum of their outer measures. The repeated application of this result implies that the finite union of measurable sets is measurable and μ^* is finitely additive on $\mathcal{L}(\mathbb{R}^n)$.

Next, we we want to show that the countable union of measurable sets is measurable. It is sufficient to consider disjoint unions. To see this, note that if

$$\{A_i \in \mathcal{L}(\mathbb{R}_n) | i \in \mathbb{N}\}.$$

is a countably infinite collection of measurable sets, then

$$B_j := \bigcup_{i=1}^{j} A_i, \text{ for } j \ge 1$$

form an increasing sequence of measurable sets, and

$$C_j := B_j \setminus B_{j-1}, \text{ for } j \ge 2, \ C_1 = B_1$$

form a disjoint measurable collection of sets. Moreover

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{j=1}^{\infty} C_j.$$

Suppose that

$$\{A_i \in \mathcal{L}(\mathbb{R}^n) | i \in \mathbb{N}\}$$

is a countably infinite, disjoint collection of measurable sets, and define

$$B_j := \bigcup_{i=1}^{j} A_i, \quad B = \bigcup_{i=1}^{\infty} A_i.$$

Let $E \subseteq \mathbb{R}^n$. Since A_j is measurable and $B_j = A_j \cup B_{j-1}$ is a disjoint union (for $j \geq 2$),

$$\begin{aligned} \mu^*(E \cap B_j) &= \mu^*(E \cap B_j \cap A_j) + \mu^*(E \cap B_j \cap A_j^c) \\ &= \mu^*(E \cap A_j) + \mu^*(E \cap B_{j-1}). \end{aligned} \tag{79}$$

Also $\mu^*(E \cap B_1) = \mu^*(E \cap A_1)$. It follows by induction that

$$\mu^*(E \cap B_j) = \sum_{i=1}^{j} \mu^*(E \cap A_i).$$

Since B_j is a finite union of measurable sets, it is measurable, so

$$\mu^*(E) = \mu^*(E \cap B_j) + \mu^*(E \cap B_j^c),$$

and since $B^c \subseteq B_j^c$, we have

$$\mu^*(E \cap B_j^c) \geq \mu^*(E \cap B^c).$$

It follows that

$$\mu^*(E) \geq \sum_{i=1}^{j} \mu^*(E \cap A_i) + \mu^*(E \cap B^c).$$

Taking the limit of this inequality as $j \to \infty$ and using the subadditivity of μ^*, we get

$$\begin{aligned} \mu^*(E) &\geq \sum_{i=1}^{\infty} \mu^*(E \cap A_i) + \mu^*(E \cap B^c) \\ &\geq \mu^*(\bigcup_{i=1}^{\infty} E \cap A_i) + \mu^*(E \cap B^c) \\ &\geq \mu^*(E \cap B) + \mu * (E \cap B^c) \\ &\geq \mu^*(E). \end{aligned} \tag{80}$$

Therefore, we must have equality in 80, which shows that $B = \bigcup_{i=1}^{\infty} A_i$ is measurable. Moreover,

$$\mu^*(\bigcup_{i=1}^{\infty} E \cap A_i) = \sum_{i=1}^{\infty} \mu^*(E \cap A_i),$$

so taking $E = \mathbb{R}^n$, we see that μ^* is countably additive on $\mathcal{L}(\mathbb{R}^n)$. □

An open rectangle R° is a union of an increasing sequence of closed rectangles whose volumes approach $\mu(R)$; for example

$$(a_1,b_1)\times(a_2,b_2)\times...\times(a_n,b_n)=\bigcup_{k=1}^{\infty}[a_1+1/k,b_1-1/k]\times[a_2+1/k,b_2-1/k]\times...\times[a_n+1/k,b_n-1/k].$$

Thus, R° is measurable and

$$\mu(R^\circ)=\mu(R).$$

Moreover if $\partial R = R\smallsetminus R^\circ$ denotes the boundary of R, then $\mu(\partial R)=\mu(R)-\mu(R^\circ)=0$.

Sets of measure zero play a particularly important role. First, we show that all sets with outer Lebesgue measure zero are Lebesgue measurable.

Proposition 8.69. *If $N\subseteq\mathbb{R}^n$ and $\mu^*(N)=0$, then N is Lebesgue measurable. If in addition $M\subset N$ then M is also measurable with $\mu(M)=0$*

Proof. If $N\subseteq\mathbb{R}^n$ has outer Lebesgue measure zero and $E\subseteq\mathbb{R}^n$, then $0\le\mu^*(E\cap N)\le\mu^*(N)=0$, so $\mu^*(E\cap N)=0$. Therefore, since $E\cap N^c\subseteq E$,

$$\mu^*(E)\ge\mu^*(E\cap N^c)=\mu^*(E\cap N)+\mu^*(E\cap N^c),$$

which shows that N is measurable. If N is a measurable set with $\mu(N)=0$ and $M\subseteq N$, then $\mu^*(M)=0$, since $\mu^*(M)\le\mu^*(N)=\mu(N)$. Therefore M is measurable and $\mu(M)=0$.

□

Example 8.70. Set $C_1=[0,1]$ and remove the middle third to obtain $C_2=[0,\frac{1}{3}]\cup[\frac{2}{3},1]$. In the same fashion, we can remove the middle third from both intervals in C_2 to obtain $C_3=[0,\frac{1}{9}]\cup[\frac{2}{9},\frac{1}{3}]\cup[\frac{2}{3},\frac{7}{9}]\cup[\frac{8}{9},1]$ and so on. Thus, C_n has 2^n intervals, each of length $\frac{1}{3^n}$. The **Cantor set** is defined to be $C:=\bigcap_{n=1}^{\infty}C_n$.

Clearly, $\mu^*(C)=0$ as for any $\epsilon>0$, it can be covered by a set of disjoint rectangles whose total measure is smaller than ϵ. Thus C is measurable by Proposition 8.69.

In trenary representation, the numbers in C are all numbers of the form $0.x_1x_2...$ where $x_i\ne1$ for all $i\ge1$: in trenary, $1/3=.1=.022...$, $2/3=.2$ and $1=.22...$ so $C_1=[.0,.022...]\cup[.2,.22...]$ consists of all trinary representations whose first digit is different than 1. Similarly, C_n consists of all trinary representations whose nth digit is different than 1.

Thus, C is uncountable.

A key property of Lebesgue measure is translational invariance

Proposition 8.71. *For $A\subseteq\mathbb{R}^n$ and $h\in\mathbb{R}^n$,*

$$\mu^*(A+h)=\mu^*(A)$$

and A is measurable if and only if $A+h$ is measurable.

Proof. Clearly, for every rectangle R, $\mu(R)=\mu(R+h)$ and $\{R_i\}_{i\in\mathbb{N}}$ is a cover of A by rectangle iff $\{R_i+h\}_{i\in\mathbb{N}}$ is a cover of $A+h$ by rectangles. Thus, $\mu^*(A)=\mu^*(A+h)$.

Caratheadory's definition of measurability is translation invriant since for any $E\subseteq\mathbb{R}^n$,

$$(E+h)\cap(A+h)=(E\cap A)+h.$$

Explicitly, if A is measurable then for any E,

$$\mu^*(E)=\mu^*(E\cap A)+\mu^*(E\cap A^c).$$

But then for any E

$$\begin{aligned}\mu^*((E+h)\cap(A+h)) + \mu^*((E+h)\cap(A+h)^c) &= \mu^*(E\cap A+h) + \mu^*((E+h)\cup(A+h)) \\ &= \mu^*(E\cap A) + \mu^*(E\cup A+h) = \mu^*(E\cap A) + \mu^*(E\cap A^c) = \mu^*(E) = \mu^*(E+h).\end{aligned} \tag{81}$$

Since any set $E' \subseteq \mathbb{R}^n$ can be written as $E+h$ for some E (take $E'-h$), we get that $A+h$ is measurable. Similarly if $A+h$ is measurable, then so is A.

□

8.2 Minkowski's theorem

We now turn to estimating the value of λ from above. Clearly, for any basis $\mathcal{B}$, we have $\lambda(L(\mathcal{B})) < \min_i \|b_i\|$, because each column of $\mathcal{B}$ is a nonzero lattice vector. We would like to get a better bound, and, specifically, a bound that does not depend on the choice of the basis. Clearly, lattices with arbitrarily large minimum distance can be easily obtained simply by scaling an arbitrary lattice by a constant $c > 0$ to obtain $\lambda(c\cdot\Lambda) = c\cdot\lambda(\Lambda)$. What if we normalize the lattice so that $\det(\Lambda) = 1$? By definition of determinant, these are lattices with density 1, i.e., with about one lattice point per each unit volume of space. Can the lattice still have arbitrarily large minimum distance? Equivalently, we are asking if it is possible to bound the ratio $\lambda(\Lambda)/\det(\Lambda)^{1/n}$ for any n-dimensional lattice Λ. (Notice that the quantity $\lambda(\Lambda)/\det(\Lambda)^{1/n}$ is invariant under linear scaling because $\det(c\Lambda) = c^n \det(\Lambda)$.) For historical reasons, we defin and studied the square of this quantity, which is called Hermite's constant.

Definition 8.72. The **Hermite** constant of an n-dimensional lattice Λ is the quantity

$$\gamma(\Lambda) = (\lambda(\Lambda)/\det(\Lambda)^{1/n})^2.$$

The Hermite constant in dimension n is the supremum

$$\gamma_n = \sup_{\Lambda} \gamma(\Lambda)$$

, where Λ ranges over all n-dimensional lattices.

9 Appendix

9.1 Limits

Definition 9.1. Let $\mathbb{R}$ be the ordered field $(\mathbb{R}, \leq)$, then we can define the *supremum* and *infinum* for a subset $S \subseteq \mathbb{R}$ as follows:

1. $\sup(S) \geq x$ for all $x \in S$ is the *least upper bound.*

2. $\inf(S) \leq x$ for all $x \in S$ is the *greatest lower bound.*

When S is not finitely bounded above, we denote $\sup(S) = \infty$. Likewise when it's not bounded below, we denote $\inf(S) = -\infty$.

Definition 9.2. A sequence $\{a_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}$ is said to converge to a limit $\ell \in \mathbb{R}$ if $\forall \epsilon > 0$ there exists an $\mathcal{N} \in \mathbb{N}$ such that

$$|a_n - \ell| < \epsilon \quad \forall n > \mathcal{N}$$

and we say that $\{a_n\}_{n \in \mathbb{N}}$ tends to ∞ if for all $M > 0$, there exists $\mathcal{N}$ with $a_n > M$ for all $n > \mathcal{N}$.

Example 9.3.

$$a_n = (-1)^n \cdot n \not\to \infty$$

this goes up and down, so this does not converge to infinity, whereas

$$a_n = 2^n \to \infty$$

tends to infinity, and likewise $a_n = -2^n \to -\infty$.

Definition 9.4. A function $f : X \to Y$ is called *monotonically increasing* if $a \leq b \implies f(a) \leq f(b) \quad \forall a, b \in X$

Proposition 9.5. *If $\{a_n\}_{n=1}^{\infty}$ is monotonically increasing and bounded from above, then there exists $\ell \in \mathbb{R}$ such that*

$$\lim_{n \to \infty} a_n = \ell$$

Proof. Let $\ell = \sup_{n \in \mathbb{N}} \{a_n\}$. We claim that $\lim_{n \to \infty} a_n = \ell$.

Let $\epsilon > 0$, then $\ell + \epsilon > a_n$ for all n, and also there exists $\mathcal{N} \in \mathbb{N}$ such that

$$a_{\mathcal{N}} > \ell - \epsilon$$

because ℓ is a least upper bound, so subtracting any ϵ means it is no longer the upper bound.

Since $\{a_n\}_{n \in \mathbb{N}}$ is monotonically increasing for all $n \in \mathbb{N}$, then $a_n \geq a_{\mathcal{N}} > \ell - \epsilon$. Therefore there is an $\mathcal{N}$ such that $|a_n - \ell| < \epsilon$ for all n. Thus

$$\ell = \lim_{n \to \infty} a_n$$

□

References

[BK] Balasubramanian, R. and Koblitz, N., 1998. The improbability that an elliptic curve has subexponential discrete log problem under the Menezes—Okamoto—Vanstone algorithm. Journal of cryptology, 11(2), pp.141-145.

[FR] Fine, B. and Rosenberger, G., 1997. The fundamental theorem of algebra. Springer Science & Business Media.

[FST] Freeman, D., Scott, M. and Teske, E., 2010. A taxonomy of pairing-friendly elliptic curves. Journal of cryptology, 23, pp.224-280.

[Gol] Golan, J.S., 2013. Foundations of linear algebra (Vol. 11). Springer Science & Business Media.

[Lyn] Lynn, B., 2007. On the implementation of pairing-based cryptosystems (Doctoral dissertation, Stanford University).

[Mil] Miller, V.S., 2004. The Weil pairing, and its efficient calculation. Journal of cryptology, 17(4), pp.235-261.

[Sil] Silverman, J.H., 2009. The arithmetic of elliptic curves (Vol. 106, pp. xx+-513). New York: Springer.

[Toth] Tóth, L., 2013. Subgroups of finite abelian groups having rank two via Goursat's lemma. arXiv preprint arXiv:1312.1485.

[Wash] Washington, L.C., 2008. elliptic curves: number theory and cryptography. Chapman and Hall/CRC.

[HSV] Hess, F., Smart, N.P. and Vercauteren, F., 2006. The eta pairing revisited. IEEE transactions on information theory, 52(10), pp.4595-4602.

List of symbols

$E(\mathbb{k})$	points on an Elliptic curve over a field $\mathbb{k}$
$E/\mathbb{k}$	Elliptic curve defined over a field $\mathbb{k}$
$E[n]$	n torsion points
G, H	group
$\overline{\mathbb{k}}, \overline{\mathbb{F}}$	algebraic closure over a field
$\circ$	composition
$\varnothing$	empty set
$\mathbb{C}$	complex numbers
$\mathbb{F}, \mathbb{k}$	field
$\mathbb{F}[x]$, $\mathbb{k}[x]$	polynomials over a field with one indeterminate
$\mathbb{F}_{p^n}$	field with p^n elements
$\mathbb{N}$	natural numbers
$\mathbb{Q}$	rational numbers
$\mathbb{R}$	real numbers
$\mathbb{Z}_n$	additive group of integers modulo n
$\mathbb{Z}$	integers
$\mathcal{O}$	point at infinity

$\mu_n(\mathbb{F})$	nth roots of unity over a field $\mathbb{F}$
$\prod$	product over an indexed set
$\sim$	equivalence relation
$\sum$	sum over an indexed set
τ_n	Tate pairing
div	divisor
e_n	Weil pairing