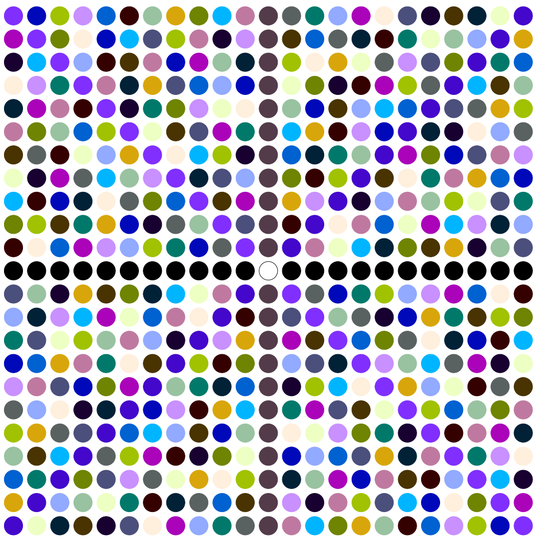
Algebra



Projective lines over the finite field $\mathbb{F}_{23}$ (see Exercice 5.4).

Rémi Géraud, 2017–2018

Introduction

Note: this is work in progress
Last compilation: 26th March 2018

This course is intended as an algebraic survival kit. Algebra is a rich, modern, and fascinating discipline; it has a reputation of being abstract, difficult. In some respects this is true.

It also has a reputation of being devoid of applications, being a mere exercise for the mind, an aesthetic pleasure and certainly not a basic engineering skill. And this is certainly not the case.

Therefore the technologically able person will necessarily face, sooner or later, the need to understand at least some algebra — be it to understand a communication device's resilience to noise, to protect a file using encryption, to work out the properties of some complicated geometric object, or simply to explore the wonderful developments of 20th and 21st century mathematics.

While certainly not complete or perfect, the aim of this booklet is to introduce the main concepts, tools, and techniques of algebra, hinting at more advanced topics as well as applications.

We expect the reader to have basic knowledge of numbers, integer and rational, of polynomials, and of Euclid's algorithm for the greatest common divisor. A familiarity with common proof techniques, such as proof by contradiction or by enumeration is advisable. The results proven in this book should hopefully help the readers to convince themselves, as well as acquire a few tricks of the trade; accordingly we avoid some very efficient, but equally obscure, shortcuts.

The course will be thematic, pointing at mountains (Galois, Grothendieck, ...) but following a gentle path. Exercises and problem sets provide additional elements. We certainly hope that the reader, equipped with this modest survival guide, with interest, patience, and determination, will find the path through these advanced topics easier.

To the reader who wonders: what *is* algebra? we can only offer a tentative answer. Algebra is a discipline that focuses on the manipulation of symbols according to arbitrary rules, and to an extent studies these rules for themselves (there are rules about rules, and so forth). This turns out to be a powerful approach to think about computing, to uncover far-reaching connections between seemingly different objects, and maybe to get a broad picture of mathematics itself.

Calendar (approximative)

1. Mathematical background
   a) Prime numbers, groups, rings, ideals, spectrum of $\mathbb{Z}$
   b) Quotients, polynomials, spectrum of $\mathbb{Z}[X]$
   c) Finite fields, Legendre–Jacobi symbol, quadratic reciprocity
   d) Diophantine equations, elliptic curves
2. Algorithmics in finite fields
   a) Euclidean algorithm, CRT, exponentiation, square roots
   b) Representation of finite fields, irreducibility tests
   c) Primality tests, smoothness
3. Applications
   a) Cryptography 1: DH key exchange, FS identification
   b) Cryptography 2: AJPS or NTRU encryption, RSA signatures
   c) Error-correcting codes: Hamming, BCH, Reed–Solomon

Note: When proofs are not given, it is usually because the reader is expected to be able to fill in the gaps. There will be very few exceptions to this rule, and they will always be explicitly indicated. Checking the provided examples is a good exercise.

Contents

Part I

Mathematical background

Chapter 1

Integers

In this chapter we set the bases of arithmetics with integers; we shall assume a familiarity with elementary operations and build upon this knowledge. The set of non-negative integers $\{0, 1, 2, \ldots\}$ is denoted $\mathbb{N}$, and it is equipped with an "addition" operation, a "multiplication" operation, and a total order $\leq$. The following fact will be used throughout this chapter:

Fact 1 *Every non-empty subset of* $\mathbb{N}$ *has a smallest element.*

And the following notion will be central to this course:

Definition 1 The set $\{\ldots, -1, 0, 1, \ldots\}$ of integers is denoted by $\mathbb{Z}$.

While relatively basic (or precisely for this reason), the reader is invited not to skip this part and to make sure that the provided examples and exercises pose no difficulty.

$$2^{77232917} - 1$$

Figure 1.1: The largest known Mersenne prime at the time of writing, found in early 2018. This number has more than 23 million decimal digits. See Exercise 1.5.

Contents

1.1 Integers and divisibility

Theorem 1 (Euclid) *Let $a, b \in \mathbb{Z}$, with $b \neq 0$. There exists a unique couple $(q, r) \in \mathbb{Z} \times \mathbb{Z}$ such that*

$$a = bq + r \qquad \text{and} \qquad 0 \leq r < |b|$$

where $|b| = \max(-b, b)$.

Definition 2 (Divisibility) If $a, b \in \mathbb{Z}$, we say that "a divides b" (and we write $a|b$) if there exists $q \in \mathbb{Z}$ such that $aq = b$.

We may equivalently say:

- b is a multiple of a;
- a is a divisor of b;
- the remainder r of the Euclidean division of b by a is zero;
- $b \in a\mathbb{Z} = \{\dots, -2a, -a, 0, a, 2a, \dots\}$;
- etc.

Definition 3 (Prime number) A number p whose only divisors are ± 1 and $\pm p$ is called a prime number.

Theorem 2 (Euclid) *The set $\mathfrak{P} = \{2, 3, 5, \dots\}$ of prime numbers is infinitely large.*

Problem 1 *Given an integer $n \in \mathbb{Z}$, how to efficiently decide whether $n \in \mathfrak{P}$?*

Definition 4 (Unit) The invertible elements of $\mathbb{Z}$, i.e., -1 and 1, are called units.

Definition 5 (Composite number) A number that is neither a unit nor a prime is called composite.

Theorem 3 (Euclid) *Every integer $n \in \mathbb{Z}$, $n \neq 0$, decomposes uniquely as*

$$n = u p_1^{n_1} p_2^{n_2} \cdots p_r^{n_r}$$

where u is a unit, $p_1 < \cdots < p_r$ are prime numbers, and $n_1, \dots, n_r > 0$ are integers.

Problem 2 *Given $\mathbb{N}$, how to efficiently find a divisor of n? A decomposition of n into its prime factors?*

Theorem 4 (Euclid) *Let $a, b \in \mathbb{Z}$ and $p \in \mathfrak{P}$. If $p|ab$ then $p|a$ or $p|b$.*

Proof: Assume $p \nmid a$, then define $A = \{n \geq 1 \text{ s.t. } p|an\}$. In particular $p \in A$ (and also $b \in A$) so that A is not empty, and $A \subset \mathbb{N}$, so by Fact 1 A has a smallest element; let this element be denoted by m. Since $p \nmid a$, $m > 1$.

Let $n \in A$, using Theorem 1 there exist integers q, r such that $n = mq + r$ with $0 \leq r < m$. Multiplying both sides by a, we have

$$(an) - (am)q = ar$$

whence $p|ar$ (indeed, $n \in A$ and $m \in A$). Since $r < m$, and m is by hypothesis the smallest element in A, $r \notin A$; therefore $r = 0$, which means that $m|n$.

As $p, b \in A$, the above applies to them in particular, i.e., $m|p$ and $m|b$. Since p is prime, and $m > 1$, the first equation gives $m = p$; therefore $p|b$. □

Corollary 1 (Largest prime factor) *Let $n \in \mathbb{N}$, $n \geq 2$. If $n \notin \mathfrak{P}$, then there exists $p \in \mathfrak{P}$ such that $p|n$ and $p^2 \leq n$.*

Algorithm 1 (Trial division algorithm) *There exists an algorithm that takes an integer $n > 2$ as input, and finds whether n is prime or composite (in which case it outputs a divisor of n) by performing at most $\sqrt{n}$ trial divisions.*

1.2 Greatest common divisor and p-adic valuation

Let $\overline{\mathbb{N}} = \mathbb{N} \cup \{+\infty\}$, where the order and operations on $\mathbb{N}$ are extended to work with the symbol $+\infty$ as follows: for all $n \in \mathbb{N}$, $+\infty \geq n$, $(+\infty) + n = n + (+\infty) = +\infty$, and $(+\infty) + (+\infty) = +\infty$.[1]

Definition 6 (p-adic valuation) Let $p \in \mathfrak{P}$. The p-adic valuation $v_p : \mathbb{Z} \to \overline{\mathbb{N}}$ is the map defined as follows:

- $v_p(0) = +\infty$
- $v_p(1) = 0$
- For all $n \geq 1$, $v_p(-n) = v_p(n)$
- For all $n \geq 2$, $v_p(n)$ is the exponent of p in the decomposition of n in prime factors.

Example 1 Let $n = 539000 = 2^3 \cdot 5^3 \cdot 7^2 \cdot 11$. Then

$$v_2(n) = 3, v_3(n) = 0, v_5(n) = 3, v_7(n) = 2, v_{11}(n) = 1, v_{13}(n) = v_{17}(n) = \cdots = 0.$$

Theorem 5 (Prime decomposition in $\mathbb{Z}$) *Every $n \in \mathbb{Z}$, $n \neq 0$, can be uniquely written as*

$$n = \epsilon \prod_{p \in \mathfrak{P}} p^{v_p(n)},$$

where ϵ is a unit. In particular, $v_p(n) \geq 1 \Leftrightarrow p|n$, and $a|b \Leftrightarrow v_p(a) \leq v_p(b)$ for all $p \in \mathfrak{P}$.

Proposition 1 (Properties of p-adic valuation) *Let $a, b \in \mathbb{Z}$, and $p \in \mathfrak{P}$. The following are easy properties of the p-adic valuation:*

- $v_p(ab) = v_p(a) + v_p(b)$
- $v_p(a + b) \geq \min(v_p(a), v_p(b))$ *(equality if $v_p(a) \neq v_p(b)$)*
- $a|b \Leftrightarrow \forall p \in \mathfrak{P}, v_p(a) \leq v_p(b)$.

Remark 1 The p-adic valuation allows one to construct the p-adic absolute value on $\mathbb{Q}$ as $|p/q|_p = p^{v_p(q) - v_p(p)}$. These norms have surprising properties (they are, in particular, ultrametric), and the completion of $\mathbb{Q}$ with $|\cdot|_p$ instead of the usual absolute value $|\cdot|$ gives the field of p-adic numbers $\mathbb{Q}_p$ instead of the field of real numbers $\mathbb{R}$. (We shall not say more at this point, as we have not introduced formally the notion of fields yet).

Theorem 6 (Gauss) *Let $a, b, c \in \mathbb{Z}$ such that a and b are coprime, and $a|bc$. Then $a|c$.*

[1] Note that multiplication by $+\infty$ is not defined.

Proof: Let $p \in \mathfrak{P}$, it suffices to show that $v_p(a) \le v_p(c)$. If $v_p(a) = 0$ this is immediate; let's therefore assume $v_p(a) > 0$, i.e., $p|a$. Since a and b are coprime, we have $v_p(b) = 0$. Finally, we know that $a|bc$ i.e. $v_p(a) \le v_p(bc) = v_p(b) + v_p(c) = v_p(c)$. □

Definition 7 (Greatest common divisor) Let $a, b \in \mathbb{N}$, $a, b, > 0$. The set $C = \{q \in \mathbb{N}$ s.t. $q|a$ and $q|b\} \subset \mathbb{N}$ is finite (by Corollary 1), therefore it has a greatest element; this element is called the greatest common divisor of a and b and denoted $\gcd(a, b)$. Equivalently,

$$\gcd(a, b) = \prod_{p \in \mathfrak{P}} p^{\min(v_p(a), v_p(b))}.$$

Definition 8 (Coprime integers) Let $a, b \in \mathbb{Z}$. We say that a and b are coprime when the following equivalent properties hold:

$$\gcd(a, b) = 1 \quad \Leftrightarrow \quad \forall p \in \mathfrak{P}, p \nmid a \text{ or } p \nmid b \quad \Leftrightarrow \quad \forall p \in \mathfrak{P}, \min\left(v_p(a), v_p(b)\right) = 0.$$

Lemma 1 *Let $a, b \in \mathbb{Z}$, $a, b \neq 0$, and $d = \gcd(a, b)$. Then $\gcd(a/d, b/d) = 1$.*

Lemma 2 *Let $a, b, c \in \mathbb{Z}$. If $\gcd(a, b) = 1$, $a|c$, and $b|c$, then $ab|c$.*

Theorem 7 (Bézout) *Let $a, b \in \mathbb{Z}$. There exists $u, v \in \mathbb{Z}$ such that $\gcd(a, b) = au + bv$.*

Proof: Let $A = \{au + bv \text{ s.t. } u, v \in \mathbb{Z}\} \cap (\mathbb{N} - \{0\})$. This is a non-empty subset of $\mathbb{N}$, therefore by Fact 1 it has a smallest element that we denote c.

By definition, $c \in A$. For every integer $k \ge 1$, $ck \in A$ as well. Let's show that these are in fact *all* the elements of A: let $n \in A$, by Theorem 1 there exist $q, r \in \mathbb{Z}$ such that $n = cq + r$, with $0 \le r < c$. Assuming $r \neq 0$, we have $r = n - cq \ge 1$; now since $c, n \in A$, $r = n - cq$ also belongs to A. But $r < c$ and c is by definition the smallest element of A, which leads to a contradiction: therefore, $r = 0$. As a consequence, $n = cq$, i.e., $A = \{ck \mid k \ge 1\}$.

If $ab \neq 0$, $|a| \in A$ and $|b| \in A$. By the above paragraph, this implies $c|a$ and $b|a$. Thus $c| \gcd(a, b)$ and since c is the smallest such number, $c = \gcd(a, b)$. At the same time, $c \in A$ means that there exists $u, v \in \mathbb{Z}$ such that $c = au + bv$. The remaining case $ab = 0$ is immediate. □

Corollary 2 (Bachet–Bézout theorem) *Let $a, b \in \mathbb{Z}$. Then a and b are coprime if and only if there exist integers u, v such that $1 = au + bv$.*

Algorithm 2 (Euclidean algorithm) *Let $a, b \in \mathbb{Z}$, $a \ge b$. The following algorithm computes $\gcd(a, b)$:*

1. *Let $r_0 = a$, $r_1 = b$*
2. *For $i \ge 1$, if $r_i \neq 0$, let r_{i+1} be the remainder of the Euclidean division of r_{i-1} by r_i.*
3. *For $n \ge 1$, if $r_n = 0$, the algorithm terminates and outputs $(r_0, \ldots, r_n)$.*

In particular, $r_n = \gcd(a, b)$.

Algorithm 3 (Extended Euclidean algorithm) *Let $a, b \in \mathbb{Z}$, $a \ge b$. Algorithm 2 can be extended to compute the integers u, v appearing in Theorem 7 (and therefore, also the gcd):*

1. *Let* $u_0 = v_0 = 0$ *and* $u_1 = v_1 = 1$
2. *For* $i \geq 1$, *let* q_i *be the quotient of the Euclidean division of* r_{i-1} *by* r_i. *Let*

$$u_{i+1} = u_{i-1} - u_i q_i \qquad \text{and} \qquad v_{i+1} = v_{i-1} - v_i q_i$$

Then $r_n = au_n + bv_n$.

Remark 2 The complexity of this algorithm can be estimated roughly using an easy result of Lamé: if Algorithm 3 terminates in n steps, then $a \geq \gcd(a,b)F_{n+2}$ and $b \geq \gcd(a,b)F_{n+1}$, where (F_k) is the Fibonacci sequence defined by

$$F_0 = 0, \qquad F_1 = 1, \qquad \forall n \geq 1, F_{n+1} = F_n + F_{n-1}.$$

The worst-case scenario consists of couples of the form $(a,b) = (F_{n+1}, F_n)$.

1.3 Base b numeration

Theorem 8 (Base change) *Let* $a \in \mathbb{Z}$ *and* $b \in \mathbb{N}$, $b \geq 2$, *then* a *can be written uniquely as*

$$a = \pm \sum_{k=0}^{A} b^k a_k,$$

where A *is the largest integer such that* $a_A \neq 0$. *The sequence* $(a_A, \ldots, a_0)$ *is called the base* b *representation of* a, *and* A *is called the* b*-ary length of* a.

Algorithm 4 (Integer exponentiation) *Let* $x \in \mathbb{Z}$ *and* $n \in \mathbb{N}$. *The naive computation of* x^n *performs* $n-1$ *multiplications; the following natural algorithm improves this by leveraging the binary (2-ary) representation of* n.

Write $n = 2^{i_0} + 2^{i_1} + \cdots + 2^{i_k}$, *with* $i_0 < i_1 < \cdots < i_k$, *and note that*

$$x^n = x^{\sum_{\ell=0}^{k} 2^{i_\ell}} = \prod_{\ell=0}^{k} x^{2^{i_\ell}}.$$

Now computing $x^{2^{i_\ell}}$ *can be done in* i_ℓ *squarings; and furthermore doing so provides* $x^{2^{i_m}}$ *for all* $m < \ell$. *As a result, to compute* x^n *we only perform* $i_k + k$ *multiplications.*

Since $2^{i_k} \leq n$, *i.e.,* $i_k \leq N$ *where* N *is the binary length of* n, *and* $k \leq i_k$, *the total number of multiplications performed is bounded by* $2N = 2\log_2(n)$.

1.4 Exercice set

Exercise 1.1. Prove Theorems 1 to 3.

Exercise 1.2. Let $a \in \mathbb{Z}$ and $n \in \mathbb{N}$ with $a, n \neq 0$. Prove that $(a-1)|(a^n-1)$.

Exercise 1.3. Find all $n \in \mathbb{N}$ such that $(n+1)|(n^2+1)$.

Exercise 1.4. Prove Corollary 1. Hint: use Theorem 4.

Exercise 1.5.

1. Show that if $2^n - 1 \in \mathfrak{P}$ then $n \in \mathfrak{P}$. (Such numbers are called Mersenne primes).

   Open question (2018): Are there infinitely many Mersenne primes?

   Show that $2^{11} - 1 \notin \mathfrak{P}$.

   Open question (2018): Are there infinitely many Mersenne non-primes?

2. Show that if $2^n + 1 \in \mathfrak{P}$, then n is a power of 2. (Such numbers are called Fermat primes).

   Open question (2018): Only 5 Fermat prime are known (3, 5, 7, 257, and 65537). Are there more? Are there infinitely many Fermat primes?

   Let $n = 2^{32} + 1$, we are going to show that $641|p$ (this result is originally due to Euler). Let $p = 641 = 5^4 + 2^4 = 5 \cdot 2^7 + 1$. Show that there exists $k \in \mathbb{Z}$ such that $(p-1)^4 = 1 + kp$ and conclude.

Exercise 1.6. Show that $1 + 2 + 2^2 + 2^3 + \cdots + 2^{26} \notin \mathfrak{P}$.

Exercise 1.7. Let $a, b \in \mathbb{Z}$. Show that $\prod_{p \in \mathfrak{P}} p^{\max(v_p(a), v_p(b))}$ is the least common multiple (lcm) of a and b.

Exercise 1.8. Let $p, q \in \mathfrak{P}$, $p \neq q$. Show that $pq \mid \left(p^{q-1} + q^{p-1} - 1\right)$.

Exercise 1.9. Let $a, b, c \in \mathbb{Z}$. Show that $\gcd(\gcd(a, b), c) = \gcd(a, \gcd(b, c))$.

Exercise 1.10. Let $a, b \in \mathbb{Z}$, $a \geq 2$, $b \geq 1$, show that

$$\gcd\left(\frac{a^b - 1}{a - 1}, a - 1\right) = \gcd(a - 1, b).$$

Exercise 1.11. Write 42 in bases 2, 3, 4, 8, and 16.

Exercise 1.12. Let $(a_A, \ldots, a_0)$ be the base-10 representation of $a \in \mathbb{N}$. Show that $11|a$ if and only if

$$\sum_{k=0}^{A} (-1)^k a_k = 0 \bmod 11.$$

Can you derive similar ways to check for divisibility by 3? by 7? by 13?

Exercise 1.13. Write a computer program that implements Algorithms 2 to 4.

Exercise 1.14. Let $d : \mathbb{Z} \to \mathbb{Z}$ be the operation defined as follows:

1. $d(p) = 1$ for all $p \in \mathfrak{P}$;
2. $d(ab) = d(a)b + ad(b)$ for any $a, b \in \mathbb{Z}$;
3. $d(-a) = -d(a)$ for any $a \in \mathbb{Z}$.

This operation is called the arithmetic derivative.

1. Prove that $d(0) = d(1) = 0$.
2. Let $a \in \mathbb{Z}$ and $n \in \mathbb{N}$, show that $d(a^n) = na^{n-1}d(a)$.
3. Show that $d(a) = 0$ if and only if $a \in \{0, 1\}$.
4. Let $a \in \mathbb{Z}$, show that
$$d(a) = \sum_{i=1}^{k} \frac{e_i}{p_i} a,$$
where $a = p_1^{e_1} \cdots p_k^{e_k}$ is the prime factorisation of a.
5. Write a program that computes the arithmetic derivative of an integer, and compute the derivatives of $a = 0, 1, 2, \ldots, 100$. This sequence begins with $0, 0, 1, 1, 4, 1, 5, \ldots$
6. Show that for any $a, b \in \mathbb{Z}$, $b \neq 0$,
$$d\left(\frac{a}{b}\right) = \frac{d(a)b - ad(b)}{b^2}.$$
7. Show that the logarithmetic derivative, defined by $L : a \mapsto d(a)/a$ is a totally additive function, i.e., for any $a, b \in \mathbb{Z}$, $a, b \notin \{0, 1\}$, $L(ab) = L(a) + L(b)$.

   Open question (2018): Let $D = d(\mathbb{Z})$ be the set of integers obtained by arithmetic derivation. Is there an even number $k \notin D$? This would disprove Goldbach's conjecture: '*Every even integer greated than 2 can be expressed as the sum of two primes*'.

   Open question (2018): Let $B = \{a \in \mathbb{Z} \mid d(d(a)) = 1\}$. Is B finite? This would disprove the twin prime conjecture: '*There are infinitely many twin primes, i.e., pairs of primes (p, q) s.t. $|p - q| = 2$*'.

Exercise 1.15. The product of prime numbers $p_1 = 2, \ldots, p_k$ is known as the k-th primorial, and is denoted $\#P_n$.

1. Write a program to compute $\#P_n$ for small values of n. The first values are $2, 6, 30, 210, 2310, \ldots$
2. Show that for all $n \leq p_n$, $\gcd(n, \#P_n + 1) = 1$. Deduce from this a constructive proof of the infinitude of prime numbers.
3. Show that $2^n \leq \#P_n \leq p_n^n$. We will later see that a consequence of the prime number theorem gives $p_n \approx n \ln n$. Compute a lower and an upper bound for $\#P_{75}$. In fact, $\#P_{75} \approx 2^{512}$.
4. The n-th fortunate number is the smallest positive integer $f_n > 1$ such that $\#P_n + f_n$ is prime. For instance, $\#P_7 = 2 \cdot 3 \cdot \cdots \cdot 17 = 510510$, and $f_7 = 19$ since 510529 is the first prime immediately after $\#P_7$.

5. Show that $f_n > p_n$, and in fact that if $d|f_n$, then $d > p_n$.
6. Write a program that computes f_n for small values of n. The first values are $3, 5, 7, 13, 23, 17, 19, 23, 37, \dots$

   Open question (2018): Is there any f_n that is not prime?

Exercice 1.16. The goal of this exercise is to show that $\sqrt{3} \notin \mathbb{Q}$ in two different ways.

1. Assume that $\sqrt{3} \in \mathbb{Q}$, and show that this implies a solution to the equation $3a^2 = b^2$ with non-zero integers a, b. Show such an equation has no solution by computing $v_3 \bmod 2$ on each side. (This is essentially Euclid's method).
2. Let E be the set
$$E = \left\{ b \in \mathbb{N} - \{0\} \mid \exists a \in \mathbb{N}, \sqrt{3} = \frac{a}{b} \right\}.$$
   a) Show that if $\sqrt{3} = a/b$, then $\sqrt{3} = a'/b'$ with
$$\begin{aligned} a' &= 3b - a \\ b' &= a - b \end{aligned}$$
   b) Show that $a' > 0$ and $b' < b$
   c) Show that $b' > 0$ and $a' < a$
   d) Assuming that E is not empty, it has a smallest element b_0. Using the above three points, derive a contradiction.

   (This is essentially Fermat's method)

Chapter 2

Groups

The notion of group is ubiquitous, as it captures a very common structure in mathematics. In particular, integers form a group.

The discipline of group theory is very large however, so we will not explore everything. In fact, we will focus our attention to finite groups, because these are the ones that arise the most in applications. Nevertheless, important examples of infinite groups are discussed.

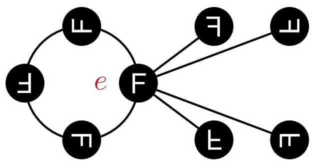


Figure 2.1: Cycle graph for the dihedral group D_4 (see Exercise 2.17).

Contents

2.1 Basic definitions

Definition 9 (Group) A non-empty set G together with an operation $G \times G \to G$, $(x, y) \mapsto x \cdot y$, and such that

- $\forall x, y, z \in G$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ (Associativity)
- $\exists e \in G, \forall x \in G, e \cdot x = x \cdot e = x$ (Existence of a neutral element)
- $\forall x \in G, \exists y \in G, x \cdot y = y \cdot x = e$ (Existence of all inverses)

is called a group. If furthermore, for all $x, y \in G$, $x \cdot y = y \cdot x$, then G is said to be a commutative or Abelian group.

Remark 3 In the above definition, we use the "multiplicative" notation for group composition. The inverse of an element $x \in G$ is usually denoted x^{-1}. Exponentiation by a positive integer n is denoted $x^n = x \cdot x \cdot \cdots \cdot x$, and extended with $x^0 = e$ and $x^{-n} = (x^{-1})^n$.

An alternative is the "additive" notation, where composition of x and y is written $x + y$, and the inverse of x is denoted $-x$. Exponentiation by a positive integer n is denoted nx or $[n]x = x + x + \cdots + x$ ($n > 0$ terms), and extended with $0x = e$ and $[-n]x = -[n]x$.

Definition 10 (Group morphism) Let G and H be groups. An application $f : G \to H$ is a group morphism if f preserves the group structure, i.e.,

- $f(e_G) = e_H$;
- $\forall x \in G, f(x)^{-1} = f\left(x^{-1}\right)$;
- $\forall x, y \in G, f(xy) = f(x)f(y)$.

Remark 4 The collection of groups, together with group morphisms, constitutes the category of groups, Grp. Similarly one can define the category of Abelian groups.

Definition 11 (Order) If G is finite, its order is the number of its elements (its cardinal).

Example 2

- The trivial group $0 = \{e\}$, with composition law $e \cdot e = e$. This is the smallest of all groups.[1] Its order is 1.
- The set $\mathbb{Z}$ (resp. $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$) with composition law the usual addition, with neutral element 0, is an Abelian group called the additive group of integers (resp. rational, real, complex numbers). It has infinite order.
- The set $\mathbb{N}$, endowed with the usual addition law, is not a group: indeed, not all element have inverses.
- The set $\mathbb{Z}$, endowed with the special composition law $(x, y) \mapsto x - y$ is not a group: indeed, $x - (y - z) \neq (x - y) - z$.
- The set $\mathbb{Q}^*$ (resp. $\mathbb{R}^*$, $\mathbb{C}^*$) obtained by removing 0 from $\mathbb{Q}$ (resp. etc.), with composition law the usual multiplication and neutral element 1, is an Abelian group called the multiplicative group of rationals (resp. etc.). It has infinite order.

[1]There is no "set of all groups"; but 0 is the initial and terminal object in the category of groups, which gives a precise and rigorous sense to the notion that it is "smallest".

- The set $\mathbb{Z}^*$ does not form a multiplicative group, since some non-zero integers do not have an integer inverse.
- Let X be a set, and $S = \mathrm{Aut}(X)$ the set of all permutations (i.e., reorderings) of X. Then S, along with the composition law of functions, and the identity application for neutral element, is a group called the symmetric group of X and usually denoted $\mathfrak{S}_X$. If $|X| = n$, then $\mathfrak{S}_X$ is of order $n!$; for $n \geq 3$, $\mathfrak{S}_X$ is non-Abelian.
- The set of invertible $n \times n$ matrices, together with the operation of matrix multiplication and matrix inverse, for a group called the general linear group of degree n, denoted GL_n.

Definition 12 (Direct product) Let $G_1, \ldots, G_n$ be groups. The direct product $G = G_1 \times \cdots \times G_n$ is a group, whose elements are of the form $x = (x_1, \ldots, x_n)$ with $x_i \in G_i$, the composition law is the element-wise application of each G_i's law, i.e.,

$$(x_1, \ldots, x_n) \cdot (y_1, \ldots, y_n) = (x_1 y_1, x_2 y_2, \ldots, x_n y_n)$$

and the neutral element is $e = (e_1, \ldots, e_n)$ where e_i is G_i's neutral element. The inverse of $x = (x_1, \ldots, x_n)$ is $x^{-1} = (x_1^{-1}, \ldots, x_n^{-1})$.

Remark 5 For Abelian groups, this notion is also called the direct sum $G_1 \oplus \cdots \oplus G_n$.

2.2 Subgroups

Definition 13 (Subgroup) Let $(G, \cdot, e)$ be a group and $H \subset G$, such that

- $e \in H$; (Existence of the neutral element)
- $\forall x, y \in H, xy \in H$; (Stability under multiplication)
- $\forall x \in H, x^{-1} \in H$; (Stability under inversion)

then H is said to be a subgroup of G, and we write $H \leq G$. Equivalently, H is a subgroup of G if and only if H is non-empty and for all $x, y \in H$, $xy^{-1} \in H$.

Example 3

- Let G be a group. Then G and $0 = \{e\}$ are subgroups of G. 0 is called the trivial subgroup of G. If $H \leq G$ and $H \neq \{e\}, G$ then H is called a proper subgroup of G.
- Let G be a group and $x \in G$. Then $\langle x \rangle = \{x^k \mid k \in \mathbb{Z}\}$ (the group generated by x) is an Abelian subgroup of G. The order of $x \in G$ is by definition the order of the subgroup it generates.
- A particularly important class of groups are the groups $\langle x \rangle$ (generated by a single element x) that are finite. These are called cyclic groups.
- $\{q \in \mathbb{Q}^* \text{ s.t. } q > 0\}$ is a subgroup of $\mathbb{Q}^*$.
- $\mu_2(\mathbb{Q}^*) = \{\pm 1\}$ is a subgroup of $\mathbb{Q}^*$.
- Let $f : G \to H$ be a group morphism, and define $\ker f = \{g \in G \text{ s.t. } f(g) = e_H\}$ the kernel of f. Then $\ker f$ is a subgroup of G, and $\ker f = 0 = \{e\}$ if and only if f is injective.
- Let $n \in \mathbb{Z}$, then $n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$ is a subgroup of $\mathbb{Z}$.

- The derived group (or commutator group) G' (or $[G,G]$) is the subgroup of G generated by all commutators (elements of the form $[g,h] = g^{-1}h^{-1}gh$ where $g,h \in G$). A perfect group is a group G such that $G = G'$.

Proposition 2 (Unions and intersections of subgroups) *Let G be a group, and $(H_i)_{i\in I}$ be a family of subgroups of G.*

- *$\bigcap_{i\in I} H_i$ is a subgroup of G;*
- *$H_i \cup H_j$ is a subgroup of G if and only if ($H_i \subset H_j$ or $H_j \subset H_i$).*

Corollary 3 *Let $a,b \in \mathbb{Z}$, $a,b \neq 0$, then $a\mathbb{Z} \cap b\mathbb{Z} = \mathrm{lcm}(a,b)\mathbb{Z}$.*

Amongst subgroups, some are particularly interesting:

Definition 14 (Normal subgroup) A subgroup N of a group G is called a normal subgroup[2] if it is invariant under conjugation, i.e., for all $n \in N$ and all $g \in G$, $gng^{-1} \in N$. When that is the case, we write $N \triangleleft G$.

Example 4

- Let G be a group, then $G \triangleleft G$ and $0 = \{e\} \triangleleft G$. A group that has no other normal subgroups is called a simple group.[3]
- The center of G is defined as $Z(G) = \{z \in G \mid \forall g \in G, zg = gz\}$; it is a normal subgroup of G: $Z(G) \triangleleft G$. A group G is Abelian if and only if $Z(G) = G$.
- The derived group is a normal subgroup.
- If G is an Abelian group, and $H \leq G$, then $H \triangleleft G$.

2.3 Group quotients

Lemma 3 *Let G be a group and $H \leq G$. The relation*

$$x \sim_H y \qquad \Leftrightarrow \qquad x - y \in H$$

is an equivalence relation.

Definition 15 (Quotient group) Let G be a group, and $H \triangleleft G$. The set of equivalence classes under $\sim_H$ is denoted G/H and called the quotient group of G by H. (Informally, we may say "G mod H".)

Remark 6 One way to think about G/H is to consider elements of H as "neutral", i.e., we remain in the same equivalence class by multiplying or dividing by elements of H (resp., in additive notation, adding or subtracting).

Remark 7 The requirement that $H \triangleleft G$ (and not just $H \leq G$) makes G/H a group, and not merely a set. Indeed, the group law is compatible with this quotient. The reader will easily check that H being normal is equivalent to the group law being compatible with the quotient.

---

[2] In the literature, they are sometimes also referred to as "invariant" or "distinguished" subgroups. Galois, who first stressed the importance of such subgroups, called them "propres".

[3] All finite simple groups have been classified; the proof was computer-assisted, and counts tens of thousands of pages, published in several hundred journal articles, written by about 100 authors, and written mostly between 1955 and 2004.

Theorem 9 (Universality) *Let G be a group and $H \triangleleft G$. The (categorical) quotient of G by H is a group Q, together with a morphism $u : G \to Q$ such that $\ker u = H$ which is universal in the following sense.*

If $\phi : G \to G'$ is any morphism such that $H \subset \ker \phi$, then there is a unique induced morphism f which makes the following diagram commute:

$$\begin{tikzcd} G \arrow[r, "\phi"] \arrow[d, "u"'] & G' \\ Q \arrow[ur, dashed, "f"] & \end{tikzcd}$$

Definition 16 (Short exact sequence) A diagram of groups and morphisms

$$0 \to A \xrightarrow{f} B \xrightarrow{g} C \to 0$$

is a short exact sequence if the following properties hold:

- f is injective;
- $\ker g = \operatorname{Im} f$;
- g is surjective.

This is equivalent to stating that $A \triangleleft B$, and $C \simeq B/A$.

Theorem 10 (First isomorphism theorem) *Let $f : G \to H$ be a group morphism. Then $\ker f \triangleleft G$, and $\operatorname{Im} f \simeq G/\ker f$. Conversely, if $N \triangleleft G$, the kernel of the quotient map $q : G \to G/N$ is N itself; therefore the normal subgroups are precisely the kernels of morphisms with domain G.*

Example 5

- $G/G \simeq 0$ and $G/\{e\} \simeq G$.
- $\mathbb{R}/\mathbb{Z}$ is isomorphic to the circle group $\mathbb{S}^1$.
- $\mathbb{R}^2/\mathbb{Z}^2$ is isomorphic to a torus. More generally, if $\vec{u}, \vec{v} \in \mathbb{R}^2$ are non-colinear, then $\Lambda = \{a\vec{u} + b\vec{v} \mid a, b \in \mathbb{Z}\}$ is an additive subgroup of $\mathbb{R}^2$, i.e., an Euclidean lattice, and $\mathbb{R}^2/\Lambda$ is a torus.
- The Abelianisation of G is the group $G^{\text{ab}} = G/G'$; it is an Abelian group. In particular, a group G is Abelian if and only if $G' \simeq 0$.[4]
- Let X be a topological space, and $x \in X$. The set of loops $\gamma : [0,1] \to X$, $\gamma(0) = \gamma(1) = x$ can be endowed with a natural group structure; quotiented by the homotopy equivalence relationship ($\gamma \sim \eta \Leftrightarrow \gamma$ can be continuously deformed into η), this gives the first homotopy group $\pi_1(X)$. The Abelianised of $\pi_1(X)$, denoted $H_1(X)$, is called the first homology group. Both $\pi_1(X)$ and $H_1(X)$ are central objects of interest in algebraic topology.

Theorem 11 (Lagrange) *Let G be a finite group, and H a subgroup of G. We have $|G| = |H| \cdot |G/H|$, and in particular, the order of H divides the order of G.*

[4] An important interpretation of G^{ab}, which is beyond the scope of this course, is as $H_1(G, \mathbb{Z})$: the first homology group of G with integral coefficients.

Proof: Let $x \in G$, let $xH = \{xh \mid h \in H\}$. The map $H \to xH$ defined by $h \mapsto xh$ is a bijection, therefore $H \simeq xH$. In other terms, all the equivalence classes have the same cardinal. Since G is the disjoint union of its equivalence classes, the result immediately follows. $\square$

Corollary 4 *Let G be a finite group of prime order. Then the only subgroups of G are 0 and G.*

2.4 $\mathbb{Z}$ as a group

Lemma 4 *Let $n \in \mathbb{N}$, $n > 0$ then $(n) = n\mathbb{Z}$ is a normal subgroup of $\mathbb{Z}$.*

Theorem 12 (Subgroups of $\mathbb{Z}$) *Let $H \leq \mathbb{Z}$, then there exists $n \in \mathbb{Z}$ such that $H \simeq (n)$.*

Proof: The result is immediate if $H = \{0\} = 0$, by taking $n = 0$; therefore let's assume $H \neq 0$. Let $A = H \cap \{\mathbb{N} - \{0\}\}$, which is a non-empty subset of $\mathbb{N}$. By Fact 1 there is a smallest element in A that we denote by n.

Since H is a subgroup of $\mathbb{Z}$, and since $n \in H$, we have $(n) \subset H$.

Conversely, let $h \in H$, then there exists by Theorem 1 two integers $q, r \in \mathbb{Z}$ such that $h = nq + r$ with $0 \leq r < n$. As $n \in H$, so does $nq \in H$. Therefore $r = h - nq \in H$.

Assume $r > 0$, then $r \in H$ and $r < n$ lead to a contradiction, since by definition n is the smallest element of H. Therefore $r = 0$. As a consequence, $h = nq \in (n)$. $\square$

Definition 17 (Integers modulo n) Let $n \in \mathbb{N}, n > 0$. The group $\mathbb{Z}/(n)$ (also occasionally denoted $\mathbb{Z}/n\mathbb{Z}$) is called the group of integers modulo n. The corresponding short exact sequence is

$$0 \to (n) \hookrightarrow \mathbb{Z} \twoheadrightarrow \mathbb{Z}/(n) \to 0$$

Remark 8 We may equivalently say the following:

- $x - y \in (n)$
- $x - y = 0$ in $\mathbb{Z}/(n)$
- $x \equiv y \bmod n)$
- $n|(x - y)$
- x and y are congruent modulo n
- etc.

Proposition 3 *Let $n \in \mathbb{N}$, $n > 0$. Then $\mathbb{Z}/(n)$ has order n, and*

$$\mathbb{Z}/(n) = \{\overline{0}, \overline{1}, \ldots, \overline{n-1}\}$$

where $\overline{x} = \{x + nk \mid k \in \mathbb{Z}\}$ is the equivalence class of x.

Proof: Let $a \in \mathbb{Z}$. There are by Theorem 1 integers $q, r \in \mathbb{Z}$ such that $a = nq + r$ with $0 \leq r < n$. Since $a - r \in (n)$, we have $\overline{a} = \overline{r}$.

Furthermore, for any distinct a, b in $\{0, 1, \ldots, n-1\}$, $n \nmid (a - b)$, i.e., $\overline{a} \neq \overline{b}$. $\square$

Remark 9 A direct product of rings can be constructed, analogously to the direct product of groups.

Theorem 13 (Subgroups of $\mathbb{Z}/(n)$) *Let $n > 1$ and $H \leq \mathbb{Z}/(n)$. Then H has order $d|n$. Conversely, for every $d|n$, there is a unique subgroup of $\mathbb{Z}/(n)$ which has order d, and is isomorphic to $\mathbb{Z}/(d)$.*

Proof: The first statement is just Theorem 11. Let $G = \mathbb{Z}/(n)$ and assume $H \leq G$. Let $\psi : Z \to G$ be the canonical surjection (i.e., reduction mod n) and $\phi : \mathbb{Z} \to G \to G/H$. Then $\ker\phi \leq \mathbb{Z}$, therefore by Theorem 12 there exists $d \in \mathbb{Z}$ such that $\ker\phi \simeq (d)$. Then by Theorem 10 $H \simeq \mathbb{Z}/(d)$. Finally, $\ker\psi \leq \ker\phi$ so that $d|n$. □

Corollary 5 *Let $n \in \mathbb{Z}$, $n > 1$, and $k \in \mathbb{Z}/(n)$. The cyclic group generated by k can be generated by $\gcd(k, n)$, and it has order $n/\gcd(k, n)$.*

Proof: Since $\gcd(k, n)|k$ we have $(k) \subset \gcd(k, n)$. Conversely, write $uk + vn = \gcd(k, n)$: by Theorem 7 this has a solution $u, v \in \mathbb{Z}$, therefore $\gcd(k, n)$ belongs to the group generated by k modulo n. Hence, $\gcd(k, n) \subset (n)$. Finally, Theorem 13 gives the order as $n/\gcd(k, n)$. □

Corollary 6 *Let $n \in \mathbb{Z}$, $n > 1$, and $k \in \mathbb{Z}/(n)$. Then k generates all of $\mathbb{Z}/(n)$ if and only if $\gcd(k, n) = 1$.*

Theorem 14 (Chinese remainder theorem) *Let $m, n \in \mathbb{Z}$, $n, m > 1$ and $\gcd(n, m) = 1$. Then the map $\mathbb{Z} \to \mathbb{Z}/(n) \times \mathbb{Z}/(m)$, which maps an integer to its equivalence classes modulo n an m, induces an isomorphism*

$$\mathbb{Z}/(nm) \simeq \mathbb{Z}/(n) \times \mathbb{Z}/(m).$$

Proof: Notice that the orders are equal, so we only need to show that the map, let's call it ϕ, is injective. Let $k \in \ker\phi$, then $n|k$ and $m|k$; as $\gcd(n, m) = 1$, we thus have $nm|k$. Hence $\ker\phi \subset (nm)$; and it is immediate that $(nm) \subset \ker\phi$. □

Remark 10 An important use of this theorem is as follows. Let $a \in \mathbb{Z}/(n)$ and $b \in \mathbb{Z}/(m)$, we want to find $c \in \mathbb{Z}/(nm)$ such that $\phi(c) = (a, b)$. To that end, start from the Bézout equation $un+vm = 1$, which has solutions $u, v \in \mathbb{Z}$ by Theorem 7. Then let $c = unb+vma$.

Remark 11 The Chinese theorem (CRT for short) often appears as a practical way to speed up computations. Instead of working in a full-sized ring, we may break the computation into several independent and smaller tasks, only to reassemble the results in a final and simple stage.

For instance, to multiply 1234 by 5678 modulo $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 = 30030$, we compute the products modulo each subfactor. Calling X the result of multiplication:

$$\begin{aligned} X &\equiv 1234 \cdot 5678 \equiv 0 \times 0 \equiv 0 \bmod 2 \\ X &\equiv 1234 \cdot 5678 \equiv 1 \times 2 \equiv 2 \bmod 3 \\ \vdots &\equiv \vdots \\ X &\equiv 1234 \cdot 5678 \equiv 9 \times 9 \equiv 4 \bmod 11 \\ X &\equiv 1234 \cdot 5678 \equiv 1 \times 3 \equiv 3 \bmod 13 \end{aligned}$$

then use the CRT to construct X modulo 30030 from the above information.

2.5 Cyclic groups

Lemma 5 *A finite group G of order n is cyclic if and only if there exists $x \in G$ of order n.*

Proof: Assume G is cyclic, i.e., there exists $x \in G$ such that $G = \langle x \rangle$. In particular, x has order n. Conversely, assume there exists $y \in G$ of order n; then $\langle y \rangle$ is a subgroup of G of order n, therefore $G = \langle y \rangle$, i.e., G is cyclic. □

Example 6

- The group $\mathbb{Z}$ is cyclic, with generator 1.
- Let $n \in \mathbb{Z}$, $n > 1$, the group $(\mathbb{Z}/(m), +)$ is cyclic, of order n.
- Let $n \in \mathbb{Z}$, $n > 1$, and let $\mu_n = \{\exp(2ki\pi/n) \mid 0 \leq k < n\}$. Then $\{\mu_n, \times\}$ is a cyclic subgroup of $\mathbb{C}^*$, of order n, called the group of complex n-th roots of unity.
- The group $(\mathbb{Z}/(2) \times \mathbb{Z}/(3), +)$ is cyclic, of order 6, and has generator $(1, 1)$.
- More generally, $G_1 \times G_2$ is cyclic if and only if $\gcd(n_1, n_2) = 1$, where n_1 and n_2 are the orders of G_1 and G_2 respectively. If furthermore $G_1 = \langle x_1 \rangle$ and $G_1 = \langle x_2 \rangle$, then $G_1 \times G_2 = \langle (x_1, x_2) \rangle$.

Corollary 7 (Groups of prime order are cyclic) *Let G be a finite group of prime order p. Then G is cyclic.*

Proof: Let $x \neq e$, then the order of x divides p by Theorem 11. Since p is prime, this means that x has order p. In particular, every element other than e generates G. □

Theorem 15 (Subgroups of a cyclic group) *Let G be a cyclic group of order n. Then,*

1. *Every subgroup of G is cyclic.*
2. *For every $d|n$, $d \geq 1$, $H_d = \{x \in G \mid x^d = e\}$ is a subgroup of G, of order d.*
3. *The map $d \mapsto H_d$ is a bijection between the (positive) divisors of n and the subgroups of G.*

In particular, for every $d|n$, H_d is the unique subgroup of G of order d.

Proof: Let G be a cyclic group of order n, and x a generator of G.

1. Let $H \leq G$, and δ be the smallest integer such that $x^\delta \in H$. Then $\langle x^\delta \rangle \leq H$. Let $y \in H$, there exists an integer m such that $y = x^m$ (because G is cyclic). Furthermore, there exist integers $q, r \in \mathbb{Z}$, $0 \leq r < \delta$, such that $m = \delta q + r$ (Theorem 1). Therefore, $x^r \in H$, which implies $r = 0$. Thus, $m = \delta q$, i.e., $y = (x^\delta)^q \in \langle x^\delta \rangle$, i.e., H is cyclic.
2. Let $d|n$, $d \geq 1$. One easily checks that $H_d \leq G$ as it satisfies all the axioms of being a subgroup. Observe that $\left(x^{n/d}\right)^d = x^n = e$, so that $x^{n/d} \in H_d$. The order of $x^{n/d}$ is $n/\gcd(n/d, n) = d$, so that d divides the order of H_d (Theorem 11). Since $H_d \leq G$, it is cyclic by the previous bullet point. If y is a generator of H_d, then $y^d = e$, so that the order of y divides d. But since the order of y and the order of H_d are the same, we conclude that H_d has order d.
3. Let $H \leq G$. For every $h \in H$, we have $h^d = e$, so that $H \subset H_d$. Since H_d has

order d, in fact $H = H_d$. Now let d, d' be two positive divisors of n such that $H_d = H'_d$; these groups have same order, therefore $d = d'$.

□

Definition 18 (Euler totient) Let $n \in \mathbb{Z}$, $n \geq 1$, the number of integers smaller than n and coprime with n is called Euler's totient and denoted $\varphi(n)$.

Remark 12 There are $\varphi(n)$ generators of a cyclic group of order n.

Example 7

- $\varphi(1) = \varphi(2) = 1$, $\varphi(3) = \varphi(4) = 2$
- For $p \in \mathfrak{P}$, $\varphi(p) = p - 1$. For $p, q \in \mathfrak{P}$, $p \neq q$, $\varphi(pq) = (p-1)(q-1)$
- For $p \in \mathfrak{P}$, $r \in \mathbb{N}$, $r \geq 1$, we have $\varphi(p^r) = p^r - p^{r-1}$.

Lemma 6 *Let $n \in \mathbb{Z}$, $n \geq 1$, then*

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

Proof: Define the following sets:

$$F = \left\{ \frac{i}{n} \mid 1 \leq i \leq n \right\}$$
$$F_d = \left\{ \frac{i}{d} \mid 1 \leq i \leq d \text{ and } \gcd(i, d) = 1 \right\}.$$

In particular, $F = \bigcup_{d|n} F_d$: indeed, every element $i/d \in F_d$ can be written as ki/n with $kd = n$, which belongs to F (as $i \leq d$); conversely every element of F has an irreductible counterpart in one of the F_d. Furthermore, if $i/d = i'/d'$ belongs to $F_d \cap F_{d'}$, then $d'i = di'$ which gives $i = i'$ and $d = d'$ since i and d (resp. i' and d' are coprime). Therefore the union is disjoint and this gives the result, since $|F| = n$ and $|F_d| = \varphi(d)$. □

Theorem 16 (Euler) *Let $a, b \in \mathbb{N}$, $a \geq 1$, $b \geq 2$, and $\gcd(a, b) = 1$. Then $a^{\varphi(b)} \equiv 1 \bmod b$.*

Proof: Let $t = \varphi(b)$, and $b_1, \ldots, b_t$ be the integers between 1 and b that are coprime with b. For each $i = 1, \ldots, t$ there exist integers q_i, r_i, $0 \leq r_i < b$, such that $ab_i = bq_i + r_i$. We have

$$\prod_{i=1}^{t} ab_i \equiv \prod_{i=1}^{t} r_i \bmod b.$$

Furthermore, $\gcd(ab_i, b) = 1$, therefore $\gcd(r_i, b) = 1$, and $r_i \neq 0$.

Assume there are $i \neq j$ such that $r_i = r_j$, then $ab_i \equiv ab_j \bmod b$, i.e., $b|a(b_j - b_i)$. Since $\gcd(a, b) = 1$ by hypothesis, then $b|(b_j - b_i)$, which entails $b_i = b_j$. This is a contradiction, and therefore $r_i \neq r_j$.

As a result,

$$\prod_{i=1}^{t} ab_i \equiv \prod_{i=1}^{t} b_i \bmod b, \qquad \text{i.e.,} \qquad (a^t - 1) \prod_{i=1}^{t} b_i \equiv 0 \bmod b.$$

Since each b_i is coprime with b, so is their product, and we have $a^t \equiv 0 \bmod b$. □

Corollary 8 (Fermat's little theorem) *Let $a \in \mathbb{N}$, $a \geq 1$, and $p \in \mathfrak{P}$. Then $p|(a^p - a)$, i.e., $a^{p-1} \equiv 1t \bmod p$.*

Theorem 17 (Representation of cyclic groups) *Let G be a cyclic group of order n. Then $G \simeq (\mathbb{Z}/(n), +)$.*

Proof: Let G be a cyclic group of order n, and let $g \in G$ be a generator. Let $f : G \to \mathbb{Z}/(n)$ be the map defined by $f(g^k) = k \bmod n$, for every $k \in \mathbb{Z}$. This map is well-defined, as whenever $g^k = g^{k'}$ we have $n|(k - k')$, i.e., $k \equiv k' \bmod n$. Furthermore,

$$f(g^{k+k'}) = (k + k') \bmod n = (k \bmod n) + (k' \bmod n) = f(g^k) + f(g^{k'})$$

so that f is a group morphism. It is surjective (by design), and injective (as G and $\mathbb{Z}/(n)$ have the same order). □

Corollary 9 *Two cyclic groups are isomorphic if and only if they have the same order.*

Corollary 10 (Representation of prime order groups) *Let G be a finite group of prime order p. Then $G \simeq (\mathbb{Z}/(p), +)$.*

Proof: Corollary 7 and Theorem 17. □

Theorem 18 (Cauchy) *Let G be an Abelian finite group of order n. Let p be a prime divisor of n, then there exists $x \in G$ of order p.*

Proof: Denote $a_1, \ldots, a_n$ the elements of G, and $\alpha_1, \ldots, \alpha_n$ their respective orders. Let the map $f : \Gamma = \prod_{i=1}^{n} \mathbb{Z}/(\alpha_i) \to G$ be defined by $f((k_1, \ldots, k_n)) = a_1^{k_1} \cdots a_n^{k_n}$. One easily checks that f is well-defined and is in fact a group morphism. Furthermore, f is surjective since

$$f((0, \ldots, 1, \ldots, 0)) = a_i$$

where the "1" is placed in i-th position. By Theorem 10, $G \simeq \Gamma/\ker f$ and in particular

$$|\Gamma| = |\ker f| \cdot |G|.$$

Since p divides $|G|$, this implies that p divides $|\Gamma| = \alpha_1 \cdots \alpha_n$. Therefore, there exists $i \in \{1, \ldots, n\}$ such that $p|\alpha_i$, so that

$$a_i^{\alpha_i/p} \neq e \qquad \text{and} \qquad \left(a_i^{\alpha_i/p}\right)^p = e.$$

Hence $a_i^{\alpha_i/p}$ has order p. □

Corollary 11 (Abelian groups of square-free order are cyclic) *Let G be a finite group of order n such that n is square-free, i.e., $v_p(n) \leq 1$ for all $p \in \mathfrak{P}$. Then G is cyclic, of order n.*

Proof: Let $n = p_1 \cdots p_r$ where each $p_i \in \mathfrak{P}$. By Theorem 18 there exists $a_i \in G$ of order p_i for each $i = 1, \ldots, r$. Furthermore, for every $1 \leq s \leq r$, the element $a_1 \cdots a_s$

has order $p_1 \cdots p_s$. In particular, $a_1 \cdots a_r$ has order n, which concludes the argument using Lemma 5. □

Theorem 19 *Let G be a finite group of order n. Then G is cyclic if an only if* $\gcd(n, \varphi(n)) = 1$.

Definition 19 (Group of units) Let $n \in \mathbb{Z}$, $n > 1$. The integers modulo n that are coprime to n form a multiplicative group called the group of units and denoted $(\mathbb{Z}/(n))^\times$.

Remark 13 The order of $(\mathbb{Z}/(n))^\times$ is $\varphi(n)$.

Corollary 12 (Exponentiation in a cyclic group) *Let G be a cyclic group of order n, and $g \in G$. Let $k \in \mathbb{Z}$. Then $g^k = g^{k \bmod \varphi(n)}$.*

We will not prove the following theorem, but it highlights why the study of cyclic groups is fundamental to understanding the structure of general Abelian groups:

Theorem 20 (Kronecker) *Let G be a finite abelian group. Then*

$$G \simeq \bigoplus_{i=1}^{n} C_{q_i}$$

where C_q is the cyclic group of order q, and q_i are powers of prime.

Example 8

- Let $G = \mathbb{Z}/(15)$, then by the CRT $G \simeq \mathbb{Z}/(3) \times \mathbb{Z}/(5)$. One concrete realisation of this division is $\mathbb{Z}/(15) \simeq \{0, 5, 10\} \oplus \{0, 3, 6, 9, 12\}$.
- Let G be an Abelian group of order 8. Then G is isomorphic either to $\mathbb{Z}/(8)$, $\mathbb{Z}/(4) \times \mathbb{Z}/(2)$, or $(\mathbb{Z}/(2))^3$.

Historical note. Group theory as it is today has several sources. One such source was the question of solving high degree polynomials (degree > 4), for which Lagrange introduced a "permutation" approach which was refined by Ruffini and Galois to yield the notion of permutation group. Galois is the first to use the name group, in French *groupe*. He developed the notion of a normal subgroup and used this to give an elegant proof that general polynomials of degree 5 and above couldn't be solved by radical techniques. Cayley formalised the notion of an abstract group around 1854.

Another source is Klein's *Erlangen programme* (1872), who introduced groups as a way to organise geometric knowledge. As a result, many new groups were introduced and studied, both discrete and continuous, and found their way at the heart of representation theory. Lie groups in particular are ubiquitous in modern Physics.

The classification of finite simple groups, possibly the largest mathematical effort ever made in history, started in the 1950s and was successfully completed around 2004 (an earlier 1983 announcement was premature). Because of the extreme length of the complete proofs for that classification, much effort has been devoted to finding a simpler proof. It is estimated that the new proof will eventually fill approximately 5000 pages.

2.6 Exercise set

Exercise 2.1. Let G be a group with neutral element e. Show that if for all $x \in G$, $x^2 = e$, then G is Abelian.

Exercise 2.2. Let G be a finite group of even order, with neutral element e. Show that there exists $x \in G$, $x \neq e$, such that $x^2 = e$.

Exercise 2.3. Let G be a group, and A and B be subgroups of G. Let

$$AB = \{ab \mid a \in A, b \in B\}. \qquad \text{and} \qquad BA = \{ba \mid a \in A, b \in B\}.$$

Show that $AB = BA$ if and only if AB is a subgroup of H.

Exercise 2.4. Let $n > 1$ and $k|n$ be integers. Prove that if $k|n$ then $(n) \subset (k)$. Does the converse hold?

Exercise 2.5. Let $n \in \mathbb{Z}$ such that $n \equiv 3 \bmod 7$ and $n \equiv 4 \bmod 11$. What is the equivalence class of n modulo 77?

Exercise 2.6. Prove the short five lemma: In the diagram

$$\begin{tikzcd}
0 \arrow[r] & A \arrow[r] \arrow[d, "g"] & B \arrow[r] \arrow[d, "f"] & C \arrow[r] \arrow[d, "h"] & 0 \\
0 \arrow[r] & A' \arrow[r] & B' \arrow[r] & C' \arrow[r] & 0
\end{tikzcd}$$

where every row is a short exact sequence, if g and h are isomorphisms, then so is f.

Exercise 2.7. Show that for every $n \in \mathbb{Z}$, n odd, $\varphi(n) = \phi(2n)$.

> **Open question (2018)**: Is there, for every even integer n, an integer $m \neq n$ such that $\varphi(n) = \phi(m)$?
>
> **Open question (2018)**: Is there an integer $n \notin \mathfrak{P}$, such that $\varphi(n)|(n-1)$?

Exercise 2.8. Let $n \in \mathbb{Z}$, $n \geq 1$. Show that $\varphi(n)|n!$.

Exercise 2.9. Find all the $p \in \mathfrak{P}$ such that $p|(2^p + 1)$.

Exercise 2.10. Show that $a \mid b$ implies $\varphi(a) \mid \varphi(b)$. Does the converse hold?

Exercise 2.11. Every commercial book is given an International Standard Book Number, or ISBN. Since 2007, this number consists in 13 digits (describing the publisher, country group, etc.) $x_1, \ldots, x_{13}$. An ISBN is valid if

$$x_1 + x_3 + \cdots + x_{13} \equiv -3\,(x_2 + x_4 + \cdots + x_{12}) \bmod 10.$$

Write a program that checks whether a given ISBN is valid. What is the 13-th digit in the ISBN `978041547370X`?

Exercise 2.12. Implement fast modular addition and multiplication using the Chinese remainder theorem.

Exercise 2.13. Compute $2^{2017} \bmod 64$. Implement fast modular exponentiation using Euler's theorem.

Exercise 2.14. Can one combine the two previous exercises to perform exponentiation even faster? Implement this.

Exercise 2.15. Let $A_1, A_2, \dots$ be a sequence of Abelian groups, and assume that we have an exact sequence

$$A : \cdots \xrightarrow{d_4} A_3 \xrightarrow{d_3} A_2 \xrightarrow{d_2} A_1 \xrightarrow{d_1} A_0 \xrightarrow{d_0} A_{-1} \xrightarrow{d_{-1}} \cdots$$

i.e., $\ker d_k = \operatorname{im} d_{k+1}$, i.e., $d_k d_{k+1} = 0$. Such a sequence is called a chain complex. Show that for every $n \in \mathbb{Z}$, $\operatorname{im} d_{n+1}$ is a normal subgroup of $\ker d_n$, and call H_n the quotient group $\ker d_n / \operatorname{im} d_{n+1}$. H_n is called the n-th homology group of A.

Consider $X = \mathbb{R}^3$, on which we can define the operators "gradient", "divergence", and "curl":

$$\operatorname{grad} f = \nabla f, \qquad \operatorname{div} f = \vec{\nabla} \cdot f, \qquad \operatorname{curl} f = \vec{\nabla} \times f$$

1. Let F_1 be the space of smooth functions on X with values in $\mathbb{R}$, and F_3 be the space of smooth functions on X with values in $\mathbb{R}^3$. Show that these are Abelian groups, and in fact, $\mathbb{R}$-vector spaces.
2. Show that each of the three operations above are group morphisms, and in fact, linear operations.
3. Show that for any $f \in F_1$, $\operatorname{curl} \operatorname{grad} f = 0$ (loosely written $\vec{\nabla} \times \nabla = 0$)
4. Show that for any $f \in F_3$, $\operatorname{grad} \operatorname{div} f = 0$ (loosely written $\nabla \vec{\nabla} = 0$)
5. Show that for any $f \in F_3$, $\operatorname{div} \operatorname{curl} f = 0$ (loosely written $\vec{\nabla} \cdot \vec{\nabla} = 0$)
6. Write down a chain complex using F_1, F_3, and the above operations.

The above was the starting point of the development of de Rham (co)homology, one of the most important examples of a (co)homology theory.

Exercise 2.16. Let $n \in \mathbb{N}$, define the set

$$E_n = \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \mid x^2 + y^2 = 5^n\}.$$

1. Compute E_0.
2. Let $a \in E_0$. For any b such that $0 \le a \le n$, define
$$Z_{a,b} = a(2 + i)^b (2 - i)^{n-b},$$
where $i^2 = -1$. Show that $|Z_{a,b}|^2 = 5^n$, and that the application
$$(a, b) \mapsto Z_{a,b}$$
is an injective map from $E_0 \times \{0, \dots, n\}$ to E_n.
3. Let $(x, y) \in E_n$. Show that either
$$\begin{cases} 2x - y \equiv 0 \bmod 5 \\ x + 2y \equiv 0 \bmod 5 \end{cases}$$
or
$$\begin{cases} 2x + y \equiv 0 \bmod 5 \\ -x + 2y \equiv 0 \bmod 5 \end{cases}$$

4. Deduce from the above that either $z/(2+i)$ or $z/(2-i)$ gives an element of E_{n-1} (taking its real and imaginary parts).

5. Show that $(a,b) \mapsto Z_{a,b}$ is in fact bijective, and deduce the number of elements in E_n, i.e., the number of solutions to the Diophantine equation

$$x^2 + y^2 = 5^n.$$

Exercise 2.17. The set of operations that leave a unit square unchanged are called symmetries. For instance, a 90 degree rotation clockwise is such a symmetry.

1. List all the symmetries for the unit square. They form a group, called the dihedral group D_4.

2. Is D_4 a commutative group? Show that any element of D_4 is generated by a (clockwise rotation of 90 degrees) and b (left-right flip). Check that

$$ab = ba^3, \quad a^2b = ba^2, \quad \text{and} \quad a^3b = ba.$$

   All this information can be condensed in the presentation $D_4 = \langle a, b \mid a^4, b^2 \rangle$, meaning that D_4 is formed of any sequence of the symbols a and b, where $a^4 = b^2 = e$ the unit.

3. Identify the five cycles of D_4 (we do not count the powers of the unit as a cycle). You can represent this information as a cycle graph as in Figure 2.2. Label each node of this figure with the appropriate transformations (the identity is indicated).

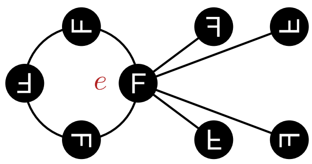


Figure 2.2: Cycle graph for the dihedral group D_4.

4. Work out the cycle graph for the symmetry groups of the segment, the triangle, and the hexagon, respectively D_2, D_3, and D_6. Are they commutative?

Chapter 3

Rings

We have seen in the previous chapter that $\mathbb{Z}$, together with the addition, forms a group. But there is another important familiar operation on integers: multiplication. Taking this new operation into account reveals an immensely rich and fertile structure.

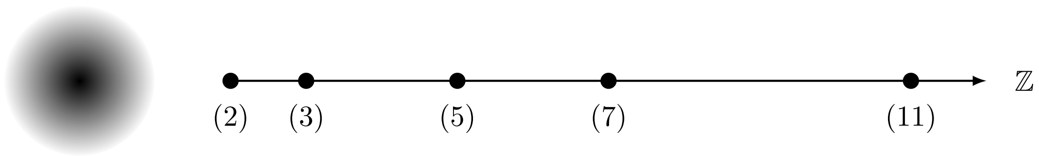


Figure 3.1: The spectrum of $\mathbb{Z}$. See Example 15.

Contents

3.1 Basic definitions

Definition 20 (Ring) A ring is a set R, together with two binary operations (denoted $+$ and $\cdot$ respectively) such that

1. $(R, +)$ is an abelian group; the neutral element for $+$ is denoted 0.
2. For every $x, y, z \in R$, $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.
3. There exists a neutral element for $\cdot$ which is denoted 1, and $1 \neq 0$.[1]
4. For every $x, y, z \in R$,

$$a \cdot (b + c) = a \cdot b + a \cdot c$$
$$(b + c) \cdot a = b \cdot a + c \cdot a.$$

Example 9

- $\mathbb{Z}/(2) = \{0, 1\}$ is the smallest possible ring.
- $\mathbb{Z}, \mathbb{Q}, \mathbb{R}$, and $\mathbb{C}$ have a ring structure, given by the usual addition and multiplication operations, and the usual neutral elements 0 and 1. However, $\mathbb{N}$ is not a ring, as $(\mathbb{N}, +)$ is not a group.
- Let R be a ring, then $n \times n$ matrices together with matrix addition and multiplication form the ring of matrices, denoted $M_n(R)$.

Definition 21 (Characteristic) Let R be a ring and $r \in R$. Let $n \in \mathbb{N}$, $n > 1$. We write $r + \cdots + r$ (n times) as nr. The characteristic of R, denoted $\operatorname{char} R$ is the least positive integer n such that $nr = 0$ for all $r \in R$. If no such integer exists, we say that the characteristic is zero.

Definition 22 (Group of units) Let R be a ring. The group of units of R, denoted $R^\times$, is the group of elements that have a multiplicative inverse in R.

Definition 23 (Subring) A subset $S \subset R$ such that $1 \in S$, that is closed under addition, additive inverses, and multiplication is called a subring of R.

Example 10

- $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ forms a tower of subrings.
- The set $A = \{a/b \mid a \in \mathbb{Z}, b \in 2\mathbb{Z} - \{0\}\}$ is a subring of $\mathbb{Q}$.
- For every integer $n > 1$, the group $Z/(n)$ is in fact a ring.
- The set of functions from a set X to a ring Y is itself a ring called the ring of functions from X to Y.
- The set $M_n(R)$ of $n \times n$ matrices with coefficients in a ring R is itself a ring, called the ring of matrices with coefficients in R.

Definition 24 (Commutative ring) We say that a ring $(R, +, \cdot)$ is commutative if the operation $\cdot$ is commutative.

[1] Some authors leave out the requirement that $1 \neq 0$ and distinguish "rings with identity" from "rings (without identity)"; we will always assume rings "with identity".

Definition 25 (Field) Let R be a commutative ring. If $R - \{0\}$ is a group under multiplication, then we say that R is a field.

Example 11

- If K is a field, then $K^\times = K - \{0\} = K^*$.
- $\mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$ is a tower of fields.
- However $\mathbb{Z}$ is not a field, as some integers do not have integer multiplicative inverses.
- $M_n(\mathbb{R})$ is not a field for $n \geq 2$, as some matrices are singular and thus do not have multiplicative inverses.
- Let
$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$
where $i^2 = -1$. These matrices satisfy the following relations:
$$\begin{gathered} \mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}, \\ \mathbf{ij} = \mathbf{k}, \quad \mathbf{jk} = \mathbf{i}, \quad \mathbf{ki} = \mathbf{j} \\ \mathbf{ji} = -\mathbf{k}, \quad \mathbf{kj} = -\mathbf{i}, \quad \mathbf{ik} = -\mathbf{j} \end{gathered}$$
Let $\mathbb{H}$ consist of matrices of the form $a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, where $a, b, c, d \in \mathbb{R}$. Observe that if $a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \in \mathbb{H}$, then
$$(a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k})(a\mathbf{1} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}) = a^2 + b^2 + c^2 + d^2$$
which is zero if and only if $a = b = c = d = 0$. Thus, if $a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} \neq 0$, we can define
$$(a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k})^{-1} = \frac{a\mathbf{1} - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}}{a^2 + b^2 + c^2 + d^2}.$$
The ring $\mathbb{H}$, called the quaternions, is however not a field because it is not commutative (it is a "division ring").
- Let M be a finite set, and K a field, then the set of all invertible functions form a field, the function field of M in K.

Theorem 21 (Finite fields of prime order) *Let $p \in \mathfrak{P}$, then $\mathbb{Z}/(p)$ is a field. It is called the finite field of order p, and denoted $\mathbb{F}_p$.*

Proof: The only point to check is that every non-zero element of $\mathbb{Z}/(p)$ has a multiplicative inverse. Let $a \in \mathbb{Z}/(p)$, $a \neq 0$. In particular, $\gcd(a, p) = 1$ so by Theorem 7 there exists $r, s \in \mathbb{Z}$ such that $ar + ps = 1$. Since $ps = 1 - ar$, we have $ar = 1 \bmod p$, i.e., r is the inverse of a in $\mathbb{Z}/(p)$. $\square$

Lemma 7 *Let R be a ring, and $a \in R$. If $a^{-1} \in R$, then a is not a zero divisor, i.e., $a \neq 0$ and there is no $b \in R$, $b \neq 0$ satisfying $ab = 0$ nor $ba = 0$.*

Definition 26 (Integral domain) A commutative ring that has no zero divisor is called an integral domain.

In other words, R is an integral domain if and only if $ab = 0$ implies either $a = 0$ or $b = 0$, where $a, b \in R$.

Example 12

- $\mathbb{Z}$ is an integral domain.
- $M_n(\mathbb{R})$ is an integral domain for $n = 1$, but is not an integral domain for $n \geq 2$.

Lemma 8 (Cancellation in integral domains) *Let R be a commutative ring. Then R is an integral domain if and only if, for every $a \in R - \{0\}$, $ab = ac$ implies $b = c$.*

Theorem 22 (Wedderburn) *Every finite integral domain is a field.*

Proof: Let R be a finite integral domain, and R^* be the set of non-zero elements of R. Let $a \in R^*$, define the map $\phi_a : R^* \to R^*$ by $\phi_a(r) = ar$; this makes sense because there are no zero divisors in R^*, so $ar \neq 0$. For every $r_1, r_2 \in R^*$, $\phi_a(r_1) = \phi_a(r_2)$ implies $ar_1 = ar_2$, which implies $r_1 = r_2$ by Lemma 8; hence ϕ_a is injective. Since R^* is finite and ϕ_a is injective, it is bijective, and in particular surjective. Therefore, there exists $b \in R^*$ such that $\phi_a(b) = 1 = ab$. Thus a has a left inverse, and since R is commutative, a right inverse. This being true of every non-zero element $a \in R$, we have shown that R is a field. □

Theorem 23 (Newton) *Let R be a ring, $a, b \in R$ such that $[a, b] = ab - ba = 0$. Then for any integer $n > 0$,*

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}.$$

3.2 Ring morphisms and ideals

Definition 27 (Ring morphism) Let $\phi : R \to S$ be a function between two rings R and S, such that $\phi(1_R) = 1_S$ and for every $a, b \in R$,

$$\phi(a+b) = \phi(a) + \phi(b)$$
$$\phi(a \cdot b) = \phi(a) \cdot \phi(b).$$

Then ϕ is called a ring morphism. The kernel of ϕ is the inverse image of 0.

Example 13

- The map $R[X] \to R[X]$ that sends $f(X)$ to $f(X+1)$ is a ring morphism.
- The map $\sigma : \mathbb{C} \to \mathbb{C}$ that sends a complex number to its conjugate is a ring morphism.

Definition 28 (Ideal) Let R be a ring, and $I \subset R$. We say that I is an ideal of R if I is an additive subgroup of R such that for all $a \in I, r \in R$,

$$ar \in I \qquad \text{and} \qquad ra \in I.$$

Lemma 9 *Let $\phi : R \to S$ be a ring morphism. Then $\ker \phi$ is an ideal of R.*

Proposition 4 (Quotient ring) *Let R be a ring and $I \neq R$ be an ideal of R. Then R/I is a ring, and there is a natural morphism $R \to R/I$ that maps $r \mapsto r + I$.*

Theorem 24 (Universality) *Let R be a ring, $I \neq R$ be an ideal of R, and $u : R \to R/I$ as above. Then u is universal amongst all ring morphisms whose kernel contains I, in the following sense.*

Let $\phi : R \to S$ such that $I \subset \ker \phi$, then there exists a unique morphism f that makes the following diagram commute:

$$\begin{tikzcd} R \arrow[r, "\phi"] \arrow[d, "u"'] & S \\ R/I \arrow[ur, dashed, "f"'] & \end{tikzcd}$$

Definition 29 (Principal ideal) Let R be a commutative ring and $a \in R$. The set $(a) = aR$ is an ideal of R, and any ideal of this form is called a principal ideal.

Example 14

- Let R be a ring. Then $\{0\}$ and R are ideals of R, called the trivial ideals.
- The additive subgroup $(n) = n\mathbb{Z}$ is a principal ideal of $\mathbb{Z}$. The corresponding quotient ring is $\mathbb{Z}/(n)$.
- Let X be a set, $Y \subset X$, and R be a ring. The set of functions $X \to R$ forms a ring C, and the subset of functions that vanish on Y form an ideal V of this ring. Furthermore, the ring C/V is isomorphic to the ring of functions $Y \to R$.
- $\mathbb{Z}$ is not an ideal of $\mathbb{Q}$.
- Let R and S be rings. Then the ideals of the ring $R \times S$ are exactly the ideals $I \times J$, where I is an ideal of R and J an ideal of S respectively.

Lemma 10 *Let R be a ring, and I an ideal of R. If I contains a unit, then $I = R$.*

Proof: Suppose that $u \in R^\times \cap I$. Then $uv = 1$ for some $v \in R$, therefore $1 = uv \in I$. Then for any $a \in R$, $a = a \cdot 1 \in I$. $\square$

Corollary 13 *Let K be a field, then the only ideals of K are (0) and K.*

Definition 30 (Maximal ideal) Let R be a ring, and M be an ideal of R. If for every ideal $I \neq M$ such that $M \subset I$, we have $I = R$ we say that M is a maximal ideal of R.

Theorem 25 (Maximal ideals and fields) *Let M be a commutative ring and M an ideal of R. Then M is a maximal ideal of R if and only if R/M is a field.*

Proof: Let M be a maximal ideal of R. In particular R/M is commutative and $1 + M$ acts as the identity of R/M. We must show that every non-zero element of R/M has a multiplicative inverse. If $a + M \in R/M$ is non-zero, then $a \notin M$; let $I = \{ra + m \mid r \in R, m \in M\}$, which is non-empty since $0 \in I$. If $r_1a + m_1$ and $r_2a + m_2$ are in I, then

$$(r_1a + m_1) - (r_2a + m_1) = (r_1 - r_2)a + (m_1 - m_2)$$

also belongs to I. Also, for every $r \in R$, $rI \subset I$. Therefore, I is an ideal of R; furthermore $I \neq M$, and $M \subset I$. Since M is maximal, this implies $I = R$. Therefore

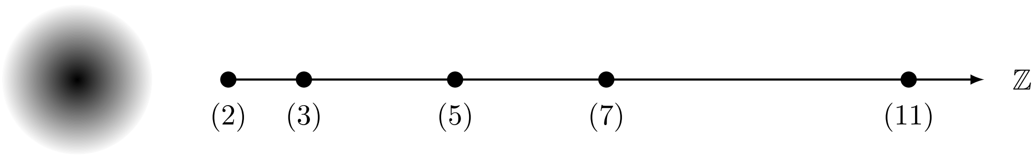


Figure 3.2: Illustration of $\operatorname{Spec}\mathbb{Z}$, composed of the prime ideals (p) and of (0). The point (0) is special (it is an "open point" in the sense of the Zariski topology).

there exists $m \in M$ and $b \in R$ such that $1 = ab + m$, and

$$1 + M = ab + M = (a + M)(b + M)$$

Conversely, suppose that R/M is a field; then it contains at least $0 + M = M$ and $1 + M$, and therefore $M \neq R$. Let I be an ideal of R such that $M \subset I$, and let $a \in I$, $a \notin M$. Since $a + M$ is non-zero, there exists an inverse $b + M$ in R/M such that $(a + M)(b + M) = ab + M = 1 + M$. Thus there exists an element $m \in M$ such that $ab + m = 1$; and $1 \in I$. Since 1 is a unit of R, this entails $I = R$. Hence M is maximal. □

Remark 14 Since $\mathbb{Z}/(p)$ is a field, we have that (p) is a maximal ideal of $\mathbb{Z}$.

Definition 31 (Prime ideal) Let R be a commutative ring. Let I be a proper ideal of R. We say that I is a prime ideal if whenever $ab \in I$, then either $a \in I$ or $b \in I$.

Theorem 26 (Prime ideals and integral domains) *Let R be a commutative ring, and I be a proper ideal of R. Then I is a prime ideal of R if and only if R/P is an integral domain.*

Corollary 14 *Every maximal ideal in a commutative ring is prime.*

Definition 32 (Spectrum of a ring) The set of all prime ideals of a ring R is called the spectrum of R and denoted $\operatorname{Spec} R$.

Example 15 $\operatorname{Spec}\mathbb{Z} = \{0\} \cup \{(p) \mid p \in \mathfrak{P}\}$. See Figure 3.2.

Definition 33 (Finite field) Let $p \in \mathfrak{P}$. The quotient $\mathbb{Z}/(p)$ is a field called the finite field of order p, and denoted $\mathbb{F}_p$.

Remark 15 The spectrum of a ring can be endowed with a topology, making it a scheme, one of the central objects of interest in algebraic geometry.

Definition 34 (Radical of an ideal) Let R be a commutative ring and I be an ideal of R. The radical of I, denoted $\operatorname{rad} I$, is the ideal defined by

$$\operatorname{rad} I = \{r \in R \mid r^n \in I \text{ for some } n > 0\}$$

Definition 35 (Reduced ring) A ring R is reduced if it has no nilpotent elements.

Theorem 27 (Reduced rings and radical ideals) *The following statements are equivalent:*

1. *An ideal I in R is radical;*
2. *The quotient ring R/I is reduced;*
3. *The nilradical of R is zero;*
4. $\bigcap_{I \in \operatorname{Spec} R} I = (0)$.

Example 16

- Every integral domain is a reduced ring since a nilpotent element is a fortiori a zero divisor, and integral domains do not have any zero divisor.
- $\mathbb{Z}/(n)$ is reduced if and only if $n = 0$ or n is a square-free integer.

3.3 Field extensions, trace and norm

Definition 36 (Field extension) Let L a field and K be a subfield of L (we say that L is an extension of K). Then L can be considered as K-vector space, and the dimension $\dim_K L$ is called the degree of the extension. An extension is finite if its degree is finite.

Example 17

- The field $\mathbb{C}$ of complex numbers is a field extension of $\mathbb{R}$ of degree 2 — such extensions are called quadratic. As a particular consequence, there are no non-trivial fields between $\mathbb{R}$ and $\mathbb{C}$.
- The ring $\mathbb{Q}[\sqrt{2}]$ of numbers of the form $a + b\sqrt{2}$ is actually a field. As such, it is a quadratic extension of $\mathbb{Q}$, denoted $\mathbb{Q}(\sqrt{2})$. Such finite extensions of $\mathbb{Q}$ are called number fields.
- The ring $\mathbb{Q}[\sqrt{2}, \sqrt{3}]$ can also be turned into a field; it is made of numbers of the form $a + b\sqrt{2} + c\sqrt{3} + d\sqrt{6}$ and is therefore of degree 4 over $\mathbb{Q}$.

Definition 37 (Field trace and norm) Let L be a finite extension of a field K. For every $a \in L$, the application $m_a : L \to L$ defined by

$$m_a : \ell \mapsto a\ell$$

is a K-linear transformation. Let $\Delta_{L/K}(a)$ be the matrix of this transformation. The field trace of a on L over K is the trace of $\Delta_{L/K}(a)$ and denoted $\operatorname{Tr}_{L/K}(a)$. The field norm of a on L over K is the determinant of $\Delta_{L/K}(a)$ and denoted $N_{L/K}(a)$.

Example 18

- A basis of $\mathbb{Q}(\sqrt{2})$ as a quadratic extension of $\mathbb{Q}$ is $\{1, \sqrt{2}\}$. Let $a \in Q(\sqrt{2})$, $a = \alpha + \beta\sqrt{2}$. Multiplication by a sends 1 to a, and $\sqrt{2}$ to $2\beta + \alpha\sqrt{2}$; the matrix of $\ell \mapsto a\ell$ is thus
$$\Delta_{\mathbb{Q}(\sqrt{2})/\mathbb{Q}} = \begin{pmatrix} \alpha & 2\beta \\ \beta & \alpha \end{pmatrix}$$
thus $N_{\mathbb{Q}(\sqrt{2})/\mathbb{Q}}(a) = \alpha^2 - 2\beta^2$ and $\operatorname{Tr}_{\mathbb{Q}(\sqrt{2})/\mathbb{Q}}(a) = 2\alpha$.
- The same reasoning with $\mathbb{C}$ as an extension of $\mathbb{R}$ gives, for any $z = x + iy \in \mathbb{C}$,
$$N_{\mathbb{C}/\mathbb{R}}(z) = x^2 + y^2, \qquad \operatorname{Tr}_{\mathbb{C}/\mathbb{R}} = 2x.$$

Remark 16 Note that the field norm is not a norm in the usual sense (that of normed vector spaces); for instance, the element $1 + \sqrt{2}$ has norm -1 in $\mathbb{Q}(\sqrt{2})$ over $\mathbb{Q}$.

3.4 Field of fractions

Let R be an integral domain. There is a generic construction of a field from R, obtained as follows. Let $\sim$ be the relation defined on $R \times R$ by $(a, b) \sim (c, d) \Leftrightarrow ad = bc$.

Lemma 11 *The relation $\sim$ is an equivalence.*

Lemma 12 *Let (a_i, b_i) be couples of $R \times R$ for $i = 1, 2, 3, 4$, such that*

$$(a_1, b_1) \sim (a_2, b_2) \quad \textit{and} (a_3, b_3) \sim (a_4, b_4).$$

Then

$$(a_1b_3 + a_3b1, b_1b_3) \sim (a_2b_4 + a_4b_2, b_2b_4) \quad \textit{and} \quad (a_1a_3, b_1b_3) \simeq (a_2a_4, b_2b_4).$$

Corollary 15 *Let $K = (R \times R)/\sim$. Define the (equivalence class of the) product as*

$$[(a, b)][(c, d)] = [(ac, bd)]$$

and the (equivalence class of the) sum as

$$[(a, b)] + [(c, d)] = [(ad + bc, bd)].$$

This makes K into a ring, and in fact a field, since for any non-zero $[(a, b)]$, $[(a, b)]^{-1} = [(b, a)]$. We call this field the field of fractions of R, denoted $\operatorname{Frac} R$.

Remark 17 There is a natural injective ring morphism $\iota : R \hookrightarrow \operatorname{Frac} R$ ('fractions with denominator 1').

Example 19

- $\operatorname{Frac} \mathbb{Z} \simeq \mathbb{Q}$
- $\operatorname{Frac} \mathbb{Z}[i] \simeq \mathbb{Q}[i]$
- For any field K, $\operatorname{Frac} K \simeq K$.

3.5 Integral elements

TODO

3.6 Exercice set.

Exercise 3.1. A ring R is called Boolean if for every $r \in R$, we have $r^2 = r$. Prove that such a ring is commutative. Let X be a set, and 2^X be the set of subsets of X (the 'power set'); using symmetric difference as addition, and set intersection as multiplication, show that 2^X is a ring, and in fact a Boolean ring.

Exercise 3.2. Let R be a ring, such that $r^3 = r$ for all $r \in R$. Show that R is commutative.

Exercise 3.3. Find two matrices A and B with coefficients in $\mathbb{Z}$ such that $AB = 0$. Now find two non-zero matrices A and B with coefficients in $\mathbb{Z}$ such that $AB = 0$.

Exercise 3.4. Let $p \in \mathfrak{P}$ and define

$$\mathbb{Z}_{(p)} = \left\{ \frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } \gcd(b, p) = 1 \right\}.$$

Prove that $\mathbb{Z}_{(p)}$ is a ring. It is called the ring of integers localised at p.

Exercise 3.5. Find the kernels of the morphisms given in Example 13.

Exercise 3.6. Let R be a ring. An element $r \in R$ is nilpotent if there exists $n > 0$ such that $r^n = 0$. Show that the set of nilpotents of R forms an ideal. It is called the nilradical of R.

Exercise 3.7. Using Bézout's theorem, show that (p) is a maximal ideal of $\mathbb{Z}$ when p is prime.

Exercise 3.8. Let i be a square root of -1, and $\mathbb{Z}[i]$ be the set of numbers of the form $a + ib$, where $a, b \in \mathbb{Z}$. Show that $\mathbb{Z}[i]$ is a ring. It is referred to as the ring of Gaussian integers.

Exercise 3.9. Let X be a set, and R be a ring. Let $f : X \to A$ be a function. Show that f is invertible if and only if $f(X) \subset A^\times$.

Exercise 3.10. Let R and S be rings. Show that $(R \times S)^\times \simeq R^\times \times S^\times$.

Exercise 3.11. Show that a ring morphism $R \to S$ induces a morphism of spectra in the reverse direction: $\operatorname{Spec} S \to \operatorname{Spec} R$.

Exercise 3.12. Equations of the form $X^2 - DY^2 = 1$, where D is a square-free integer, are called Fermat–Pell equations.[2] Show that a solution to such an equation is, in a precise sense, 'just' a unit of $\mathbb{Z}[\sqrt{D}]$. From there, show that solutions form a group and make the group law explicit.

[2]The attribution to Pell (due to Euler) is a historical mistake, but the name is now standard. We will revisit these equations later in the course.

Exercise 3.13. Let $P = X^2 - 3$.

1. Show that P is a prime ideal of $\mathbb{Z}[X]$. We use the notation $\mathbb{Z}[\sqrt{3}]$ to denote $\mathbb{Z}[X]/(P)$.
2. Recall $\operatorname{Spec}\mathbb{Z}$.
3. Show that while (3) is a prime ideal of $\mathbb{Z}$, it cannot be a prime ideal in $\mathbb{Z}[\sqrt{3}]$. Show that (2) is also a square in $\mathbb{Z}[\sqrt{3}]$. Are there more?
4. Show that $11 = -(4-3\sqrt{3})(4+3\sqrt{3})$ and that $13 = (4+\sqrt{3})(4-\sqrt{3})$, and conclude that similarly (11) and (13) are not in $\operatorname{Spec}\mathbb{Z}[\sqrt{3}]$. The same decomposition happens for every p which is a square modulo 3.
5. Thus in general, if $(p) \in \operatorname{Spec}\mathbb{Z}$, three situations can arise in $\operatorname{Spec}\mathbb{Z}[\sqrt{3}]$:
   a) (p) could turn out to be a square (we say that p "ramifies" in $\mathbb{Z}[\sqrt{3}]$);
   b) (p) could turn out to be the product of two distinct ideals;
   c) (p) could give an actual prime ideal of $\mathbb{Z}[\sqrt{3}]$.

   We will admit here that these are the three only possibilities. Explain Figure 3.3.

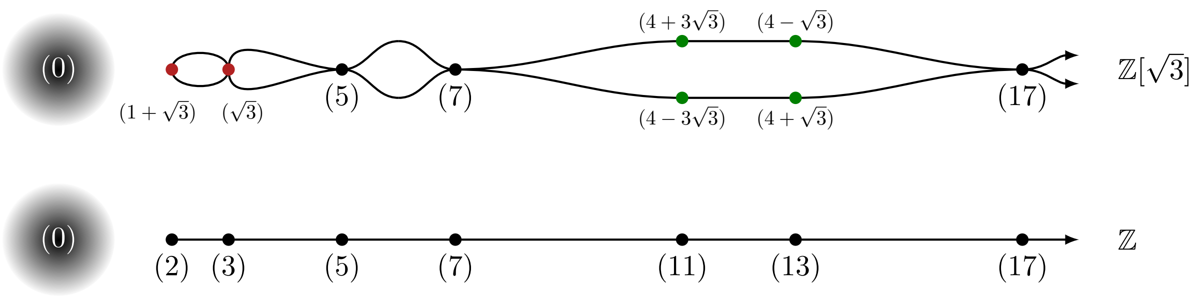


Figure 3.3: $\operatorname{Spec}\mathbb{Z}[\sqrt{3}] \to \operatorname{Spec}\mathbb{Z}$ (see Exercise 3.11).

Chapter 4

Polynomial rings

In this chapter, we develop further the theory of rings by focusing on another familiar example: polynomials. It serves both as an application of the general notions developed for rings in the previous chapter, and as a useful basis for constructing new rings and fields.

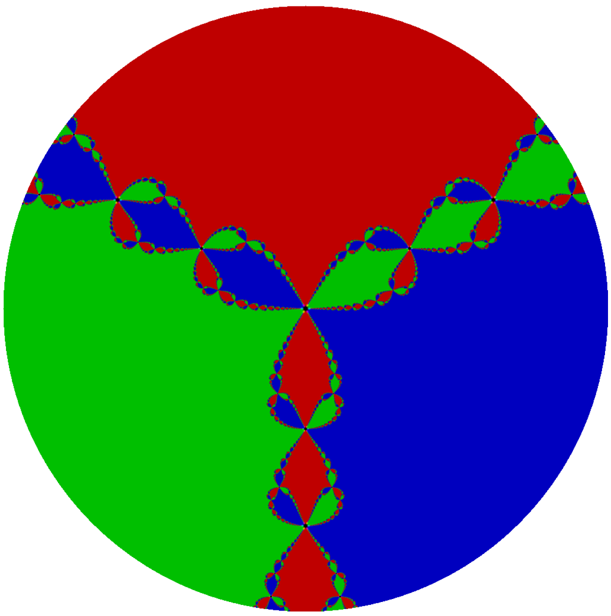


Figure 4.1: A Newton attractor for the polynomial $P = X^3 - 1$. See exercise 4.11.

Contents

4.1 Basic definitions

Definition 38 (Polynomial ring) Let R be a ring. The polynomial ring $R[X]$ is defined to be the set of all formal sums

$$\sum a_i X^i$$

where $a_i \in R$ and all the a_i are zero, except maybe for a finite number of them. Elements of $R[X]$ are called polynomials.

Definition 39 (Degree of a polynomial) Let $P \in [X]$, $P = a_0 + \cdots + a_k X^k$. The degree of P, denoted $\deg P$, is the largest k such that $a_k \neq 0$; the degree of the zero polynomial is $-\infty$.

Remark 18 We can define an ordered set $\underline{\mathbb{N}} = \mathbb{N} \cup \{-\infty\}$, by setting $-\infty \leq n$ for all $n \in \mathbb{N}$, $(-\infty) + n = n + (-\infty) = -\infty$, and $(-\infty) + (-\infty) = -\infty$. Then the degree is an application

$$\deg : K[X] \to \underline{\mathbb{N}}.$$

Proposition 5 *Let $P, Q \in R[X]$, we have*

- $\deg(P + Q) \leq \max(\deg P, \deg Q)$, *with equality if* $\deg P \neq \deg Q$*;*
- $\deg(PQ) = \deg P + \deg Q$.

Definition 40 (Monic polynomial) A polynomial $P = a_0 + \cdots + a_k X^k$ of degree k is monic if $a_k = 1$.

Corollary 16 (Units of a polynomial ring) *The ring $R[X]$ is an integral domain, and $R[X]^\times = R^\times$.*

Proof: $R[X]$ is a non-empty commutative ring. By Proposition 5, $PQ = 0$ implies either $\deg P = -\infty$ or $\deg Q = -\infty$, i.e., $P = 0$ or $Q = 0$. Hence $R[X]$ has no zero divisor. Furthermore, if $PQ = 1$, then $\deg P + \deg Q = 0$, but since these are non-negative integers, $\deg P = \deg Q = 0$. □

Theorem 28 (Euclidean division of polynomials) *Let $A, B \in S[X]$, $B \neq 0$. Then there exists a a unique couple (Q, R) of polynomials in $S[X]$ such that*

$$A = BQ + R \qquad \textit{and} \qquad \deg R < \deg B.$$

We say that Q is the quotient, and R is the remainder of the Euclidean division of A by B. If $R = 0$, we say that B divides A, and write $B|A$.

Proposition 6 *Let $A, B \in R[X]$, $A, B \neq 0$, such that $A|B$ and $B|A$. Then there exists $\lambda \in R$, $\lambda \neq 0$, such that $A = \lambda B$.*

Proof: By Theorem 28, there exists $Q_1, Q_2 \in R[X]$ such that $A = BQ_1$ and $B = AQ_2$. Therefore $A(1 - Q_1 Q_2) = 0$. Since $R[X]$ is an integral domain and $A \neq 0$, this implies $Q_1 Q_2 = 1$, i.e., Q_1 is invertible (Corollary 16). □

Remark 19 The notion of irreducibility is profoundly dependent on the underlying ring R, and the following example shows: Let $R_1 = \mathbb{Z}$ and $R_2 = \mathbb{Z}[\sqrt{2}]$. The polynomial $X^2 - 2$ is irreducible in $R_1[X]$, but it can be decomposed as $(X - \sqrt{2})(X + \sqrt{2})$ in $R_2[X]$.

4.2 Spectrum of a polynomial ring

Remark 20 In this section, we assume that R is a field.

Theorem 29 (Ideals of a polynomial ring) *Let R be a field, and I be a non-zero ideal of $R[X]$. Then there exists a unique monic polynomial $P \in R[X]$ such that $I = (P)$. As a consequence, all the ideals of $R[X]$ are principal.*

Proof: Since $I \neq 0$, there exists a non-zero $P \in I$ of minimal degree. Up to multiplication by a unit, we may assume P is monic. By definition, P and all its multiples belong to I. Conversely, if $Q \in I$, then by Theorem 28 there are unique A, B such that $Q = AP + B$ with $\deg B < \deg P$. But then $B \in I$, and if $B \neq 0$ then this contradicts that P is minimal. Hence $B = 0$, and $P|Q$, which means that $Q \in (P)$.

Now if there are polynomials $P_1, P_2 \in R[X]$ such that $I = (P_1) = (P_2)$, we have $P_2|P_1$ and $P_2|P_1$, hence $P_1 = \lambda P_2$ for some $\lambda \in R^\times$ (Proposition 6). But since P_1 and P_2 are monic, $\lambda = 1$ and $P_1 = P_2$. □

Corollary 17 (Greatest common divisor) *Let $A, B \in R[X]$, $A, B \neq 0$. Then*

$$I = \{AU + BV \mid U, V \in R[X]\}$$

is an ideal of $R[X]$, that is non-empty. Then by Theorem 29 there exists a unique monic polynomial $D \in R[X]$ such that $I = (D)$. We say that D is the greatest common divisor of A and B, and write $D = \gcd(A, B)$.

Definition 41 (Coprime polynomials) Two polynomials $A, B \in R[X]$ are coprime if $\gcd(A, B) = 1$, or equivalently, if there exist $U, V \in R[X]$ such that $AU + BV = 1$.

Remark 21 Theorem 29 relies critically on the assumption that R is not only a ring, but a field.

Definition 42 (Irreducible polynomial) A polynomial $P \in R[X]$ is said to be irreducible if $P \notin R^\times$, and if $Q|P$, $Q \neq P$ implies $Q \in R^\times$.

Example 20

- The polynomial $X^2 - 1$ is reducible in any ring, since it equals $(X - 1)(X + 1)$
- The polynomial $X^2 - 2$ is irreducible in $\mathbb{Z}[X]$ or $\mathbb{Q}[X]$, as otherwise there would be a rational (or integer!) square root of 2.
- The polynomial $X^p - 1$, where $p > 2$ is prime, is reducible since it has an obvious root. But it we remove this factor, we get

$$\Phi_p = \frac{X^p - 1}{X - 1} = 1 + X + \cdots + X^{p-1}$$

which is irreducible in $K[X]$ if K doesn't have roots of unity besides 1. Φ_p is known as the p-th cyclotomic polynomial.

Theorem 30 (Prime decomposition) *Let $P \in R[X]$, $P \neq 0$. Then P can be uniquely written (up to reordering) as*

$$P = u \prod_{P_i \text{ monic irred.}} P_i^{n_i}$$

where all but a finite number of the integers n_i are zero, and $u \in R^\times$.

Remark 22 A ring in which all the ideals are principal and there are no zero divisors is called a principal ideal domain. In particular, elements of such rings always admit a unique factorisation into "irreducible" elements, and a notion of GCD, similar to the situation in $\mathbb{Z}$.

Remark 23 The spectrum of $R[X]$, i.e., the prime ideals of $R[X]$, are (0) and (P_i), where P_i are the monic irreducible polynomials of $R[X]$.

Remark 24 If R is not a field, there can be other ideals. For instance, $\mathbb{Z}[X]$ is not a principal ideal domain. Its spectrum is composed of

- (0);
- (p) for each $p \in \mathfrak{P}$;
- (P) for each monic polynomial P irreducible in $\mathbb{Q}$ (hence, in $\mathbb{Z}$);
- (q, Q) for $q \in \mathfrak{P}$ and Q a monic polynomial irreducible in $\mathbb{F}_q$.

The following result is remarkably useful:

Theorem 31 (Eisenstein's criterion) *Let $P \in \mathbb{Z}[X]$, $P = c_nX^n + c_{n-1}X^{n-1} + \cdots + c_1X + c_0$. If there exists a prime p such that all the three following conditions hold:*

1. $p \nmid c_n$;
2. $p \mid c_{n-1}, c_{n-2}, \ldots, c_0$;
3. $p^2 \nmid c_0$;

then P is irreducible in $\mathbb{Q}[X]$, hence in $\mathbb{Z}[X]$.

Example 21 Let $X^3 - 6X + 3$. By Eisenstein's criterion (using $p = 3$) this polynomial is irreducible in $\mathbb{Q}$.

Proof: Assume that P is reducible, and write $P = gh$ with $\deg g$ and $\deg h$ strictly smaller than $\deg P$. Let $g = a_0 + \cdots + a_kX^k$ and $h = b_0 + \cdots + b_\ell X^\ell$.

Since $c_n = a_kb_\ell$, and $p \nmid c_n$, we have that $p \nmid a_k$ and $p \nmid b_\ell$. Similarly, p divides one of a_0 or b_0, but not both, since $p^2 \nmid c_0$; without loss of generality assume $p|a_0$.

Finally, $P = fg$ holds modulo p, i.e.,

$$c_nX^n \equiv P \equiv (a_0 + \cdots + a_kX^k)(b_0 + \cdots + b_\ell X^\ell) \bmod p$$

(we work in the ring $(\mathbb{Z}/(p))[X]$). The right-hand side of this equation would imply that, for any $m < n$, the coefficient of X^m is $a_mb_0 \not\equiv 0 \bmod p$. This contradicts the left-hand side. □

4.3 Roots of a polynomial

Definition 43 (Polynomial function) Let $P \in R[X]$, $P = a_0 + \cdots + a_kX^k$, and $x \in K$. The map $\mathrm{eval}_x : K[X] \to K$, defined by

$$\mathrm{eval}_x(P) = \sum_{k=0}^{\deg P} a_kx^k$$

is a ring morphism (the evaluation morphism). This defines an application $K \to K$ that we (abusively, but conventionally) denote P, and that we call the polynomial function associated to P.

Remark 25 Many polynomials will in often determine the same function. As an example, there are exactly 4 functions from $\mathbb{Z}/(2)$ to itself, while there are infinitely many polynomials in $(\mathbb{Z}/(2))[X]$.

Definition 44 (Root) Let $P \in R[X]$. An element $a \in R$ such that $P(a)$ is a root of P.

Lemma 13 *Let $P \in R[X]$ and $a \in R$. Then a is a root of P if an only if $(X-a)|P$.*

Remark 26 An irreducible polynomial may have no root: consider for instance $(X^1+1)^2$ which is reducible but has no real roots in $\mathbb{R}$.

Definition 45 (Derived polynomial) Let $P \in R[X]$, $P = a_0 + \cdots + a_k X^k$. The derived polynomial of P is

$$P' = \sum_{k=1}^{\deg P} k a_k X^{k-1}.$$

Remark 27 If $a \in R$ is a root of P but not a root of P', then a is called a simple root of P; in particular, $(X-a)^2 \nmid P$.

Lemma 14 *Let $P \in R[X]$, $\deg P = n$. Then P has at most n distinct roots.*

Proposition 7 *Let $p \in \mathfrak{P}$. Then every $a \in \{1, 2, \ldots, p-1\}$ is a root of the polynomial $X^{p-1} - 1 \in (\mathbb{Z}/(p))[X]$.*

Proof: This is a consequence of Fermat's little theorem. □

Corollary 18 (Wilson's theorem) *Let $p \in \mathbb{Z}$, $p > 1$. Then $p \in \mathfrak{P}$ if and only if $p|((p-1)!+1)$.*

Proof: Using the previous lemma, and the fact that $P = X^{p-1} - 1$ has degree $p-1$, we can write it as

$$P = \prod_{a=1}^{p-1} (X-a).$$

In particular, $P(0) = -1$ (left hand side) and $P(0) = \prod_{a=1}^{p-1}(-a)$ (right hand side), which gives the result.

Conversely, assume $p|((p-1)!+1)$. If $q|p$ and $q \neq p$, then $q|(p-1)!$ which gives $q = 1$; therefore p is prime. □

4.4 Quotients

Definition 46 (Module) Let R be a ring. A set E endowed with

- An Abelian group structure on E (denoted additively);
- An external law ("scalar multiplication") $\times : R \times E \to E$ such that for all $a, b \in R$ and all $x, y \in E$,

$$\begin{aligned}
a \times (x+y) &= a \times x + a \times y \\
(a+b) \times x &= a \times x + b \times x \\
a \times (b \times x) &= (ab) \times x \\
1 \times x &= x
\end{aligned}$$

is called an R-module. Since there is no ambiguity, we omit the symbol $\times$ for scalar multiplication.

Remark 28 If R is a field, then R-modules are typically called R-vector spaces.

Remark 29 If E is a ring and an R-module, it is called an R-algebra.

Example 22

- An Abelian group is exactly the same thing as a $\mathbb{Z}$-module.
- A commutative ring is exactly the same thing as a $\mathbb{Z}$-algebra.
- A polynomial ring $R[X]$ has a natural R-algebra structure, with scalar multiplication by "constants". As a result, we may refer to $R[X]$ as the polynomial algebra on R. It is an associative and commutative algebra.
- The ring of square $n \times n$ matrices with coefficients in R, $M_n(R)$, has a natural R-algebra structure.
- Let R be a ring and $Z(R)$ be its center, then R is an $Z(R)$-module.

Theorem 32 *Let R be a ring, and I be an ideal of $R[X]$. Then $R[X]/I$ is an R-algebra.*

Proof: As a ring quotient, $R[X]/I$ is a ring. Multiplication by an element of R gives it an R-algebra structure, as it it easily checked that this operation is compatible with the quotient. □

Theorem 33 (Units of a polynomial ring) *Let R be a field and $P \in R[X]$, $\deg P \geq 0$. Then $(R[X]/(P))^\times$ is made of the (classes of) polynomials coprime with P.*

Corollary 19 (Prime ideals of a polynomial ring) *The prime ideals of $R[X]$ are its irreducible polynomials.*

Theorem 34 *If P is irreducible, then $R[X]/(P)$ is an integral domain, and in fact, a field.*

Proof: We already know that the quotient of a ring by a prime ideal is an integral domain. Let show that any non-zero polynomial in $Q \in R[X]/(P)$ is invertible. Since P is irreductible and $P \nmid Q$ (otherwise Q would be zero in the quotient ring) P and Q are coprime polynomials. By Theorem 33 Q is therefore invertible. □

Theorem 35 *Let R be a field, and $P \in \operatorname{Spec} R[X]$. There exists a field K, such that*

- *R is a subfield of K;*
- *There is a root of P in K;*
- *K is an R-vector space*
- *As a vector space, K has finite dimension, and $\dim K = \deg P$.*

Proof: Denote $A = R[X]/(P)$; by Theorem 34, A is a field. Define

$$Z = \{a \in A \mid a \notin R + (P)\}$$

and $K = Z + R$. Then the application $f : A \to L$ defined for all $x \in R$ by

$$f(x) = x \bmod (P)$$

and for all $z \in Z$ by $f(z) = z$ is a bijection from A to K. This allows us to define the following operations on K: for every $u, v \in K$ and every $r \in R$, define

$$\begin{aligned} u + v &= f\left(f^{-1}(u) + f^{-1}(v)\right) \\ u \cdot v &= f\left(f^{-1}(u) \cdot f^{-1}(v)\right) \\ ru &= f(rf^{-1}(u)) \end{aligned}$$

which makes K an R-algebra; in particular, $A \simeq K$ as R-algebras. We conclude by showing that A, hence K, is an R-vector space of dimension $\deg P$.

Let $\beta = X + (P)$, and define the family $B = \{\beta^0, \beta^1, \dots, \beta^{\deg P - 1}\}$. Let $\lambda_0, \dots, \lambda_{\deg P-1} \in R$ such that

$$\sum_{i=0}^{\deg P-1} \lambda_i \beta^i = 0,$$

which in particular implies

$$\sum_{i=0}^{\deg P-1} \lambda_i X^i \in (P).$$

But P is of degree $\deg P$, so the λ_i are necessarily zero, therefore the elements of B are linearly independent.

Now let $Q \in R[X]/(P)$, there exists $S \in R[X]$ such that $Q = S + R$. Since P is non-zero, there exists by Theorem 28 $U, V \in R[X]$ such that $Q = PU + V$ and $\deg V < \deg P$. As a result, $Q = V \bmod (P)$, which shows that B is a generating family. □

Remark 30 The notion of polynomial ring extends the base ring. Then taking a quotient restricts the polynomial ring. For well chosen quotients, this enables us to build precisely controlled extensions. The process is usually as follows:

1. Start from some base ring or field K which you want to build an extension of.
2. Consider the corresponding polynomial ring $K[X]$
3. Choose an irreducible polynomial of $K[X]$, and consider the quotient $K[X]/(P)$.

Remark 31 Taking quotients is not a harmless operation. For instance, in the ring $\mathbb{Z}[\sqrt{-5}] \simeq \mathbb{Z}[X]/(X^2+5)$,

$$(1 - \sqrt{-5})(1 + \sqrt{-5}) = 6 = 2 \cdot 3$$

and both $2, 3, (1 - \sqrt{-5})$, and $(1 + \sqrt{-5})$ are irreducible; so this means we do not have unique factorisation! That being said, the ideal (6) can be uniquely decomposed as a product of prime ideals (which ones?). This remarkable feature was Kummer's motivation for introducing ideals in the first place.

Example 23

- $\mathbb{R}[X]/(X^2+1) \simeq \mathbb{R}[i] \simeq \mathbb{C}$ is a possible definition of the field of complex numbers, and $\dim_{\mathbb{R}} \mathbb{C} = \deg(X^2+1) = 2$. (Which tells us, for instance, that any complex number is entirely described by a couple (x, y) of reals).

- Every polynomial is reducible over $\mathbb{C}$, so we cannot construct algebraic extensions of $\mathbb{C}$ that are not $\mathbb{C}$ itself.
- If P is an irreducible polynomial of degree n in some polynomial ring $K[X]$, then it is often practical to introduce a formal root α of P and think of $K[X]/(P)$ as an extension $K[\alpha, \alpha^2, \ldots, \alpha^{n-1}]$ which is a concrete realisation of the corresponding K-algebra. For instance, $\mathbb{Z}[X]/(X^2 + D) \simeq \mathbb{Z}[\sqrt{-D}]$, where $\sqrt{-D}^2 = -D$.
- It is possible to construct extensions of extensions. For instance, $X^2 - 2$ is irreducible in $\mathbb{Q}[X]$, and we construct from it $\mathbb{Q}(\sqrt{2}) \simeq \mathbb{Q}[X]/(X^2-2)$. Now X^2-3 is irreducible in that new field, and we can get $\mathbb{Q}(\sqrt{2}, \sqrt{3}) = \mathbb{Q}(\sqrt{2})[\sqrt{3}] \simeq \mathbb{Q}(\sqrt{2})[X]/(X^2 - 3)$. In particular, this shows that $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ has dimension 4 over $\mathbb{Q}$.
- Let $p \in \mathfrak{P}$, $p \neq 2$, then the p-th cyclotomic polynomial is monic and irreducible in $\mathbb{Q}[X]$. Therefore $\mathbb{Q}[X]/(\Phi_p)$ is a field, called the p-th cyclotomic field. If we introduce a formal p-th root of unity $\zeta \neq 1$, then $\mathbb{Q}[X]/(\Phi_p) \simeq \mathbb{Q}[\zeta]$, which we write $\mathbb{Q}(\zeta)$ to insist that it is a field. We can similarly define $\mathbb{Z}[\zeta]$.

HISTORICAL NOTE. The modern theory of rings was born as an effort to capture various observations about burgeoning number theory and the study of polynomials; the axiomatic definition in use today is due to Emmy Noether who published her foundational work in 1921, whereas the name "ring" is due to Hilbert (in German, *Zahlring*).

One of the driving forces behind the development of ring theory was the question of factorisation in "extended" number systems, i.e. number fields, which would readily solve many problems in a way similar to the introduction of complex numbers. This question asked for a generalisation of "prime" numbers, which is captured by prime ideals (ideals usually behave better than numbers).

Another leap was made when Lasker and Macauley found that for any variety (e.g., the circle) one can associate a ring and an ideal (e.g. the ideal $(x^2 + y^2)$ of $\mathbb{R}[x, y]$). In doing so, geometry can be translated into facts about different kinds of ideals in polynomial rings. This correspondence was extended by Noether: to any algebraic variety corresponds a ring, and many difficult geometric problems can be readily translated into easy algebraic facts (e.g. irreducibility).

Grothendieck's ambitious programme in the 1960s has been to claim the equivalence: to any ring, corresponds a geometric object: a scheme. Some schemes can be visualised as usual varieties (e.g., the circle), but many have no obvious representation at all (e.g., $\operatorname{Spec} \mathbb{Z}[\sqrt{3}]$). Extending the notion of "what geometry is" enabled major advances in the treatment and understanding both of algebraic, geometric, and arithmetic problems. This fruitful intuition is at the foundation of modern algebraic geometry: it was for instance essential to proving the Fermat–Wiles theorem.

4.5 Exercise set

Exercise 4.1. Let $\mathbb{F}_2 = \mathbb{Z}/(2)$. Compute the Euclidean division of $X^3 + X^2 + 1$ by $X^2 + X + 1$ in $\mathbb{F}_2[X]$.

Exercise 4.2. Let $m, n \in \mathbb{N}$, $m, n > 0$. Show that $\gcd(X^n - 1, X^m - 1) = X^{\gcd(n,m)} - 1$.

Exercise 4.3. Let $p \in \mathfrak{P}$ and $\mathbb{F}_p = \mathbb{Z}/(p)$. How many monic polynomials are there in $\mathbb{F}_p[X]$? Show that there are $p(p-1)/2$ irreducible monic polynomials of degree 2 in that ring.

Exercise 4.4. Compute $\operatorname{Spec} \mathbb{R}[X]$, $\operatorname{Spec} \mathbb{C}[X]$, $\operatorname{Spec} \mathbb{F}_q[X]$.

Exercise 4.5. Show that if $p \in \mathfrak{P}$, $W(p) = ((p-1)! + 1)/p$ is an integer. If $p | W(p)$ then p is called a Wilson prime.

Open question (2018): Are there other Wilson primes besides 5, 13, and 563?

Exercise 4.6. Let $\mathbb{F}_2 = \mathbb{Z}/(2)$, and $P = X^3 + X + 1 \in \mathbb{F}_2[X]$.

1. Show that P is irreducible.
2. What is the dimension of the the field $\mathbb{F}_2[X]/(P)$?
3. How many elements does the field $\mathbb{F}_2[X]/(P)$ contain?
4. Let α be the class of X modulo (P), show that

   a) $P(\alpha) = 0$
   b) $\alpha^4 = \alpha^2 + \alpha$
   c) α^2 is a root of $X^2 + \alpha X + 1 + \alpha^2$
   d) α^4 is a root of $X^2 + \alpha X + 1 + \alpha^2$

   Conclude that
   $$P = (X + \alpha)(X + \alpha^2)(X + \alpha + \alpha^2)$$

Exercise 4.7. The sequence of Fibonacci numbers is defined by $F_0 = F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$.

1. Show that there is a closed-form formula for F_n, given by
   $$F_n = \frac{\phi^n - \overline{\phi}^n}{\sqrt{5}},$$
   where $\phi = (1 + \sqrt{5})/2$ is the golden ratio, and $\overline{\phi} = (1 - \sqrt{5})/2 = -1/\phi$.
2. What is the approximate binary length of F_n?
3. Write a program computing F_n using this formula. What is a concrete issue with this approach? What happens for large values of n?

4. Show that
$$\begin{pmatrix} F_{n+2} \\ F_{n+1} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}.$$
and therefore that
$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n = \begin{pmatrix} F_{n+1} & F_n \\ F_n & F_{n-1} \end{pmatrix}.$$
Write a program that uses this equality to compute the exact value of F_n.

5. Use the above equation to prove Cassini's identity:
$$F_{n+1}F_{n-1} - F_n^2 = (-1)^n.$$

6. Show that for all $n \geq 1$,
$$F_{2n-1} = F_n^2 + F_{n-1}^2$$
$$F_{2n} = (F_{n-1} + F_{n+1})F_n = (2F_{n-1} + F_n)F_n = 2F_{n-1}F_n + F_n^2.$$
Use this to write a program computing the exact value of F_n. Using memoization (i.e., keeping track of already-computed values), this is faster than computing F_n using the matrix formula, and much faster than using the definition.

7. Let $n \in \mathbb{N}$, $n > 1$. Show that $F_k \bmod n$ is a periodic sequence. The period of $F_k \bmod n$ is known as the Pisano period. Write a program that computes the Pisano period for small values of n; this gives: $1, 3, 8, 6, 20, 24, \ldots$ Observe (or better, prove) that the period mod mn is the least common multiple of the period mod m and the period mod n (Hint: it suffices to consider prime powers, and invoke the CRT).

Exercise 4.8. The goal of this exercise is to prove Mason's theorem. For any polynomial $P \in \mathbb{C}[X]$, define its radical as
$$\operatorname{rad} P(X) = \prod_{\alpha \text{ s.t. } P(\alpha)=0} (X - \alpha) \in \mathbb{C}[X].$$

1. Show that if $P \in \mathbb{C}^*$, then $\operatorname{rad} P = 1$.

2. Let $A, B, C \in \mathbb{C}[X]$ non zero and pairwise coprime polynomials, satisfying
$$A + B + C = 0.$$
Show that
$$\begin{vmatrix} A' & B' \\ A & B \end{vmatrix} = \begin{vmatrix} B' & C' \\ B & C \end{vmatrix} = \begin{vmatrix} C' & A' \\ C & A \end{vmatrix}.$$

3. Deduce from this that if $AB \neq C$, then
$$\deg\left(\frac{ABC}{\operatorname{rad} ABC}\right) < \deg(AB).$$

4. Show that if $\deg(ABC) > 0$ then
$$\max(\deg A, \deg B, \deg C) < \deg \operatorname{rad} ABC.$$

5. Conclude that for $n \geq 3$, the curve $x^n + y^n + z^n = 0$ cannot be parametrised by elements of $C[X]$

Exercise 4.9. A party trick consists in the following: ask a challenger to choose a polynomial $P \in \mathbb{Z}[X]$, and keep it secret from you. Your goal is to guess P, by querying your challenger with values x, and having her reply with the value $P(x)$. Show that you only need two queries. What if she had chosen a polynomial in $\mathbb{Q}$? In $\mathbb{F}_p$ for given p?

Exercise 4.10. Let $P \in \mathbb{Z}[X]$ of degree n, $P = a_n X^n + \cdots + a_1 X + a_0$, and $H = \max_{i \neq n} |a_i / a_n|$. We want to show the following result:

> *If there exists $N \in \mathbb{N}$, $N > H + 1$ such that $P(N) \in \mathfrak{P}$, then P is irreducible in $\mathbb{Z}[X]$.*

1. Assume that P has a root α (one may consider working in $\mathbb{C}[X]$). Show that $|\alpha| < H + 1$.
2. Assume that P is reducible. Show that therefore $P = QR$ satisfies $R(N) = \pm 1$ and write
$$R = c \prod_i (X - \alpha_i)$$
with $\alpha_i \in \mathbb{Z}$ and c is the leading coefficient of R. Use the previous result to show that $|R(N)| > 1$. Conclude.
3. Consider the polynomial $P = X^4 + 6X^2 + 1$. Show that one cannot use Eisenstein's criterion on P; show that $P(8)$ is prime, and therefore that P is irreducible in $\mathbb{Z}[X]$.
4. Consider the polynomial $P = (X - 9)(X^2 + 1)$, show that $P(10) = 101 \in \mathfrak{P}$. P is clearly reducible, what does this say about the statement at hand?
5. There exists a slightly stronger version of this result, known as Cohn's theorem: if P is constructed from that base 10 representation of a prime, such as
$$p = 65537 \in \mathfrak{P} \mapsto P_p = 6X^4 + 5X^3 + 5X^2 + 3X + 7$$
then P_p is irreducible in $\mathbb{Z}[X]$. Show that there exist at least a number $q \notin \mathfrak{P}$ such that P_q is irreducible.

Exercise 4.11. In this exercise we consider a fixed polynomial $P \in \mathbb{C}[Z]$ of degree $r > 0$.

1. Show that P' has at most $r - 1$ roots, and that therefore there is only a finite set of points S where the rational fraction $1/P'$ is not defined.
2. Consider the rational fraction $Q(z) = z - P(z)/P'(z)$, which is also defined on $\mathbb{C} - S$. Show that if $z \notin S$ and z is a root of P, then z is a fixed point of z.
3. Let $z_0 \in \mathbb{C} - S$, define the sequence (z_n) as $z_{n+1} = Q(z)$. Show that if this sequence converges, then its limit is a root $R(z_0)$ of P. Are there situations in which the sequence (z_n) does not converge? Do we need to know the roots of P to detect convergence?
4. Let $\epsilon > 0$, and consider a value z_0 for which the above sequence converges. Let $N_\epsilon(z_0)$ be the smallest integer such that for any $n > N_\epsilon(z_0)$, $|R(z_0) - z_n| \leq \epsilon$. We say that z_0 "Newton converges to $R(z_0)$ after $N_\epsilon(z_0)$ iterations".
5. Implement a program that takes a polynomial P, a value z_0, and a value ϵ as inputs, and returns $R(z_0)$ and $N_\epsilon(z_0)$.
6. Generate a picture such as Figure 4.2, where each colour corresponds to $R(z_0)$, as z_0 spans a portion of $\mathbb{C}$. This fractal is called the Newton attractor of P.

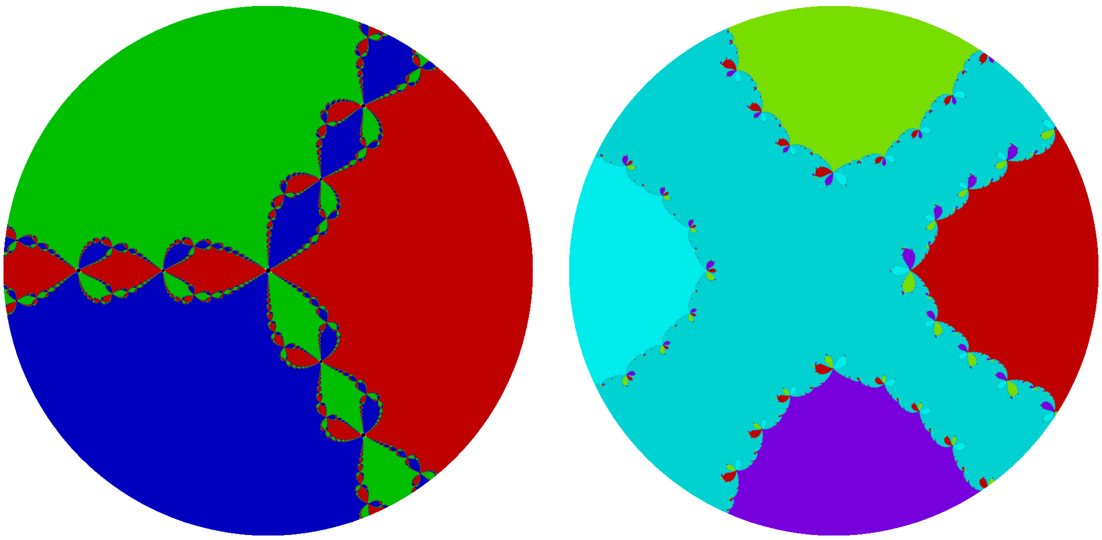


Figure 4.2: Newton attractor for the polynomials $P_1 = X^3 - 1$ (left), and $P_2 = (X - \frac{1}{2})(X^4 - X - 2)$ (right), over the complex disc of radius 4. The colours correspond to the root's angle.

Chapter 5

Finite fields

We met finite fields *en passant* as a possible byproduct of certain ring quotients. In this relatively short chapter, we develop a systematic treatment of these structures, and in particular how to construct them. Finite fields are of central importance to many algebraic constructions (e.g., in cryptography).

TODO

Figure 5.1: TODO: illustration.

Contents

5.1 Finite fields of prime order

Recall that by Wedderburn's theorem (Theorem 22) every finite integral domain is a field. Thus in particular every finite field is a (commutative) field.

Lemma 15 (Characteristic of an integral domain) *Let A be an integral domain, then either* $\operatorname{char} A = 0$ *or* $\operatorname{char} A \in \mathfrak{P}$.

Proof: Let $f : \mathbb{Z} \to A$ the application defined by $f(m) = 1m$. It is a ring morphism, and therefore its kernel is an ideal of $\mathbb{Z}$. Since $\mathbb{Z}$ is a principal ideal domain, this means that there exists a unique $n \in \mathbb{Z}$ such that $\ker f = (n)$ — by definition, $n = \operatorname{char} A$.

As a result, $\mathbb{Z}/(n)$ is isomorphic to a subring of A, and is therefore an integral domain.

If n is non-zero, $\mathbb{Z}/(n)$ is a finite integral domain, hence it is a field (Theorem 22). Therefore $n \in \mathfrak{P}$. □

Lemma 16 *If K is a field of characteristic zero, then it contains a subfield that is isomorphic to $\mathbb{Q}$ (which is infinite).*

Proof: If K is of characteristic zero, then the map $\mathbb{Q} \to K$ defined by $a/b \mapsto a1(b1)^{-1}$ is a field morphism. Thus its image is a subfield of K which is isomorphic to $\mathbb{Q}$. □

Lemma 17 (Prime subfield) *If K is a field of characteristic p, then it contains a subfield that is isomorphic to $\mathbb{F}_p = \mathbb{Z}/(p)$. This subfield is called the prime subfield of K.*

Proof: We have $m1 = 0$ if and only if $p|m$; the image of the map $m \mapsto m1$ is thus a subfield of K, isomorphic to $\mathbb{F}_p$. □

Corollary 20 (Characteristic of finite fields) *A finite field has prime characteristic.*

Corollary 21 (Order of finite fields) *Let K be a finite field. There exists an integer $n > 0$ and a prime $p \in \mathfrak{P}$ such that K has p^n elements.*

Proof: Let $p = \operatorname{char} K$. By the above lemma, K contains a subfield isomorphic to $\mathbb{F}_p$. This gives K a natural $\mathbb{F}_p$-vector space structure, and since K is finite, its dimension n as a vector space is also finite. Therefore as vector spaces $K \simeq \mathbb{F}_p^n$, and $|K| = p^n$. □

5.2 Finite fields of prime power order

Theorem 36 (Constructing finite fields of prime power order) *Let K be a finite field with p^n elements. Then there exists an irreductible polynomial $F \in \mathbb{F}_p[X]$ of degree n such that $K \simeq \mathbb{F}_p[X]/(P)$.*

Proof: Let α be a generator of $K^\times$, and define the application $f : \mathbb{F}_p[X] \to K$ by

$$f\left(\sum a_i X^i\right) = \sum a_i \alpha^i.$$

Then f is a ring morphism, and it is surjective because $\langle\alpha\rangle = K^\times$. The kernel of f is an ideal I of $\mathbb{F}_p[X]$, and therefore $\mathbb{F}_p[X]/I \simeq K$ as rings. Note that I is non-zero, because this would imply $K \simeq F_p[X]$ which is not a field. Therefore there exists $F \in \mathbb{F}_p[X]$ such that $I = (F)$; since $\mathbb{F}_p[X]/(P)$ is a field, the polynomial P is irreducible. Finally, the degree of P must be n so that the size of K matches p^n. $\square$

Remark 32 What the above entails is that there exists a finite field with p^n elements if, and only if, there exists an irreducible polynomial of degree n in $\mathbb{F}_p[X]$. We will admit the following result.

Theorem 37 (Möbius) *Let $\mu : \mathbb{Z} \to \{-1, 0, 1\}$ be defined as*

$$\mu(n) = \begin{cases} (-1)^r & \text{if } n \text{ is the product of } r \text{ distinct prime numbers} \\ 0 & \text{otherwise} \end{cases}$$

then the number of irreductible polynomials of degree d in $\mathbb{F}_p[X]$ is

$$\frac{1}{n}\sum_{d|n} \mu(d) p^{\frac{n}{d}}.$$

In particular, for every $n \geq 1$ and every $p \in \mathfrak{P}$, this quantity is strictly positive.

Corollary 22 *For every $n > 0$ and $p \in \mathfrak{P}$, there exists a finite field with p^n elements.*

Remark 33 Using the formula from Theorem 37, we see that there are 22517997465744 ways to construct a field of size 2^{50}. As we will now see, they are all isomorphic to one another.

Theorem 38 (Fields of same order are isomorphic) *If K and L are finite fields and $|K| = |L|$, then $K \simeq L$.*

Proof: We have $|K| = p^n$ with $p \in \mathfrak{P}$ by Corollary 21. Let $\alpha \in K^\times$ of order $p^n - 1$ (such an element necessarily exists) and consider the evaluation morphism $\mathfrak{a}$ which maps polynomials in $\mathbb{F}_p[X]$ to K by mapping X to α. This is a surjective morphism, and its kernel is an ideal, which can be generated by a monic irreducible polynomial $P \in \mathbb{F}_p[X]$. Then $P|(X^{p^n} - X)$.

Now if $|L| = |K| = p^n$, we have for every $\ell \in L$, $\ell^{p^n} = \ell$, i.e., the sets of roots of the polynomial $X^{p^n} - X$ is exactly L. If r is any root of P, the evaluation morphism at r factors through $\mathbb{F}_p[X]/(P)$ and induces an isomorphism $K \simeq L$. $\square$

Remark 34 We will henceforth talk of "the" finite field of size p^n, and write $\mathbb{F}_{p^n}$ to denote any construction of this field. The finite fields $\mathbb{F}_q$ are sometimes referred to as Galois fields, and denoted $GF(q)$.

Remark 35 Beware! The field $\mathbb{F}_2$ is isomorphic to $\mathbb{Z}/(2)$, but the field $\mathbb{F}_8 = \mathbb{F}_{2^3}$ is not isomorphic to $\mathbb{Z}/(8)$, which is not a field!

Remark 36 Although all finite fields are isomorphic to some $\mathbb{F}_q$, this isomorphism is not unique!

Example 24

- $\mathbb{F}_2 \simeq \mathbb{Z}/(2)$
- $\mathbb{F}_4 \simeq \mathbb{F}_2[X]/(X^2 + X + 1)$

- $\mathbb{F}_8 \simeq \mathbb{F}_2[X]/(X^3+X+1)$
- $\mathbb{F}_{16} \simeq \mathbb{F}_2[X]/(X^4+X+1)$; the polynomial $X+2+X$ generates a subfield of order 4.
- $\mathbb{F}_9 \simeq \mathbb{F}_3[X]/(X^2+X-1)$

Remark 37 It is typical to represent a finite field element in $\mathbb{F}_q \simeq \mathbb{F}_p[X]/(P)$ by introducing a formal root α of P. This in turn allows for a compact machine representation of finite field elements as sequence of elements of $\mathbb{F}_p$. For instance, in $\mathbb{F}_4$ using the isomorphism $\mathbb{F}_4 \simeq \mathbb{F}_2[X]/(X^2+X+1)$,

Element	With a formal root	Implemented as
0	0	$(0,0)$
1	1	$(0,1)$
X	α	$(1,0)$
$X+1$	$\alpha+1$	$(1,1)$

This representation is practical, but care should be taken when performing operations; as an example, $X \times X = X^2 = -X-1 = X+1$, which means $(1,0)\times(1,0) = (1,1)$. In the common case that $p=2$, one can represent elements directly in binary, e.g., $1+X$ will be represented as 11.

Theorem 39 (Subfields of $\mathbb{F}_{p^n}$) *Let $p \in \mathfrak{P}$ and $n > 0$. If K is a subfield of $\mathbb{F}_{p^n}$ then $K \simeq \mathbb{F}_{p^k}$ with $k|n$.*

Proof: Since K is a subfield of a finite field, it is finite, and therefore it is of order p^k for some k. By Lagrange's theorem (Theorem 11) if $\mathbb{F}_{p^k}$ is a subfield of $\mathbb{F}_{p^n}$, then $p^n = (p^k)^r$ for some integer r, and thus $k|n$.

Conversely, assume $k|n$, i.e. $n = kr$ for some integer r, then $p^n - 1 = (p^r - 1)\left(1+\cdots+(p^r)^{k-1}\right) = (p^r-1)N$. Now

$$\begin{aligned} X^{p^n} - X &= X\left(X^{p^{rk}-1} - 1\right) \\ &= X\left(X^{p^r-1}-1\right)\left(1+\cdots+X^{(p^r-1)(N-1)}\right). \end{aligned}$$

But since the roots of $X^{p^r} - X$ is exactly $\mathbb{F}_{p^r}$, it is contained in the field $\mathbb{F}_{p^n}$. □

5.3 Multiplicative group of finite fields

Theorem 40 (Multiplicative groups of finite fields) *The group $\mathbb{F}_q^\times$ is cyclic, of order $q-1$.*

Definition 47 (Primitive element) There are $\varphi(q-1)$ many choices for the generator of the group $\mathbb{F}_q^\times$. Such a generator is called a primitive element of $\mathbb{F}_q$.

Example 25 The field $\mathbb{F}_{2^5}$ has 32 elements, and the group $\mathbb{F}_{2^5}^\times$ is cyclic of order $2^5 - 1 = 31 \in \mathfrak{P}$ elements (this is a Mersenne prime); it is therefore isomorphic to $\mathbb{Z}/(31)$, and has 30 generators.

Remark 38 If α is a primitive element of $\mathbb{F}_q^\times$, then $\{\alpha, \alpha^2, \ldots, \alpha^{q-1}\}$ is a permutation of the elements in $\mathbb{F}_q^\times$. Given an element $x \in \mathbb{F}_q^\times$, an integer y such that $\alpha^y = x$ is called a discrete logarithm of x in base α.

Example 26 Let $p \in \mathfrak{P}$ and g be a primitive element of $\mathbb{F}_p^\times$. Consider the following game, played between Alice and Bob:

1. Alice picks a number $a \in \mathbb{F}_p^\times$ and computes $A = g^a$;
2. Bob picks a number $b \in \mathbb{F}_p^\times$ and computes $B = g^b$;
3. Alice send A to Bob ; Bob sends A to Alice; (over a possibly insecure channel)
4. Alice computes $K_A = B^a$;
5. Bob computes $K_B = A^b$.

Observe that $K_A = B^a = (g^b)^a = g^{b+a} = g^{a+b} = (g^a)^b = A^b = K_B$. Hence, Alice and Bob both know $K = K_A = K_B$. This is the principle of the Diffie–Hellman key exchange protocol (DHE).

If an eavesdropper captures the conversation between Alice and Bob, i.e., if A and B are known to the adversary, then computing K from this information is challenging — in fact, it seems there is no substantially faster way than solving the discrete logarithm problem in base g in $\mathbb{F}_p^\times$.

The security of Diffie–Hellman key exchange, as employed today to secure Internet connections, relies on the assumption that this is intractable for large enough primes p (or for large enough groups that we use instead of $\mathbb{F}_p^\times$, such as the groups of rational points of an elliptic curve).

Lemma 18 (Order of elements) *Since $\mathbb{F}_q^\times$ is cyclic of order $q-1$, for every $d|(q-1)$ there are $\varphi(d)$ elements of multiplicative order d. In particular, there are $\varphi(q-1)$ primitive elements.*

Example 27 Consider $\mathbb{F}_{16} \simeq \langle \alpha \rangle$ with α of order 15. There are

- $\varphi(1) = 1$ element of order 1 (namely, 1);
- $\varphi(3) = 2$ elements of order 3 (namely, α^5 and α^{10});
- $\varphi(5) = 4$ elements of order 5 (namely, α^3, α^6, α^9, and α^{12});
- $\varphi(15) = 8$ primitive elements, of order 15 (namely α, α^2, α^4, α^7, α^8, α^{11}, α^{13}, and α^{14}).

5.4 Irreducible polynomials

We already discussed irreducibility of polynomials in the general context; polynomials with coefficients in a finite field deserve a slightly special treatment, not the least because arbitrary maps between finite fields can always be thought of as polynomials.

5.4.1 Product of all monic irreducible polynomials

We make use of the following technical lemma in some proofs:

Lemma 19 *Let K be a field and L be an extension of K. Let $P, Q \in K[X]$ with P irreducible. If there exists $\alpha \in L$ such that $P(\alpha) = Q(\alpha) = 0$ then $P \mid Q$.*

Proof: If we had $P \nmid Q$ then these polynomials would be coprime and Bézout's theorem for polynomials would lead to a contradiction. Hence $P \mid Q$. □

The following elementary lemma, due initially to Gauss, is of importance:

Lemma 20 (Product of irreducibles) *Let $q = p^k$ be a prime power, and $n > 1$. Then the polynomial $P = X^{q^n} - X \in \mathbb{F}_q[X]$ is the product of all monic irreducible polynomials $Q \in \mathbb{F}_q[X]$ such that $\deg Q \mid n$.*

Proof: P has no repeated factor (its derivative is -1) so it is sufficient to show that $Q \mid P$ iif Q is irreducible and $\deg Q \mid n$.

Assume Q is irreducible of degree d, and let $K = \mathbb{F}_q[X]/(Q)$. Note that X is a root of Q in K and also of $X^{q^d} - X$ in K, so by Lemma 19 we have $Q \mid X^{q^d} - X$. Thus, if $d \mid n$, we have $Q \mid P$.

Conversely, assume $Q \mid P$. The map $a \mapsto a^{q^n}$ is an automorphism of K, and in particular the set of its fixed points is a subfield of K. Since X is a fixed point, in fact the set of fixed points is K itself. Thus for any $a \in K$, we have $a^{q^n-1} = 1$. Thus, the order of a (i.e., $q^d - 1$) divides $q^n - 1$, and therefore $d \mid n$. □

We will make use of this result later to design an algorithm for polynomial irreducibility.

5.4.2 Minimal polynomial

Let K be a finite field of characteristic p, then it has a prime subfield isomorphic to $\mathbb{F}_p$ by Lemma 17. We admit the following result (which is an elementary result of Galois theory):

Theorem 41 *Let $q = p^k$ be a prime power. Every element of $\mathbb{F}_q$ is algebraic over $\mathbb{F}_p$.*

In other terms, for any $a \in bbFq$, there exists a non-zero polynomial $f_a \in \mathbb{F}_p[X]$ such that $f_a(a) = 0$.

Definition 48 (Minimal polynomial) The monic polynomial of $\mathbb{F}_p[X]$ of smallest degree having root a is called the minimal polynomial of a.

In particular, this polynomial is irreducible (and therefore generates a prime ideal). Conversely, any irreducible monic polynomial is the minimal polynomial for some element in an extension. Furthermore the degree of the minimal polynomial of an element in $\mathbb{F}_{p^n}$ is at most n.

Example 28 Consider $\mathbb{F}_3[i]$. We can easily find the minimal polynomials for all its elements:

$$\begin{array}{rl}
0 & X \\
1 & X-1 \\
-1 & X+1 \\
\pm i & X^2+1 \\
1 \pm i & X^2+X-1 \\
-1 \pm i & X^2-X-1
\end{array}$$

Finding the minimal polynomial of a given element has interesting applications, so we give another example.

Example 29 Let $q = 2^5$, and let α be the root of the polynomial $X^5 + X^2 + 1$. Say we are looking for the minimal polynomial of $\beta = a^3$.

We know that the degree of that polynomial will be at most 5, and therefore we only need to compute $\beta^0, \ldots, \beta^5$, and look for a linear relationship between these values, i.e., numbers $\lambda_i \in \mathbb{F}_2$ such that

$$\sum_{i=0}^{5} \lambda_i \beta^i = 0.$$

Concretely, we express the powers of β in terms of α and use the simplification that $\alpha^5 + \alpha^2 + 1 = 0$, then turn the equation above in a linear system in the powers of α; the appropriate coefficients λ_i make the α powers disappear.

We find $\lambda_2 = 0$ and all other coefficients $\lambda_i = 1$. Thus the minimal polynomial of β is $1 + X + X^3 + X^4 + X^5$.

HISTORICAL NOTE. The notion of field was introduced by Galois (as he was working on quotients of polynomial rings); Dedekind called them *Körper*, whereas Moore used the English *field*. The formal definition of a field is due to Weber (1893), and the structure of finite fields was studied by Dickson around 1905.

The context of finite fields makes the study of many problems much simpler. In 1960, Bose and Ray-Chaudhuri show that finite fields of characteristic 2 are a natural context for information transmission. Their work, together with Hocquenghem, led to the BCH error-correcting codes (which we'll study later). In 1974, the Weil conjectures (a finite field analogue of the still open Riemann hypothesis) were proven by Dwork, Grothendieck, and Deligne, and had a profound effect on mathematics. Finally, the study of elliptic curves over finite fields for cryptography was suggested in 1985 by Koblitz and Miller.

5.5 Exercise set

Exercise 5.1. Let $P = X^4 + X + 1$ in $\mathbb{F}_2[X]$.

1. Show that P is irreducible, and therefore that $\mathbb{F}_{16} \simeq \mathbb{F}_2[X]/(P)$. Can you find another irreducible polynomial of degree 4 in $\mathbb{F}_2[X]$?

2. Denote by α a formal root of P in $\mathbb{F}_{16}$. Write the multiplication table of $\mathbb{F}_{16}$.

3. The Galois counter mode (GCM) is an efficient and provably secure mode of operation for cryptographic block ciphers for use in authenticated encryption. It relies on multiplications in the finite field $\mathbb{F}_{2^{128}} \simeq \mathbb{F}_2[X]/(X^{128} + X^7 + X^2 + X + 1)$.

   a) Compute 2^{128}.

   b) Write a program that computes additions and multiplications in $\mathbb{F}_{2^{128}}$. You may represent the field elements as hexadecimal numbers.

   c) Show that multiplication in $\mathbb{F}_{2^{128}}$ can be computed in parallel.

   Since 2010, Intel and AMD microprocessors include a dedicated `PCLMULQDQ` for fast operations of this kind.

Exercise 5.2. Let $R = X^8 + X^4 + X^3 + X + 1 \in \mathbb{F}_2[X]$ (also known as the Rijndael polynomial, as it appears in the design of the cryptographic block cipher of the same name, better known as the AES). Show that $\mathbb{F}_2[X]/(R) \simeq \mathbb{F}_{256}$. Let $X = \mathtt{53}$ and $Y = \mathtt{CA}$ where X and Y are written in hexadecimal (base 16) representation. Show that $XY = \mathtt{01} = 1$, i.e., Y is the inverse of X.

Exercise 5.3. Let $p \in \mathfrak{P}$, $q = p^k$ for some integer $k > 0$, and an integer $m > 0$. Let $\alpha \in \mathbb{F}_{q^m}$. Show that the trace and norm of α are given by

$$\mathrm{Tr}_{\mathbb{F}_{q^m}/\mathbb{F}_q}(\alpha) = \sum_{i=1}^{m} \alpha^{q^i} \quad \text{and} \quad N_{\mathbb{F}_{q^m}/\mathbb{F}_q}(\alpha) = \prod_{i=1}^{m} \alpha^{q^i}.$$

Exercice 5.4. Let $q = p^k$ be a power prime, and $\mathbb{F}_q$ be the finite field with q elements. Let $\sim$ be the relation defined by

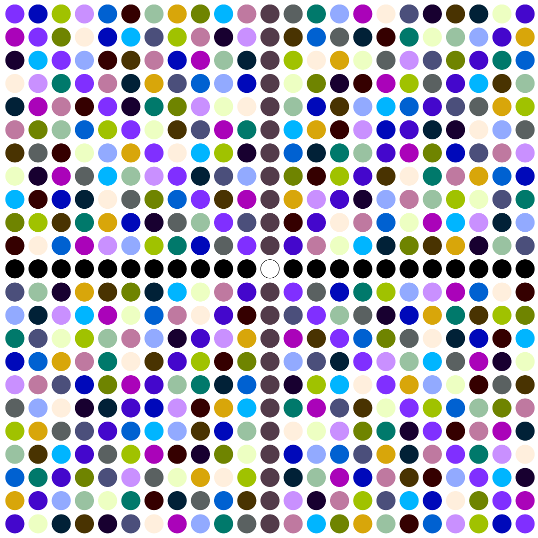
$$x \sim y \qquad \Leftrightarrow \qquad \exists u \in \mathbb{F}_q^\times, ux - y = 0$$

1. Show that $\sim$ defines an equivalence relation.

2. Let $\mathbb{P}^1(\mathbb{F}_q) = (\mathbb{F}_q \times \mathbb{F}_q)/\sim$, the set of projective lines over $\mathbb{F}_q$. Compute $\mathbb{P}^1(\mathbb{F}_3)$.

3. Show that $\#\mathbb{P}^1(\mathbb{F}_q) = q + 1$.

4. Write a computer program that computes $\mathbb{P}^1(\mathbb{F}_p)$ for primes p. The following picture shows the map $\mathbb{F}_{23}^2 \to \mathbb{P}^1(\mathbb{F}_{23})$, where each colour corresponds to an equivalence class:



What is your algorithm's complexity?

5. Compute $\mathbb{P}^1(\mathbb{F}_4)$ and $\mathbb{P}^1(\mathbb{F}_6)$. How could they be represented graphically?

Exercise 5.5. Let $q = 16$ and γ be a root of $P = x^4 + x + 1$ in $\mathbb{F}_q$.

1. Is P the minimal polynomial of γ?
2. Compute γ^k for $k = 0$ to 15.
3. Find the minimal polynomials of γ^3, γ^4, and γ^5.

Exercise 5.6. Establish the multiplication table of $\mathbb{F}_8 \simeq \mathbb{F}_2[x]/(x^3 + x + 1)$. Compute its primitive elements. Find the minimal polynomials for them.

Exercise 5.7. Let $q = p^k$ be a prime power.

1. Let $y_1, y_2 \in \mathbb{F}_q$. Show that there is a unique polynomial P of degree 2 such that $P(1) = x_1$ and $P(2) = x_2$. Give an explicit formula for P.
2. It is possible to use the above result to perform secret sharing: two users (labeled U_1 and U_2 respectively) receive each a share x_1 and x_2 respectively. The "secret" is the value $P(0)$, and can be reconstructed if the users reveal their share to one another. What happens if a user only known one share?
3. Show that given any $s \in \mathbb{F}_q$, we can easily generate shares x_1, x_2 such that $P(0) = s$.
4. Generalise the above construction to n shares $x_1, \ldots, x_n$.
5. Modify this construction so that the secret is recovered provided at least k out of n shares are known.
6. Implement and test this algorithm (initially due to Shamir).

Chapter 6

Quadratic residues and reciprocity

This chapter is concerned with the computation of "square roots" in finite fields and rings. Such roots do not always exist, but when they do there are algorithms to compute them. There are abundant uses of the notions introduced in this chapter, both theoretical and practical.

TODO

Figure 6.1: TODO: illustration.

Contents

6.1 Basic definitions

Definition 49 (Legendre symbol) Let $p \in \mathfrak{P}$, $p \geq 3$, and $n \in \mathbb{Z}$. The Legendre symbol of n by p is defined as

$$\left(\frac{n}{p}\right) = \begin{cases} 0 & \text{if } p|n \\ 1 & \text{if there exists } b \text{ such that } a \equiv b^2 \bmod n \\ -1 & \text{otherwise} \end{cases}$$

Example 30 We have $\left(\frac{2}{7}\right) = 1$, since $3^2 \equiv 2 \bmod 7$. However, $\left(\frac{3}{7}\right) = 0$.

Remark 39 Elements a of $\mathbb{F}_p$ such that $\left(\frac{a}{p}\right) = 1$ are called quadratic residues modulo p.

Proposition 8 (Multiplicative property) *Let $p \in \mathfrak{P}$ and $a, b \in \mathbb{Z}$, then*

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Remark 40 As a result, multiplying a quadratic residue with a quadratic residue is again a quadratic residue; multiplying non-quadratic residues together gives a quadratic residue; multiplying a quadratic residue together with a non-quadratic residue yields a non-quadratic residue.

Remark 41 The Legendre symbol can be defined on finite extensions: for instance $\mathbb{F}_q[X] \simeq F_p[X]/(P)$, by defining for every $Q \in \mathbb{F}_p[X]$

$$\left(\frac{Q}{P}\right) = \begin{cases} 0 & \text{if } P|Q \\ 1 & \text{if } Q \text{ is a non-zero square modulo } P \\ -1 & \text{otherwise} \end{cases}$$

Definition 50 (Character) Let G be a group. A character of G is a group morphism $G \to \mathbb{C}^*$.

Remark 42 This notion is extended to fields as follows: the character of a field K is the character of its multiplicative subgroup $K^\times$. The Legendre symbol is a character of $\mathbb{F}_p$ in that sense.

Lemma 21 (Roots of 1 in $\mathbb{F}_p$) *Let $p \in \mathfrak{P}$, $p > 2$, and $\beta \in \mathbb{F}_p$. Then $\beta^2 = 1$ if and only if $\beta \in \{-1, 1\}$.*

Proof: Only the reverse direction is not trivial. Assume $\beta^2 = 1$, and lift to $\mathbb{Z}$ as $b \equiv \beta \bmod p$. Then $b^2 \equiv 1 \bmod p$ which means that $p|(b^2-1)$; but $b^2-1 = (b-1)(b+1)$ and p is prime, therefore it divides one or the other of these factors. Hence $b \equiv \pm 1 \bmod p$, and the result follows. □

Corollary 23 *Let $p \in \mathfrak{P}$, $p > 2$, and $a, b \in \mathbb{F}_p^\times$. Then $a^2 = b^2$ if and only if $a = \pm b$.*

Corollary 24 (Number of quadratic residues) *There are $(p-1)/2$ squares in $\mathbb{F}_p^\times$.*

Remark 43 In other words, for every odd prime p, exactly half the elements of $\mathbb{F}_p^\times$ are squares, and half are non-squares.

6.2 Euler's criterion

A natural question is to find efficient ways to compute the Legendre symbol. Euler's criterion provides a first tool in that direction.

Theorem 42 (Euler's criterion) *Let $p \in \mathfrak{P}$, $p > 2$, and $a \in \mathbb{F}_p^\times$. Then*

$$a^{(p-1)/2} = \left(\frac{a}{p}\right).$$

Example 31 We have $3^{(7-1)/2} = 3^3 = -1 \bmod 7$ and $4^{(7-1)/2} = 4^3 = 1 \bmod 7$. This is a more systematic way to address Example 30.

Proof: First note that, by Euler's theorem (Theorem 16), $a^{p-1} \equiv \left(a^{(p-1)/2}\right)^2 = 1 \bmod p$, which by Lemma 21 shows $a^{(p-1)/2} \equiv \pm 1 \bmod p$. If a is a quadratic residue modulo p, then there exists $b \in \mathbb{F}_p^\times$ such that $a = b^2$; using Euler's theorem again we have $a^{(p-1)/2} \equiv b^{p-1} \equiv 1 \bmod p$.

Now assume that a is not a quadratic residue. Wilson's theorem (Corollary 18) gives $\prod_{b\in\mathbb{F}_p^\times} b \equiv -1 \bmod p$. Now we will see that this product is in fact equal to $a^{(p-1)/2}$: indeed, let C be the set of pairs $(u,v) \in (\mathbb{F}_p)^2$ such that $uv = a$. Note that for any given u the corresponding v is uniquely determined, and $u \neq v$ (otherwise a would be a square, which would be a contradiction). Therefore every element of $\mathbb{F}_p^\times$ belongs to exactly one pair in C; as a result

$$\prod_{b\in\mathbb{F}_p^\times} b = \prod_{(u,v)\in C} uv = \prod_{(u,v)\in C} a = a^{(p-1)/2}$$

which finishes the proof. □

Remark 44 All the above results apply to $\mathbb{F}_q$, where $q = p^k$, by replacing $p-1$ in the exponents with $\varphi(q)$.[1]

Lemma 22 *A number a is a quadratic residue modulo $q = p^k$ if and only if it is a quadratic residue modulo p.*

Proof: Suppose that a is a quadratic residue modulo q, then $p \nmid a$ and $a \equiv b^2 \bmod q$ for some integer b. It follows that $a \equiv b^2 \bmod p$, and therefore a is a quadratic residue modulo p.

Suppose now that a is not a quadratic residue modulo q, and assume that $p \nmid a$. Then Euler's criterion gives

$$a^{\varphi(q)/2} \equiv -1 \bmod q$$

which also holds modulo p since $p|q$; by applying several times Fermat's little theorem we get

$$a \equiv a^p \equiv a^{p^2} \equiv \cdots \equiv a^{p^k-1} \bmod p$$

which gives

$$-1 \equiv a^{p^{k-1}(p-1)/2} \equiv a^{(p-1)/2} \bmod p.$$

[1] However, the proof of Lemma 21 must be completed by explaining that q, while not prime, cannot divide both $(b-1)$ and $(b+1)$, as otherwise it would divide their difference, which is impossible since p is odd.

Thus, by Euler's criterion again, a is not a quadratic residue modulo p. □

Theorem 43 (Roots of -1 in $\mathbb{F}_p$) *Let $p \in \mathfrak{P}$, $p > 2$. Then -1 is a quadratic residue modulo p if and only if $p \equiv 1 \bmod 4$.*

Proof: By Euler's criterion,

$$\left(\frac{-1}{p}\right) \equiv (-1)^{(p-1)/2} \bmod p.$$

If $p \equiv 1 \bmod 4$ then the power on the right-hand side is even, and therefore -1 is a quadratic residue. Otherwise, $p \equiv 3 \bmod 4$ and the power is odd, which shows that -1 is not a quadratic residue.

□

Corollary 25 (Congruence of primes) *There are infinitely many primes p such that $p \equiv 1 \bmod 4$; and infinitely many primes such that $p \equiv 3 \bmod 4$.*

Proof: Suppose that there are only finitely such primes and call M their product. Let $N = 4M^2 + 1$, and consider a prime p dividing N. Now p cannot satisfy $p \equiv 1 \bmod 4$, as it would otherwise divide both $4M^2$ and N, and therefore their difference $N - 4M^2 = 1$, which is impossible. Furthermore, p is odd because N is odd. Moreover, $(2M)^2 \equiv -1 \bmod p$, i.e., -1 is a quadratic residue modulo p, and by Theorem 43 this implies $p \bmod 1 \bmod 4$. Therefore there are infinitely many primes of the form $4k + 1$.

A similar argument (using $N = 4M - 1$) shows that there are infinitely many primes of the form $4k + 3$. □

Remark 45 An easy but very useful result is that

$$\left(\frac{2}{p}\right) = \begin{cases} 1 & \text{if } p \equiv \pm 1 \bmod 8 \\ -1 & \text{if } p \equiv \pm 3 \bmod 8 \end{cases} = (-1)^{(p^2-1)/8}.$$

Together with Theorem 43 and the next section's key theorem, this enables the efficient computation of any Legengre symbol.

Remark 46 Using the Chinese remainder theorem (Theorem 14), it is easy to see that if $n = p_1^{e_1} \cdot p_k^{e_k}$, a number a is a quadratic residue modulo n if an only if it is a quadratic residue modulo $p_i^{e_i}$ for each $i = 1, \ldots, k$. In particular, when this is the case, a has 2^k square roots modulo n.

Remark 47 The above remark motivates the definition of a Legendre-like symbol, defined for arbitrary moduli n. This is the Jacobi symbol defined as

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right)^{e_1} \times \cdots \times \left(\frac{a}{p_k}\right)^{e_k}$$

for non-zero $n = p_1^{e_1} \cdots p_k^{e_k}$. Note however that in general, $\left(\frac{a}{n}\right) = 1$ does not imply that a is a quadratic residue! Indeed, it may happen that $\left(\frac{a}{p_i}\right) = -1$ for a pair of indices i, so that $-1 \times -1 = 1$ in the above expression.

6.3 Quadratic reciprocity

Gauss is famous for producing the first (six!) rigorous proofs of the following result

Theorem 44 (Law of quadratic reciprocity) *Let p, q be odd primes. Then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

In other terms, $\left(\frac{p}{q}\right) = \left(\frac{q}{p}\right)$ if and only if at least one of p and q is congruent to 1 modulo 4.

To prove the theorem, following an idea of Eisenstein, we use the following easy lemma:

Lemma 23 *For any positive odd integer m,*

$$\frac{\sin mx}{\sin x} = (-4)^{(m-1)/2} \prod_{t=1}^{(m-1)/2} \left(\sin^2 x - \sin^2 \frac{2\pi t}{m}\right).$$

There are several ways to check this identity, which we leave as a rather easy exercise for the reader.

Proof: Now let $S = \{0, 1, \dots, (p-1)/2\}$, and for any $s \in S, x \in \mathbb{F}_p^\times$, define $\epsilon_s(x) \in \{\pm 1\}$ so that $\epsilon_s(x)sx \in S$. The map $\sigma : s_x \mapsto \epsilon_s(x)sx$ from S to S is injective, and therefore bijective because S is finite. Thus for any $a \in \mathbb{F}_p^\times$, $as = \epsilon_s(a)\sigma_a(s)$. Therefore

$$\begin{aligned}
\prod_{s\in S} as &= \prod_{s\in S} \epsilon_s(a)\sigma_a(s) \\
a^{(p-1)/2}\prod_{s\in S} s &= \prod_{s\in S} \epsilon_s(a) \prod_{s\in S} \sigma_a(s)
\end{aligned}$$

which gives, by Euler's criterion (Theorem 16)

$$\left(\frac{a}{p}\right) = \prod_{s\in S} \epsilon_s(a).$$

Now, since $qs = \epsilon_s(q)\sigma(q)$, we have

$$\sin\left(\frac{2\pi s}{p} q\right) = \epsilon_s(q)\sin\left(\frac{2\pi}{p}\sigma(q)\right),$$

therefore,

$$\begin{aligned}
\left(\frac{q}{p}\right) &= \prod_{s\in S} \epsilon_s(q) \\
&= \frac{\prod_{s\in S} \sin\left(\frac{2\pi s}{p} q\right)}{\prod_{s\in S} \sin\left(\frac{2\pi}{p}\sigma(q)\right)} \\
&= \prod_{s\in S} \frac{\sin\left(\frac{2\pi s}{p} q\right)}{\sin\left(\frac{2\pi}{p}\sigma(q)\right)} \\
&= \prod_{s\in S} (-4)^{(q-1)/2} \prod_{t=1}^{(q-1)/2} \left(\sin^2 \frac{2\pi s}{p} - \sin^2 \frac{2\pi t}{q}\right)
\end{aligned}$$

where in the last line we used the identity discussed at the beginning of this proof, with $m = q$ and $x = 2\pi s/p$. Introducing $T = \{0, 1, \ldots, (q-1)/2\}$, we see that

$$\left(\frac{q}{p}\right) = (-4)^{(p-1)(q-1)/4} \prod_{s\in S}\prod_{t\in T}\left(\sin^2\frac{2\pi s}{p} - \sin^2\frac{2\pi t}{q}\right).$$

Swapping p and q and following the same steps, we obtain

$$\left(\frac{p}{q}\right) = (-4)^{(p-1)(q-1)/4} \prod_{s\in S}\prod_{t\in T}\left(\sin^2\frac{2\pi t}{q} - \sin^2\frac{2\pi s}{p}\right),$$

or in other words,

$$\left(\frac{p}{q}\right) = (-1)^{|S|\cdot|T|}\left(\frac{q}{p}\right).$$

Since $|S| \cdot |T| = (p-1)(q-1)/4$, this concludes the proof. □

Algorithm 5 (Computing Legendre and Jacobi symbols) *The law of quadratic residuosity (LQR for short) makes it very practical to compute Legendre and Jacobi symbols, in a way reminiscent of Euclid's algorithm for the* gcd*:*

1. *Input a, n two integers.*
2. *Set $\sigma \leftarrow 1$*
3. *Repeat forever*
   a) $a \leftarrow a \bmod n$
   b) *If $a = 0$ then*
      - *If $n = 1$ then return σ otherwise return* 0
   c) *Compute a' such that $a = 2^h a'$ and a' is odd.*
   d) *If $h \not\equiv 0 \bmod 2$ and $n \not\equiv \pm 1 \bmod 8$ then $\sigma \leftarrow -\sigma$*
   e) *If $a' \not\equiv 1 \bmod 4$ and $n \not\equiv 1 \bmod 4$ then $\sigma \leftarrow -\sigma$*
   f) *Swap: $(a, n) \leftarrow (n, a')$*

Example 32 Let's compute the following Legendre symbol, step by step:

$$\begin{aligned}
\left(\frac{29}{43}\right) &\overset{\text{LQR}}{=} \left(\frac{43}{29}\right) \\
&\overset{\text{mod } 29}{=} \left(\frac{14}{29}\right) \\
&\overset{\text{Prop. } 8}{=} \left(\frac{2}{29}\right)\left(\frac{7}{29}\right) \\
&\overset{\text{Rem. } 45}{=} -\left(\frac{7}{29}\right) \\
&\overset{\text{LQR}}{=} -\left(\frac{-29}{7}\right) \\
&\overset{\text{mod } 7}{=} -\left(\frac{-1}{7}\right) \\
&= -1.
\end{aligned}$$

Remark 48 For odd, composite n, if we know the factorization of n, then we can also determine if a is a quadratic residue modulo n by determining if it is a quadratic residue modulo each prime divisor p of n. If we do not know the factorisation of n, in general not much can be said.

> **Open question (2018)**: Let $a_i = \left(\frac{x}{p_i}\right)$ where $p_1 = 2, p_2, \ldots \in \mathfrak{P}$ and x is some integer. Given the sequence $(a_i)_{i\geq 1}$, can one efficiently recover x?

Remark 49 The proof given here of the LQR seems unnecessarily contrived and doesn't shed must light on what's happening. Finding simpler proofs, and generalising this result to moge settings (cubic, quartic reciprocity, etc.) was one of the driving forces behind algebraic number theory.

6.4 Modular square roots

6.4.1 Square roots in $\mathbb{F}_p$

Let p be an odd prime, and a be a quadratic residue modulo p.

Proposition 9 *Assume $p \equiv 3 \bmod 4$, then $b = a^{(p+1)/4}$ is a square root of a.*

Proof: It suffices to check that $4|(p+1)$ and use Fermat's little theorem. □

Remark 50 Obviously, if b is a square root of a modulo p, then so is $-b$.

We will see another algorithm in Exercise 6.4 to compute modular square roots. In general, there are two cases: either computing the root is a straightforward application of Fermat's little theorem, or we need to distinguish several sub-possibilities. The following algorithm is due to Pocklington:

- If $p = 3 \bmod 3$, apply Proposition 9.
- Otherwise, $p = 1 \bmod 3$. We distinguish two subcases:
  - If $p = 5 + 8m$, then compute $y = (4a)^{m+1} \bmod p$. If y is even return $y/2$, otherwise return $(p+y)/2$.
  - If $p = 1 + 8m$, find (by trial and error) u_1, v_1 integers such that $u_1^2 + av_1^2 = n$ is a quadratic non-residue modulo p. Then, in a manner reminiscent to the resolution of Fermat–Pell's equations, we define the sequences

    $$u_{i+j} = u_i u_j - a v_i v_j \bmod p \qquad \text{and} \qquad v_{i+j} = u_i v_j + u_j v_i \bmod p.$$

    In particular, $u_i^2 + av_i^2 = n^i$. There is a value ℓ such that $u_\ell = 0$, at which point we have $av_\ell^2 = n^\ell$. Since a is a quadratic residue, ℓ is even and we write $\ell = 2k$. Return $u_k v_k^{-1} \bmod p$.

Example 33 As a direct application of Pocklington's algorithm (in the complicated, i.e., interesting case), let's solve $x^2 = 13 \bmod 17$. Note that $17 = 8\dot{2} + 1$, so we are in the third scenario of Pocklington's algorithm.

Choosing $u_1 = 3$ and $v_1 = 1$ gives $u_1^2 + 13 \cdot v_1^2 = 9 + 13 = 5 \bmod 17 = n$, which is a quadratic non-residue. Iterating, we get $u_8 = 0$, and the solution is $u_4 v_4^{-1} = 7 \cdot 3^{-1} \bmod 17 = 8 \bmod 17$.

We check that $8^2 = 64 \bmod 17 = 13 \bmod 17$.

Remark 51 Solving square roots in $\mathbb{F}_p$ immediately gives us the ability to solve quadratic polynomials, in a way that is very similar to the situation over $\mathbb{R}$.

Let $P = aX^2 + bX + c$ be a polynomial of degree 2 defined on $\mathbb{F}_q[X]$, where $2 \nmid q$. Let $\Delta = b^2 - 4ac$. Then P has roots in $\mathbb{F}_q$ if and only if Δ is a quadratic residue modulo q; in that case let e be a root of Δ, and we can express the roots as $\alpha_1 \pm = (-b \pm e)2^{-1}a^{-1}$.

What about modular square roots in more general settings, e.g., fields $\mathbb{F}_q$, or even rings $\mathbb{Z}/(n)$?

6.4.2 Square roots in $\mathbb{Z}/(pq)$, Rabin cryptosystem

Let $n = pq$ with p, q distinct primes. Then for any $a \in \mathbb{Z}/(n)$, we have $\left(\frac{a}{n}\right) = \left(\frac{a}{p}\right)\left(\frac{a}{q}\right)$. Thus, if a is a quadratic residue modulo n, its Jacobi symbol will be $+1$, but remember that the converse does not necessarily hold!

> **Open problem (2018).** Given a and $n = pq$ such that $\left(\frac{a}{n}\right) = 1$, is it possible to guess if a is a quadratic residue modulo n better than chance?

Remark 52 Note that if the factorisation of n is known, then the problem is easy: using the Chinese remainder theorem, we can solve over $\mathbb{Z}/(p)$ and $\mathbb{Z}/(q)$ separately, and reconstruct a solution over $\mathbb{Z}/(n)$.[2]

In fact, if there were an efficient algorithm for computing square roots modulo n, then we could immediately factor n: choose a random m and compute $a = b^2 \bmod n$, then find a square root b of a. Necessarily, $pq|(b-m)(b+m)$, so that computing $\gcd(n, b-m)$ reveals a factor of n with probability $1/2$.

The above remark is the basis for the Rabin cryptosystem: choose p, q large primes so that $p, q = 3 \bmod 4$ (for easy computation of square roots), let $n = pq$. A message is encrypted by squaring it modulo n. Decryption is equivalent to the knowledge of a factor of n.[3] As we will see in a later chapter, factorisation of large numbers is measurably difficult, which enables us to give security guarantees for this system.

6.4.3 Square roots in $\mathbb{F}_{p^k}$

Some algorithms for square root computation work equally well over the fields $\mathbb{F}_{p^k}$. When this is not the case, or when we need a more theoretical approach, the following result is very handy:

Theorem 45 (Hensel) *Let $f \in \mathbb{Z}[X]$. Let p be prime, and $k \geq 1$. If there exists r such that*

$$f(r) = 0 \bmod p^k \qquad \text{and} \qquad f'(r) \neq 0 \bmod p,$$

then there exists s and $m \leq k$ such that

$$f(s) = 0 \bmod p^{k+m} \qquad \text{and} \qquad r = s \bmod p^k.$$

Furthermore, this s is unique and $s = r - f(r)a$, with $a = f'(r)^{-1} \bmod p^m$.

The proof of this theorem is given as Exercice 6.5.

Example 34 Consider the equation $x^2 = 19 \bmod 243$. We easily check that 19 is a quadratic residue modulo $243 = 3^5$, and set out to find a solution. One way to do so is by leveraging Theorem 45 (this is known as Hensel lifting). Let $f = x^2 - 17$, then

[2] Note that if a is a quadratic residue modulo n, then it has 4 square roots.

[3] The astute reader will notice that there are 4 possible decryptions for every message, but this is easily fixed.

- First solve $f(r) = 0$ modulo $p = p^1 = 3$, which gives $r = 1 \bmod 3$.
- We have $f(1) = 0 \bmod p^1$ and $f'(1) = 2 \neq 0 \bmod p$, thus Theorem 45 applies, and we get $s = 1 \bmod 9$. We now have a solution s over $\mathbb{F}_{p^2}$, "lifted" from the solution over $\mathbb{F}_p$.
- Repeating the process, we get the solution $x = 1 + 9 = 10 \pmod{3^4}$ and finally $x = 10 + 81 = 91 \pmod{3^5}$. We check that $91^2 = 19 \bmod 243$.

6.5 Exercise set

Exercise 6.1. Let $\mathbb{F}_q$ be a finite field, the Jacobsthal matrix of $\mathbb{F}_q$ is defined as

$$Q = \left(\left(\frac{i-j}{q}\right)\right)_{i,j\in\mathbb{F}_q}$$

i.e., the matrix that indicates on row i and column j whether $i - j$ is a quadratic residue modulo q.

1. Show that Q is symmetric if $q \equiv 1 \bmod 4$. Show that Q is skew-symmetric if $q \equiv 3 \bmod 4$.

2. Write a program that computes Q for a given $q = p^k$.

3. Show that each row of Q sums to zero, i.e., $QJ = 0$, where J is the $q \times q$ matrix with all entries set to 1. Show that $JQ = 0$. Show that $QQ^\top + J \equiv 0 \bmod q$.

4. Assume that $q \equiv 3 \bmod 4$, and define the matrix

$$H = I_{q+1} + \begin{pmatrix} 0 & \mathbf{1}^\top \\ -\mathbf{1} & Q \end{pmatrix}$$

   where I_k is the $k \times k$ identity matrix, and $\mathbf{1}$ is the all-one column vector of size q. Show that $H + H^\top = 2I_{q+1}$. This makes H a (skew) Hadamard matrix of order $2(q+1)$, which can be used to design error-correcting codes. The above construction is due to Raymond Paley (1933), and was a breakthrough in the following problem:

   > **Open question (2018)**: Are there Hadamard matrices of size m, for all m divisible by 4?

   This is known as the Hadamard conjecture. Paley's construction found Hadamard matrices of all possible sizes up to 100, except 92 (which was later found using other methods). As of 2018, it is unknown whether there is a Hadamard matrix of size $m = 668$.

Exercise 6.2. The point of this exercise is to show two easy corollaries of the following result (due to Gauss):

> *A positive integer is the sum of three squares if and only if it is not of the form $4^a(8b-1)$ for integers a, b.*

1. Let $n \in \mathbb{N}$, $n > 0$. Show that the above statement implies

$$x_1^2 + x_2^2 + x_3^2 = 8n + 3$$

   has integer solutions x_1, x_2, x_3.

2. Compute the set of quadratic residues modulo 8, and show that the x_i are odd.

3. A triangular number is of the form $1 + 2 + 3 + \cdots + k$ for some k. Prove the following statement (originally due to Gauss):

   > *Every positive integer is the sum of three triangular numbers.*

4. Let $n \in \mathbb{Z}$, $n > 0$. Write $n = 4^k m$ with $4 \nmid m$. Prove the following statement (originally due to Lagrange):

   > *Every positive integer is the sum of four squares.*

Exercise 6.3. Consider the equation $E : y^2 = x^3 + x + 3$. If K is a field, we denote by $E(K)$ the set of solutions to E in $K \times K$.

1. Show that $E(\mathbb{R})$ is infinite. Plot this curve in the range $(-10, 10) \times (-10, 10)$.
2. Let $p \in \mathfrak{P}$. Show that $|E(\mathbb{F}_p)| \leq p - 1$. Write a computer program that outputs $|E(\mathbb{F}_p)|$.
3. Let $p \in \mathfrak{P}$. Show that if $P = (x, y) \in E(\mathbb{F}_p)$, then $-P = (x, -y)$ is also a solution.
4. Assume $\operatorname{char} K \neq 2, 3$. Let $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ be two points of K, $P \neq Q$ and $P \neq -Q$. Write the equation for the line through P and Q.
5. Assume $P \in E(K)$. Write the equation for the tangent line at P.
6. Assume $P, Q \in E(K)$, $P \neq Q$ and $P \neq -Q$. Show that the line through P and Q intersects with $E(K)$ at a third point R.
7. Assume $P \in E(K)$. Show that the tangent line at P intersects with $E(K)$ at a second point, which we call by convention $2P$.
8. We introduce the following notation: if $P, Q, R \in E(K)$ are distinct aligned points, we write $P+Q+R = 0$. By convention we also set $P+(-P) = 0$, and $P+P = 2P$, with $2P$ defined as in the previous point.
   a) Show that 0 is the neutral element for operation $+$ on E.
   b) Show that for any $P, Q \in E(K) \cup \{0\}$, $P + Q = Q + P$.
   c) Show that for any $P, Q, R \in E(K) \cup \{0\}$, $(P + Q) + R = P + (Q + R)$ (you may want to show this graphically).

   All these properties turn $E(K) \cup \{0\}$ into an Abelian group, called the group of rational points of the elliptic curve E.
9. Show that if $K = \mathbb{F}_q$ then this group is cyclic. How can one find a generator for it?
10. The integer $2^{1279} - 1$ is a Mersenne prime. Implement elliptic curve operations in the field $\mathbb{F}_{2^{1279}-1}$.

Exercise 6.4. The goal of this exercise is to practice and prove Cipolla's algorithm for the modular square root. Let $p \in \mathfrak{P}$, and consider the equation $x^2 = n$ in $\mathbb{F}_p$. If n is not a quadratic residue mod p then there is no solution, we therefore assume that n is a quadratic residue mod p. The algorithm works as follows:

1. Find $a \in \mathbb{F}_p$ so that $\left(\frac{a^2-n}{p}\right) = -1$ (this is done by trying several values of a)
2. Define $P = X^2 - a^2 + n$, and $K = \mathbb{F}_{p^2} \simeq \mathbb{F}_p[X]/(P)$.
3. Compute in K the expression $x = (a + X)^{(p+1)/2}$.
4. Check that $x \in \mathbb{F}_p$, and return the solutions x and $-x$.

Answer the following questions:

1. Using this method, find a solution of $x^2 = 10 \bmod 13$ (use $a = 2$).
2. Prove the correctness of this algorithm.
3. Estimate the complexity of this algorithm as a function of the binary size of p. How does this algorithm compare in the cases $p = 1 \bmod 4$ and $p = 3 \bmod 4$ to the algorithms discussed in the lecture?

Exercice 6.5. The purpose of this theorem is to prove Hensel's theorem.

1. Let $f \in \mathbb{Z}[X]$. Show that there exists $g \in \mathbb{Z}[X]$ such that

$$f(r+tp^k) = f(r) + t\,p^k\,f'(r) + p^{2k}\,t^2\,g(t)$$

2. Assume that r is a root of f modulo p^k. Assuming that s is a root of f modulo p^{k+m}, show that

$$f(r+tp^k) = f(r) + t\,p^k\,f'(r) = 0 \bmod p^{k+m}$$

3. Deduce from thee above that there exists an integer z such that

$$z + tf'(r) = 0 \bmod p^m.$$

4. Based on the above, write a complete proof of Theorem 45.

Exercice 6.6. The purpose of this exercise is to discuss the Feige–Fiat–Shamir identification protocol.

1. Assume n is a large number whose factorisation is unknown. Alice picks a secret s and computes $v \leftarrow s^2 \bmod n$. She sends v to Bob.

   Can Bob recover s from v?

2. Alice now wished to identify herself to Bob, claiming that she was the one who generated this v.

   Why can't she just reveal s?

3. Alice picks a random number r, a random $e \in \{-1, +1\}$, and computes $x \leftarrow er^2 \bmod n$. If $x = 0$ she starts over with another r. Then, she sends x to Bob.

   Can Bob say whether x is a quadratic residue modulo n?

4. Bob chooses $a \in \{0, 1\}$ and sends it to Alice. Upon receiving a, Alice computes $y \leftarrow rs^a \bmod n$, and sends y to Bob. Finally, Bob checks that $y^2 = \pm xv^a \bmod n$; if it does not hold then Bob rejects Alice's identification claim.

   Show that if Alice is honest, then Bob will not reject Alice. What would it take to be accepted, provided we are not Alice? How to fix this?

5. Show that the conversation between Alice and Bob can be simulated, i.e., that it is possible to generate exchanges (x, a, y) that are statistically indistinguishable from real exchanges (x, a, y) between Alice and Bob. This simulation does not know the secret s!

   This shows that the Feige–Fiat–Shamir protocol does not reveal any information to an adversary: it is a zero-knowledge identification protocol.

Chapter 7

Diophantine equations and elliptic curves

We mentioned elliptic curves several times in this course. In fact, elliptic curves are very much ubiquitous in mathematics (not only in Algebra!), and they are now standard in cryptography as well. As a result, there is a large collection of works on the topic, and several insightful points of views.

Instead of trying to give a systematic treatment of elliptic curves, which the interested reader will find elsewhere, we will adopt a slow pace and work through examples. In particular, we approach elliptic curves here from the angle of Diophantine equations, which is a fascinating field in itself.

TODO

Figure 7.1: The elliptic curve $y^2 = x^3 - x$ over the field $\mathbb{R}$.

Contents

7.1 Diophantine equations

The study of elliptic curves takes its roots in a much more general (and still largely open!) problem: that of solving Diophantine equations, one of the most famous of which is Fermat's:

$$a^n + b^n = c^n,$$

where a non-trivial solution is to be found in $\mathbb{Z}$, unless one can prove no such solution exist (see Problem sets). It is often more practical to work over the rationals $\mathbb{Q}$, which makes little difference.

For instance, Diophantes considered the problem of "dividing a square into a sum of cubes", say

$$3^2 = 9 = 8 + 1 = 2^3 + 1^3.$$

Natural questions follow: are there other integers a, b, c so that $a^2 = b^3 + c^3$? How many? How can we find them?

7.1.1 Degree 1 equations

To make progress, it is fruitful to develop a geometric intuition of these questions. For instance, the equation $4x - 3y = 1$ (which the reader knows how to solve using the extended Euclidean algorithm!) corresponds to a line, and we can provide a parametrisation

$$\begin{cases} x = 1 + 3t \\ y = 1 + 4t \end{cases}$$

that yields infinitely many solutions, one for each choice of $t \in \mathbb{Z}$.

This approach can be generalised to any number n of variables $x_1, x_2, \ldots, x_n$: there is an integer solution to the equation

$$a_1 x_1 + \cdots + a_n x_n = d$$

if and only if $\gcd(a_1, \ldots, a_n) \mid d$. When a solution exists, it is found by the extended Euclidean algorithm.

Remark 53 Systems of degree 1 Diophantine equations can be written as a matrix equation $AX = B$. The general strategy is to transform A into a normal form (typically Hermite or Smith normal form) from which we can iteratively solve decoupled degree 1 equations. A popular way to solve such systems automatically is to rely either on lattice reduction (Lenstra–Lenstra–Lovász algorithm) or integer linear programming.

7.1.2 The circle equation

Let's now turn our attention to a specific degree 2 equation, over $\mathbb{Q}$. In keeping with our geometric intuition, the equation $x^2 + y^2 = 1$ corresponds to a "circle". It is of course not a circle in the usual sense, since x and y are rationals, but we will temporarily ignore this detail.

Now consider a line D intersecting with the circle, as in Figure 7.2. This line usually intersects the circle at another point (except in a degenerate situation where D is tangent to the circle; we ignore this case for the moment). Assume that D intersects the circle at $(-1, 0)$, then it intersects the y axis at a point of coordinate t, and intersects the circle at another point P.

Applying Thales' theorem, we quickly find the coordinates for P:

$$P(t) = (P_x, P_y) = \left(\frac{1 - t^2}{1 + t^2}, \frac{2t}{1 + t^2} \right).$$

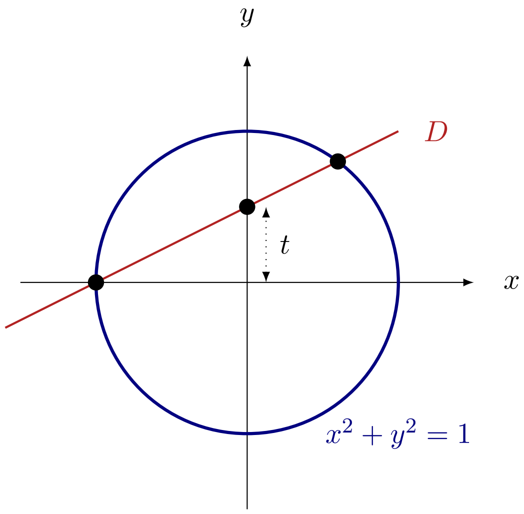


Figure 7.2: The "circle" $x^2 + y^2 = 1$ and an intersecting line D. Except in the degenerate case where D is tangent to it, D intersects the circle twice.

One immediately notices that P_x and P_y are rational numbers if and only if t is a rational number. Thus, with the notable exception of the degenerate situation, we have a parametrisation for all solutions of the degree 2 Diophantine equation $x^2 + y^2 = 1$. The degenerate situation is easy to deal with: we artificially append the point $O = (-1, 0)$.[1] Thus the complete set of solutions is

$$\{P(t), t \in \mathbb{Q}\} \cup \{O\}.$$

Remark 54 The above approach yields the complete set of solutions for circle equations that admit rational solutions. That being said, some equations have no such solutions, such as $x^2 + y^2 = -1$, or even $x^2 + y^2 = 3$. How can we know? The following result provides a useful tool:

Theorem 46 (Hasse–Minkowski theorem) *A quadratic equation has a solution in rational numbers if and only it has a solution over $\mathbb{R}$ and over the p-adic integers $\mathbb{Z}_p$ for every prime p.*[2]

7.1.3 The Fermat–Pell equation

Another important example on our way to elliptic curves is the degree 2 Fermal–Pell equation

$$x^2 - Dy^2 = 1,$$

where D is a square-free positive integer. Our geometric intuition makes this object an avatar of the "hyperbole", and we may look for techniques similar to what we used for the "circle". Instead, let's assume that we have found a solution (x, y) to this equation, and observe that

$$x' = x^2 + Dy^2$$
$$y' = 2xy$$

[1] The point O is sometimes called "the point at infinity", as it corresponds to the limit of $P(t)$ when $t \to \pm\infty$.

[2] A p-adic integer is a formal sum $a_0 + a_1 p + a_2 p^2 + \cdots$ with integer coefficients a_i. Such numbers can be added or multiplied in the obvious way, and two p-adic integer are considered equal if they match modulo p^n for all n.

satisfy

$$\begin{aligned} x'^2 - Dy'^2 &= x^4 + D^2y^4 + 2Dx^2y^2 - 4Dx^2y^2 \\ &= x^4 + D^2y^4 - 2Dx^2y^2 \\ &= (x^2 - Dy^2)^2 \\ &= 1. \end{aligned}$$

In other terms, if we have a solution (x, y), then we know how to generate a new solution (x', y'). This process can be repeated, and we in fact construct infinitely many solutions. More generally, given any two solutions (x_1, y_1) and (x_2, y_2), we get a new solution (x_3, y_3) by combining them:

$$\begin{aligned} x_3 &= x_1x_2 + Dy_1y_2 \\ y_3 &= x_1y_2 + x_2y_1. \end{aligned}$$

But where do these relations come from?

Let's factor $x^2 - Dy^2$ as $(x - \sqrt{D}y)(x + \sqrt{D}y)$. Of course, this makes no sense in $\mathbb{Q}$ to take the square root of the square-free integer D, so we actually work in the extension field $\mathbb{Q}[\sqrt{D}]$, or to make things simpler, in $\mathbb{Z}[\sqrt{D}]$. But then something magical happens! Indeed, $x^2 - Dy^2 = N(x + \sqrt{D})$ is the norm of the element $x + \sqrt{D}$ in our ring. Thus solving the Fermat–Pell equation is exactly the same thing as finding the unit group $\mathbb{Z}[\sqrt{D}]^\times$. This point of view gives us to important facts: first, that we are looking at a group, and second that we can therefore use the group operation to generate new elements from old ones.[3]

We haven't said how to get a correct solution (x, y) to begin with. This is a particularly nice question, with insightful ideas (e.g. continued fractions), and the interested reader will find many references for this. Let us only point out the following conjecture:

> **Open question (2018).** Is there any method for solving the Fermat–Pell equation, spelling out a solution (x, y) in full, that runs in polynomial time?

Remark 55 There are several algorithms to solve the Fermat–Pell equation, e.g. based on continued fraction expansions, or Cornacchia's algorithm.

7.2 Elliptic curves

7.2.1 Definition

As we increase the degree of our Diophantine equation, we eventually meet the degree 3 case. For such "curves", with very few exceptions, it is pointless to look for a parametrisation. We will restrict ourselves to a very narrow (but so important!) class of degree 3 Diophantine equations:

Definition 51 (Elliptic curve over $\mathbb{Q}$) Let $P \in \mathbb{Q}[x, y]$ of degree 3, and let C be the curve defined by $P(x, y) = 0$. If

1. There exists at least one couple $(x, y) \in \mathbb{Q}^2$ in C;
2. C does not contain any line;
3. C cannot be parametrised by any couple of rational fractions $(x(t), y(t))$;

then C is called the elliptic curve of equation $P(x, y) = 0$ over $\mathbb{Q}$.

TODO

Figure 7.3: The elliptic curve $y^2 = x^3 - x$ over the field $\mathbb{R}$.

As an example, consider $y^2 = x^3 - x$, which satisfies all these constraints (the reader is invited to perform an easy check) and is represented in Figure 7.3.

Another example is the equation $x^3 + y^3 = 9$ (with which we started the section on Diophantine equations). Well this turns out to be an elliptic curve as well! If we can find all the points of the curve, we can find all the decompositions of 9 (or any integer N) as a sum of cubes.

Remark 56 The reader may wonder at this point why we call such objects "elliptic curves". This name comes from another point of view (remember that elliptic curves appear everywhere!) concerned with the trigonometry in ellipses.

For a circle, $x^2 + y^2 = 1$, we have a parametrisation in terms of special functions $\cos(t)$ and $\sin(t)$; there are of course many ways to define these functions, one of them being as the inverse of simple integrals, such as

$$\int_0^s \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \operatorname{Arcsin}(s).$$

The reader will easily verify that the above can be interpreted geometrically as saying that the Arcsin function measures a circle's arc length; for instance $\operatorname{Arcsin}(1) = \pi/2$ measures the length of a circle's quarter.

Mathematicians in the 18th century realised that the same process can be applied to ellipses. The arc length is thus obtained as an integral. One of such integrals is:

$$F(\phi) = \int_0^\phi \frac{\mathrm{d}t}{\sqrt{1-k^2\sin^2 t}}$$

for some $k \in [0,1]$, which can be rewritten by substituting x for $\sin t$:

$$F(u) = \int_0^u \frac{\mathrm{d}x}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

As for the circle's case, we may consider this to be an "inverse" function of some elliptic trigonometric function. This intuition turns out to be correct, although not immediately workable, and the very same way that $\cos$ and $\sin$ give rise to an algebraic equation $x^2 + y^2 = 1$, elliptic functions give rise to expressions of the form $y^2 = ax^3 + bx + c$, i.e., elliptic curves.

Thus, there is a (distant and non trivial) reltionship between elliptic curves and ellipses, although they are very different objects!

7.2.2 The group law

A remarkable feature of elliptic curves, which is similar to the circle situation and the Fermat–Pell situation discussed earlier, is that if we know two points then we can construct a third. For the particular case of $x^3 + y^3 = N$ for instance, given any solution (x, y) we get a new solution

$$x' = \frac{x(y^3+N)}{x^3-y^3} \qquad \text{and} \qquad y' = -\frac{y(x^3+N)}{x^3-y^3}.$$

[3]An important theorem of Dirichlet, which we haven't discussed in this course, appropriately called the "unit" theorem, tells us that the group we are looking at the product of $\mathbb{Z}/(2)$ with an infinite cyclic group.

TODO

Figure 7.4: Chord-and-tangent operations on an elliptic curve (illustrated over $\mathbb{R}$ here).

Naturally at least two questions arise: can we find such formulas in a systematic way, for any given elliptic curve? and does this procedure give all the points on the curve?

To the first question we can answer affirmatively. Geometrically, we can do the following:

1. Take any point P where the curve's tangent is not vertical. Then a vertical line through P will intersect the curve at another point, $-P$ (the notation will become clear in a moment).

2. Take a point P where the curve's tangent is not vertical. Then the tangent at P will intersect the curve at another point, $2P$ (the notation will become clear in a moment).

3. Take two distinct points (P, Q) on an elliptic curve C that are not aligned vertically; draw the line (PQ), which intersects C on a third point $-R$.

What the above indicates is that the set of points of the curve (almost) forms an additive group; what is missing is the unit. We artificially add the unit O (which geometrically corresponds to a vertical line). The reader will verify that the introduction of O allows to resolve all the exceptions in our construction. But there is still a subtle point: associativity, which is not entirely trivial, but always holds. Thus we have a full-fledged Abelian group. See Figure 7.4.

Furthermore, all the operations above involve intersection of lines, so that starting from a rational point we can only end up in another rational point (as in the circle case). For this reason the group that we obtain is called the group of rational points of the curve. Accordingly, we use the notation "$P+Q$" to denote addition in this group.

Remark 57 This "chord-and-tangent" construction of the group of rational points of an elliptic curve is very general. In particular, and this is essential, it works exactly as well for curves defined over finite fields. In fact, the very equations for new points are identical (except for fields of characteristic 2 or 3).

Now we can turn to the second question we had: is the group of points obtained in this way the whole of the curve's points? This is answered by a famous theorem of Mordell (later extended by Weil):

Theorem 47 (Mordell–Weil) *Let E be an elliptic curve. There exists a finite set S of rational points on E, such that every point of $E(\mathbb{Q})$ is derived from S by a finite number of applications of the chord-and-tangents construction. In particular,*

- *Only finitely many points of $E(\mathbb{Q})$ have finite order, they are called torsion points;*
- *There exists $r \geq 0$ and rational points $S = \{S_1, \ldots, S_r\}$ such that any point P of C can be uniquely written $P = a_1 S_1 + \cdots + a_r S_r + R$, with $a_i \in \mathbb{Z}$ and R a torsion point.*

The proof of this result is far beyond this course's scope. The integer r is called the rank of C, and is not very well understood as of today.

> **Open question (2018).** Given the equation for an elliptic curve, how to compute its rank?

> **Open question (2018).** Are there curves of rank $r > 28$?

For cryptographic applications, and many arithmetic results, the case of elliptic curves over finite fields is the most interesting. In particular, the group of points $E(\mathbb{F}_q)$ is necessarily finite itself. More precisely

Theorem 48 (Hasse–Weil) *Let $q = p^k$ be a prime power and E an elliptic curve over $\mathbb{F}_q$. Then*

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

While not necessarily difficult, we skip the proof for this result here as it is related to the theory of zeta functions of curves, which we haven't even mentioned.

As a consequence, we know that the group of rational points of E is a direct sum of cyclic groups of prime order, namely $E(\mathbb{F}_q) \simeq C_n \oplus C_{nk}$ where n, k are positive integers. As a result, this group is either cyclic ($n = 1$), or the product of two cyclic groups.

7.2.3 The number of points

The number of actual points on a curve is not completely understood, although there are efficient algorithms to count them, such as the Schoof–Elkies–Atkin algorithm (which is beyond the scope of this course). For most cryptographic applications, one wishes to have a group as large as possible, but as of now the standard procedure is to generate curves, count their points, and repeat until there are enough of them. This seems sufficient for applications.

But it is perhaps important to insist that this question of counting points on elliptic curves has far-reaching consequences in mathematics. One of the ways that this connection is visible is through the celebrated "modularity theorem" (known before its proof in 2001 as the Taniyama–Shimura–Weil conjecture), which amongst others prove Fermat's conjecture that $a^n + b^n = c^n$ has no non-trivial solutions for $n > 2$.[4] Another related open question is the conjecture of Birch and Swinnerton-Dyer. Both the modularity theorem and the Birch–Swinnerton-Dyer conjectures rely on the theory of L-functions on elliptic curves, which we won't discuss further, but suffices to say that they are constructed from nothing more than the number of points on the curves in $\mathbb{F}_p$ for primes p.

7.3 More general equations

As we consider more general equations (e.g., higher degree), we encounter powerful obstacles. Hilbert asked whether a general algorithm can exist that would solve any given Diophantine equation (Hilbert's tenth problem). A result of Matiyasevich–Robinson–Davis–Putnam answers by the negative: no such algorithm exist.[5]

One way to go forward is to focus on specific families. One such family is called Abelian varieties, which in a sense generalise elliptic curves.

Another way to go forward is to develop new mathematical tools, trying to generalise the way we approached Fermat–Pell's equation. This is the starting point of an ambitious programme, still unfolding to this day.

[4] In fact, Wiles and Taylor's 1994 proof of Fermat's conjecture proceeded to show a special case of the modularity theorem, which was sufficient. Further work completed the proof of the full modularity theorem.

[5] In fact, a corollary of this theorem is that one can explicitly construct a Diophantine equation which has no solutions, but such that this fact cannot even be proved. There are many other profound and surprising consequences of this result. Matiyasevich was 23 when he published his results.

7.4 Exercice set

Exercise 7.1. Write out explicitly the addition and doubling law for an elliptic curve over $\mathbb{Q}$ (which also works for $\mathbb{F}_p$ if $p > 3$). What about $\mathbb{F}_2$ and $\mathbb{F}_3$? Verify associativity (you may want to use a computer for this).

Exercice 7.2. Implement the addition and doubling law on an elliptic curve. How to find a point on the curve?

Exercice 7.3. Implement Diffie–Hellman key exchange over the elliptic curve `Curve25519`

$$y^2 = x^3 + 486662x^2 + x$$

over $\mathbb{F}_p$ with $p = 2^{255} - 19$, using as base point P whose x coordinate is 9. As a cultural note, this curve was specially designed for this application. Amongst other nice properties, it is possible to perform point multiplication in a fixed amount of time, using the Montgomery ladder technique (not discussed here).

Exercise 7.4. Solve the Diophantine equation $102x + 1003y = 1$.

Exercise 7.5. Find 10 solutions to the Fermat–Pell equation $x^2 - 2y^2 = 1$.

Exercise 7.6. Consider the curve $E : y^2 = x^3 - 2$. Show that $\#E(\mathbb{F}_{3^k}) = 3^k + (\sqrt{3})^k + (-\sqrt{3})^k$ for all $k \geq 1$. Is this compatible with the Hasse–Weil bound?

Exercise 7.7. Let $k = \mathbb{F}_q$ be a finite field of prime characteristic different from 2 and 3. Let E be an elliptic curve given by the equation $y^2 = x^3 + ax + b$ (this is called the short Weierstraß form of E), such that $4a^3 + 27b^2 \neq 0$.

1. Let $j(E) = 1728 \times 4a^3/(4a^3+27b^2)$, called the j-invariant of E. Check that $j(E) = 0$ if an only if $a = 0$. Check that $j(E) = 1728$ if an only if $b = 0$.

2. Let K be an extension of k. Let $E' : y^2 = x^3 + a'x + b'$ with $a' = u^4a$ and $b' = u^6b$ for some unit $u \in K^\times$. Show that $j(E') = j(E)$.

3. Show that the map $f : E \to E'$ defined by $(x, y) \mapsto (u^2x, u^3y)$ is an isomorphism.

4. Let $j \notin \{0, 1728\}$, show that the curve

$$y^2 = x^3 + \frac{3j}{1728 - j}x + \frac{2j}{1728 - j}$$

   has j-invariant equal to j.

5. Two curves E, E' such that $j(E) = j(E')$ are isomorphic over the algebraic closure $\overline{k}$ of k. In many cases of interest, two curves are isomorphic over some degree d extension of k, but non isomorphic for any extension of degree $< d$. Such curves are called degree d twists of one another.

   Let $E : y^2 = x^3 + ax + b$ and $E' : y^2 = x^3 + c^2ax + c^3b$, where $c \in k^\times$. Under what conditions on c are E and E' quadratic twists of one another?

Exercise 7.8. The goal of this exercise is to discuss the Cantor–Zassenhaus algorithm for polynomial factorisation.

1. Prove that
$$x^p - x = \prod_{a \in \mathbb{F}_p} (x-a).$$
Therefore, if $f \in \mathbb{F}_q[X]$, then $\gcd(f, x^p - x)$ is the product of all linear factors of the form $(x - a_i)$.

2. Prove that
$$x^{p^i} - x = \prod_{a \in \mathbb{F}_{p^i}} (x-a).$$
In other terms, x^{p^i} is the product of all irreducible polynomials over $\mathbb{F}_p$ of degree $d|i$.

3. Prove that
$$x^{(p-1)/2} - 1 = \prod_{\left(\frac{a}{p}\right)=1} (x-a).$$
Therefore, taking the $\gcd$ with f we get a product of quadratic factors.

4. Assume that f has degree n, and that it is has m irreducible factors, each having the same degree d. Consider the following algorithm:
   - Pick a random polynomial a, $\deg a < \deg f$. If $\gcd(f, a) \neq 1$ return $\gcd(f, a)$.
   - Compute $a' \leftarrow a^{(p^d-1)/2} + 1$. If $a' = 0$ or $\gcd(a', f) \neq 1$ repeat the previous step.
   - Otherwise return $\gcd(a', f)$.

   Show that the last step happens with probability at least $1 - 2^{m-1}$. We refer to this algorithm as `EDF(f, d)`.

5. Show that the following algorithm gives a distinct degree factorisation of f, i.e., a factorisation $f = g_1 \cdots g_n$ where g_i is the product of all irreducible factors of f of degree i.
   - Let $f_0 \leftarrow f$ and $i \leftarrow 1$.
   - Compute $s_i \leftarrow x^{p^i} - x \bmod f_{i-1}$
   - Compute $g_i \leftarrow \gcd(f_{i-1}, s_i)$
   - Let $f_i \leftarrow f_{i-1}/g_i$.
   - If $i < n$, then let $i \leftarrow i + 1$ and go to step 2. Otherwise return $\{g_1, \ldots, g_n\}$.

6. Show that if f is not square-free, then $f/\gcd(f, f')$ is square free. Without loss of generality we will assume that f is square-free.

7. The Cantor–Zassenhaus algorithm can thus be described:
   - Compute the distinct degree factorisation of f: $\{g_1, \ldots, g_n\}$.
   - If $g_n \neq 1$ return f (f is irreducible).
   - Return $F = \{\texttt{EDF}(g_i, d) \text{ for each } g_i \neq 1\}$.

   Using this algorithm, show that $f = x^7 - x^5 + x^3 - x \in \mathbb{F}_3[X]$ can be factored as
$$x(x+1)(x-1)(x^2+x-1)(x^2-x-1).$$

Part II

Applications

Chapter 8

Primality testing and factorisation

We have highlighted the relevance of prime numbers in algebra. In this chapter we develop algorithmic tools to tell whether a number is prime. This is especially important to cryptographic applications; we will also discuss algorithms to find the factorisation of a composite number, and ask similar questions for polynomials.

TODO

Figure 8.1: TODO: illustration.

Contents

8.1 Testing primality and generating prime numbers

There are essentially two approaches to generating prime numbers: the "generative" and the "discriminative". The "generative" consists in producing numbers that are known to be prime (it is easy to come up with such a sequence). For instance,

Remark 58 In 1947, William H. Mills proved the existence of a number A such that, for every natural number n,

$$\lceil A^{3^n} \rceil \in \mathfrak{P}.$$

However... the value of A, called Mill's constant, is unknown! Under the Riemann hypothesis[1], its value can be estimated as

$$A \approx 1.3063778838630806904686144926\ldots$$

Naturally, not much is known about this number (not even whether it is rational...). As of 2018, it is known that the number obtained with $n = 11$ is indeed prime.

The "discriminative" approach, in contrast, consists in designing a test that will tell whether a given integer is prime or not. It is thereforee called a primality testing algorithm. When looking for large primes, this is the preferred method.

The simplest conceivable primality testing method is the trial division algorithm, whereby we attempt to divide the given integer n by 2, 3, etc., until either a factor is found, or no factor smaller than $\sqrt{n}$ exists. This algorithm is of course terribly inefficient, but it will always succeed.

8.1.1 Probabilistic primality testing

A much more efficient approach is to relax the requirement to provide an exact answer to the primality question. A probabilistic primality test takes as input an integer n and (usually) returns a value indicating that n is not prime, or that n is *maybe* prime. In the second case, n is called a probable prime, where the error rate is of course dependent on the algorithm. By repeating the algorithm, we can reduce the probability of a false positive to an arbitrary small value.

The general strategy and Fermat test

The general strategy to construct probabilistic primality testing algorithms consists in the following steps:

1. Taking a result that holds true for prime numbers. For instance, Fermat's little theorem tells us that if n is prime, then $a^{n-1} = 1$ for all $a \in \mathbb{Z}/(n)$.

2. The converse does not hold in general, namely, it may happen that $a^{n-1} = 1$ for all $a \in \mathbb{Z}/(n)$ but that n is not prime. Nevertheless, such a situation is relatively rare. Thus, if $a^{n-1} = 1$ for all $a \in \mathbb{Z}/(n)$, output that n is probably prime. Otherwise, output that n is not prime (with certainty).

The error rate of such an algorithm is exactly the proportion of "exceptions" in the converse statement. In the example given, counter-examples are Carmichael numbers, for instance $n = 3 \cdot 11 \cdot 17$. Carmichael numbers are exceedingly rare[2], but we know that there are an infinity of them.

---

[1] In particular, the Riemann hypothesis implies that there exists a prime between any two consecutive cubes.

[2] Between 1 and 10^{21}, there are only 20 138 200 Carmichael numbers.

Theorem 49 (Korselt) *A positive composite integer n is a Carmichael number if and only if n is square-free, and for all prime divisors p of n, it is true that $(p-1) \mid (n-1)$.*

We can test $a^{n-1} = 1 \bmod n$ in time $O(\log(n)^3)$; in practice we will not try all values of a (so that there are more false positives beyond Carmichael numbers!) but typically only a few of them: this gives the Fermat primality testing algorithm.

The Miller–Rabin algorithm

A slightly improved variant of Fermat's algorithm builds from the following observation: if $n > 2$ is prime, then $n-1$ is even and we can write $n-1 = 2^h t$ for some $h \geq 1$ and odd integer t. Applying Fermat's little theorem on n then gives the following two constraints:

$$a^{2^h t} = 1 \qquad \text{and} \qquad (a^{2^{j+1}t} = 1) \Longrightarrow (a^{2^j t} = \pm 1) \tag{8.1}$$

where $j = 0, \dots, h-1$. Implementing this test gives the Miller–Rabin algorithm, which we now describe in full length:

- Input: integers $n, k > 1$.
- Output: Prime, Probably prime, or Composite.

  1. If $n = 2$ then return Prime.
  2. If $n = 0 \bmod 2$ then return Composite.
  3. Repeat k times:
     - Pick $a \in \{2, \dots, n-1\}$ uniformly at random;
     - Apply Fermat's test using Equation (8.1). Concretely,
       * Write $n - 1 = 2^t h$
       * Let $b \leftarrow a^t$. If $b = 1$ then return Probably prime.
       * For $j = 0$ to $h - 1$,
         · If $b = -1$ then return Probably prime.
         · If $b = +1$ then return Composite.
         · Let $b \leftarrow b^2 \bmod n$
  4. Return Composite.

Theorem 50 (False positive rate) *The Miller–Rabin algorithm has a false positive rate of at most 4^{-k}.*

Proof: The proof is not complicated but is omitted here (see Shoup, p. 310, Theorem 10.3). □

In other terms, if n is prime then the Miller–Rabin algorithm detects it, and if n is composite there is a probability at most 2^{-2k} that the Miller–Rabin algorithm wrongly claims it is prime.[3] The running time is $O(k \log(n)^3)$.

Remark 59 If we have a set of small primes $\{p_1, \dots, p_u\}$, we can use trial division to quickly rule out composite numbers. This is faster than Miller–Rabin (it runs in $O(\log n)$) and can thus be used as a pre-treatment when testing many candidate n values.

[3] In fact, this estimation is overly pessimistic: the error probability is much smaller.

The Solovay–Strassen algorithm

To illustrate that the strategy to construct probabilistic primality testing algorithms is indeed general, we can contemplate another avenue. Recall that for a prime n, by Euler's theorem, for any integer a,

$$a^{(n-1)/2} = \left(\frac{a}{n}\right) \bmod n.$$

We can therefore construct a probabilistic primality test by considering n probably prime when the above equation holds. This yields the Solovay–Strassen algorithm. The error rate of this algorithm, which repeats Euler's test k times, is 2^{-k}.

8.1.2 Deterministic probability testing

The above algorithms are inherently probabilistic.[4] For some time, although it was theoretically known that deterministic primality testing algorithms were possible, no such algorithm was known.

This changed in 2002 when Agrawal, Kayal, and Saxena published a stunning paper describing a simple, unconditional, generic, deterministic, polynomial time algorithm for primality testing.[5] While this is certainly of theoretical importance, the AKS algorithm is utterly impractical[6], orders of magnitude slower than probabilistic tests for any concrete needs.

We mention its existence here, beyond a cultural point, to emphasise how powerful probabilistic methods are.

8.1.3 Verifiable probability testing

In cryptography, even a very probable prime may not be enough. We may need a proof that a given number is prime, which could be checked without having to run a costly primality test. Concretely, we would like to have an efficiently verifiable proof that n is prime (or not).

Theorem 51 (Pocklington) *Let n be an integer. If there exist integers a and q such that*

1. *q is prime, $q|(n-1)$, and $q > \sqrt{n} - 1$;*
2. *$a^{n-1} = 1 \bmod n$;*
3. *$\gcd\left(a^{(n-1)/q} - 1, n\right) = 1$;*

then n is prime.

Proof: Suppose that n is composite, then there exists a prime $p \leq \sqrt{n}$ that divides n. In particular, $q > p - 1$, and therefore $\gcd(q, p-1) = 1$, which means that q is invertible modulo $p - 1$. Let u be the inverse of q modulo $p - 1$.

By assumption, $a^{n-1} = 1 \bmod n$, and therefore $a^{n-1} \bmod p$ since $p|n$. Furthermore, $a^{n-1} = (a^{u(n-1)}) = (a^{uq})^{(n-1)/q} \bmod p$, which means $a^{(n-1)/q} = 1 \bmod p$ by Fermat's little theorem. But this contradicts the third assumption as p is a common factor between $a^{(n-1)/q} - 1$ and n. □

---

[4] In fact, Miller's algorithm was initially deterministic, but relied on a generalisation of the (unproved) Riemann hypothesis for its correctness. Rabin's made the algorithm probabilistic, and proved its correctness.

[5] The authors received the 2006 Gödel Prize and the 2006 Fulkerson Prize for this work.

[6] Its complexity is estimated to be $O(\log(n)^6 \log^k \log(n))$ for some k.

The pair (q, a) enables anyone to efficiently checking the three conditions of Pocklington's theorem, and are called a primality certificate.

Pocklington's theorem is nowadays chiefly used over elliptic curves:

Theorem 52 (Pocklington, elliptic curve version) *Let $n > 2$ be an integer, and E be the set of solutions to $y^2 = x^3 + ax + b \bmod n$, together with a special element 0. E can be endowed with the structure of an Abelian group, with neutral element 0. Let m be an integer. If there exist an integer q and a point $P \in E$ such that*

1. *q is prime, $q|m$, and $q > (\sqrt[4]{n}+1)^2$;*
2. *$[m]P = 0$;*
3. *$[m/q]P \neq 0$;*

then n is prime.

Proof: Assume that n is composite, then there exists a prime $p \leq \sqrt{N}$ that divides n.

Let E_p be the the elliptic curve defined by equation of E taken modlo p, which gives a group of order m_p. By Hasse's theorem on elliptic curves we have

$$m_p \leq p + 1 + 2\sqrt{p} = (\sqrt{p}+1)^2 \leq \left(\sqrt[4]{N}+1\right)^2 < q$$

and thus $\gcd(q, m_p) = 1$: therefore there exists an integer u such that $uq = 1 \bmod m_p$.

Let P_p be the evaluation of P modulo p. On E_p, $[m/q]P_p = [uq][m/q]P_p = [um]P_p = 0$ (recall that $p|n$). This contradicts the second assumption, because if $[m/q]P \neq 0 \bmod n$ the we would get $[m/q]P_p \neq 0$. $\square$

Elliptic curve primality testing (ECPP for short) is at the heart of the most efficient verifiable techniques available, and is therefore widely used.

The practical implementation of such techniques (such as the celebrated Atkin–Morain algorithm) requires more knowledge about elliptic curves that we are willing to expose in this course. Furthermore, the running time analysis of these algorithms is still largely conjectural.

Nevertheless, ECPP is of immense help in tackling real-world problems and turns out to be very efficient. It is in particular by using ECPP that the 11-th Mill number was proved to be prime in 2009.[7]

8.2 Testing irreducibility and generating irreducible polynomials

As the reader is now familiar with, prime numbers are symptoms of a broader phenomenon, namely that of prime ideals in rings. A natural extension of our discussion of primality tests is therefore to look for algorithms that answer the question of whether a given ideal is irreducible in a given ring. It turns out that this is also a very important practical question, with applications in error-correcting codes for instance. We will in this section only consider the case of finite fields.

[7]This number, $((((((((2^3+3)^3+30)^3+6)^3+80)^3+12)^3+450)^3+894)^3+3636)^3+70756)^3+97220$, has 20 562 decimal digits.

8.2.1 Polynomial irreducibility testing

Let $q = p^k$ be a prime power, and let $d > 1$ be an integer. Observe that if we multiply together all the monic irreducible polynomials of $\mathbb{F}_q[x]$ of degree $n|d$, we get $Q = x^{q^d} - x$ (see Lemma 20).[8] This immediately gives the following algorithm, which takes as input a polynomial P:

1. If $\gcd(Q, P) \neq P$, return Reducible.
2. Let $n \leftarrow \deg(P)$, factor $n = p_1^{a_1} \cdots p_r^{a_r}$.
3. For every p_i,
   - Let $Q_i \leftarrow x^{q^{n/p_i}} - x$.
   - If $\gcd(Q_i, P) \neq 1$, return Reducible.
4. Return Irreducible.

Note that this is a deterministic algorithm, which runs in time polynomial to $\deg(P)$ and $\log(q)$.[9] If n is prime we can avoid Step 2.

8.2.2 Generating irreducible polynomials

Similarly to prime generation algorithms, we can design an irreducible polynomial generator by picking a random polynomial, testing for irreducibility using the above algorithm, and iterating until success. As we mentioned, the proportion of irreducible polynomials of degree n in $\mathbb{F}_q[x]$ is about q^n/n, so this algorithm succeeds in $O(n)$ steps.

> **Open question (2018)**: Is there an efficient deterministic procedure to generate irreducible polynomials?[10]

8.3 Factoring integers

We now have techniques able to determine whether a given integer is prime. If such is the case, then of course we have factored n; but if n is composite then the next big question is to exhibit a divisor of n.

8.3.1 Smoothness

The trial division algorithm mentioned in the previous section can be used to find the factors of a composite number: it suffices to keep track of successfull divisions. For some numbers, this is particularly efficient:

Definition 52 (Smooth integer) Let $B > 0$. An integer $n \in \mathbb{Z}$ is B-smooth if all the prime factors of n are smaller than B.

Remark 60 For relatively small values of B, testing for B-smoothness can be done efficiently by trial division, namely in time $O(B \log(n))$.

[8] This is a result of Gauss. In particular, it serves to show that there are around q^d/d irreducible polynomials of degree d.

[9] In practice, one would use fast algorithms to compute x^{q^n} better than the naive exponentiation, yielding an even slightly lower complexity.

[10] It is known that a deterministic algorithm exists under the "extended Riemann hypothesis", but nothing is known whether it is efficient, i.e., polynomial-time.

We may expect however that not all numbers are B-smooth. This intuition can be made more precise by introducing the de Bruijn function

$$\Psi(x, B) = \#\{\text{positive } B\text{-smooth integers smaller than } x\}.$$

Remarkably, an estimate for this quantity is known:

$$\Psi(x, B) = x\rho\left(\frac{\log x}{\log B}\right) + O\left(\frac{x}{\log B}\right)$$

where $\rho(u)$ is the solution to $u\rho'(u) + \rho(u-1) = 0$ such that $\rho([0,1]) = 1$. In particular, $\rho(u) \approx u^{-u}$. As an example, if we draw at random a 1024-bit digit, then there is a probability

$$\frac{\Psi(2^{1024}, 2^{16})}{2^{1024}} \approx 64^{-64} = 2^{-384}$$

that this number is 2^{16}-smooth, which is very unlikely.[11] As a comparison, 128-bit numbers have a probability 2^{-24} to be 2^{16}-smooth.

Devising a factoring algorithm that bets on its input being smooth is therefore not in itself the best approach. Nevertheless, smoothness will prove instrumental in the design of more efficient algorithms.

8.3.2 Dixon's method

Another key ingredient is the following observation. Assume that we are given an integer n, and that we know $x \neq \pm y \bmod n$ such that $x^2 = y^2 \bmod n$. Then in particular $(x-y)(x+y)|n$, so for instance a factor of n is given by computing $\gcd(x-y, n)$.

Example 35 Consider $n = 84923$. Let $x = 505$ and $y = 16$, which are such that $x \neq \pm y \bmod n$, and $x^2 = y^2 \bmod n$. Therefore $\gcd(x-y, n) = 163$ is a factor of n, and we verify indeed that $n = 163 \cdot 521$.

Given x and y this method is algorithmically very fast, as it is essentially Euclid's algorithm. Naturally, the hard question is to find x and y in the first place.

Dixon's approach consists in looking for numbers z such that $z^2 \bmod n$ is B-smooth, i.e., $z^2 = p_1^{a_1} \cdots p_r^{a_r}$. Multiplying such numbers is equivalent to adding the corresponding exponents; in fact, given enough such numbers, we can use linear algebra to choose k of them $z_1, \ldots, z_k$

$$z_1^2 \cdot z_2^2 \cdots z_k^2 = \prod_{i=1}^{r} p_i^{a_{i,1}+\cdots+a_{i,k}}$$

so that $a_{i,1} + \cdots + a_{i,k} = 0 \bmod 2$. The left hand side can be written x^2, and the right hand side can be written y^2, yielding the desired couple.

Example 36 Consider again $n = 84923$. Let $B = 7$, so that $p_i = \{2, 3, 5, 7\}$. We look for B-smooth squares, and find

$$\begin{aligned} 513^2 &= 8400 \bmod n &&= 2^4 \cdot 3 \cdot 5^2 \cdot 7 \\ 537^2 &= 33600 \bmod n &&= 2^6 \cdot 3 \cdot 5^2 \cdot 7 \end{aligned}$$

Multiplying these numbers together, we get on the left $(513 \cdot 537)^2$, and on the right $2^{10} \cdot 3^2 \cdot 5^4 \cdot 7^2 = \left(2^5 \cdot 3 \cdot 5^2 \cdot 7\right)^2$.

We get $x = 275481 = 20712 \bmod n$, and $y = 16800 \bmod n$. Finally, $\gcd(x-y, n) = \gcd(3912, 84923) = 163$ gives us a factor of n.

[11] A result of Schroeppel states that if we pick an integer n, then with probability $\ln 2 \approx 0.69$ the largest prime factor of n will be larger than $\sqrt{n}$. This explains in part the relative rarity of smooth numbers.

The analysis of Dixon's method is beyond the scope of this course: it relies on a fine analysis of the De Bruijn function, and makes use of special efficient techniques for sparse matrices. For a well-chosen B and the best known algorithms, the optimal complexity of Dixon's algorithm is

$$\exp\left((2\sqrt{2}+o(1))\sqrt{\ln n \ln\ln n}\right)$$

8.3.3 Quadratic and general number field sieves

The most efficient methods for general factorisation are improvements on Dixon's approach. The simplest is the quadratic sieve, which is the second fastest known algorithm (the fastest for integers under about 100 decimal digits). For larger numbers, the general number field sieve (GNFS) algorithm is faster, although theoretically more subtle. Both algorithms improve on Dixon's way to find x and y.

Quadratic sieve

Essentially, the quadratic sieve speeds up the process of finding x and y in two ways. First let's introduce the quadratic polynomial $f(x) = (\lfloor\sqrt{n}\rfloor + x)^2 - n$. Solutions to $f(x) = 0 \bmod p$ can be efficiently computed for primes p.[12] We choose primes for which n is a quadratic residue modulo p, called the factor basis $\{p_1, \ldots, p_r\}$, which replaces B.

Second is the sieving step. We generate several values $z_i = f(x_i)$, keeping track of both x_i and z_i. For every p in the factor basis, we compute the solution to $f(u) = 0 \bmod p$; for every $x_i = u + kp$, we update $z_i \leftarrow z_i/p$. At the end of this process, some entries z_i are equal to 1: this means that we found a smooth number.

Example 37 Let $n = 15347$, which is a quadratic residue modulo $p_i = \{2, 17, 23, 29\}$.

1. Factor basis. We choose $p_i = \{2, 17, 23, 29\}$ as a factor basis.

2. Quadratic evaluation. We compute $z_i = f(x_i) = (x_i + 124)^2 - 15347$ for $x_i = 0, 2, \ldots, 99$. This gives $z_0 = 29$, $z_1 = 278$, $z_2 = 529$, $z_3 = 782, \ldots, z_{99} = 34382$.

3. Sieving. For every p_i in the factor base, we compute the solutions to $f(u) = 0 \bmod p$. This gives the following

p_i	2	17	23	29
u	1	3, 4	2, 3	0, 13

For every $x_i = u \bmod p$, we update $z_i \leftarrow z_i/p$, keeping track of the division.

4. At this point, every x_i for which $z_i = 1$ provides a smooth couple. In our case, this corresponds to $i = 0$, 3, and 71:

$$\begin{aligned}
x_0 &= 0 & f(x_0) &= 29 & &= 2^0 \cdot 17^0 \cdot 23^0 \cdot 29^1\\
x_3 &= 3 & f(x_3) &= 782 & &= 2^1 \cdot 17^1 \cdot 23^1 \cdot 29^0\\
x_{71} &= 71 & f(x_{71}) &= 22678 & &= 2^1 \cdot 17^1 \cdot 23^1 \cdot 29^1
\end{aligned}$$

At this point we can perform the same linear algebra step as for Dixon, we find that multiplying the three numbers gives us the equality $22678^2 = 3070860^2 \bmod n$, from whence we find that 103 is a factor of n.

[12] For instance, one may use the Shanks–Tonelli algorithm.

The basic outline of the quadratic sieve can be improved in several ways (by using several polynomials, by running in parallel, etc.), and there are non trivial tricks needed to manage operations on very large numbers n. The running time of this algorithm is heuristically

$$\exp\left((1+o(1))\sqrt{\ln n \ln\ln n}\right).$$

For instance, a 1024-bit integer n would be factored with effort about 2^{98} (against 2^{278} using Dixon's method).

Number field sieve

The (general) number field sieve (GNFS) is an improvement over the quadratic sieve. It makes use of irreducible polynomials of small degree and achieves a heuristic complexity of

$$\exp\left(\left(\sqrt[3]{\frac{64}{9}}+o(1)\right)(\ln n)^{\frac{1}{3}}(\ln\ln n)^{\frac{2}{3}}\right)$$

For the same 1024-bit integer as for the quadratic sieve and Dixon's approach, GNFS is expected to provide a factorisation in with effort about 2^86 (against respectively 2^{98} and 2^{278}). Note that at the time of writing, there is still much research being done on improving the GNFS (in particular the choice of polynomials f and g, as well as the parallelisation).

In what follows we give a quick description of the main ideas behind GNFS, keeping in mind that it builds from the quadratic sieve. Let $f(t)$ be a monic polynomial, that is irreducible over $\mathbb{Z}$. Let α be a (possibly complex) root of f.

Now assume that m is an integer such that $f(m) \equiv 0 \bmod n$. There is a natural homomorphism ϕ that transports the ring $\mathbb{Z}[\alpha]$ of algebraic integers to the ring $\mathbb{Z}/n\mathbb{Z}$, and such that $\phi(\alpha) = m \bmod n$. In other terms, if $g(t)$ is any polynomial with integer coefficients, $\phi(g(\alpha)) = g(m) \bmod n$.

Here is how this construction helps us: suppose we could find a set $\mathcal{S}$ of polynomials g with integer coefficients, such that $\prod_{g\in\mathcal{S}} g(\alpha) = \beta^2$ is a square in $\mathbb{Z}[\alpha]$, and such that $\prod_{g\in\mathcal{S}} g(m) = y^2$ is a square in $\mathbb{Z}$. Then we have found

$$x^2 \equiv \phi(\beta)^2 \equiv \phi(\beta^2) \equiv \phi\left(\prod_{g\in\mathcal{S}} g(\alpha)\right) \equiv \prod_{g\in\mathcal{S}} g(m) \equiv y^2 \bmod n. \tag{8.2}$$

From there it is easy to factor n as we did before.

The devil lies in the details. Indeed, $\mathbb{Z}[\alpha]$ is a complicated ring, in which we need a notion of smoothness and primes, etc.

Let's focus first on the right-hand side of Equation (8.2). How to find the g's that we need? Let $d > 1$ such that $n > 2^{d^2}$, and let $m = \lfloor n^{1/d} \rfloor$. Now, writing n in base m we have

$$n = c_d m^d + c_{d-1}m^{d-1} + \cdots + c_0$$

In fact it is easy to see that $c_d = 1$, by the relative size of m and n. Now let

$$f(t) = t^d + c_{d-1}t^{d-1} + \cdots + c_0.$$

Then f is a monic polynomial such that $f(m) = n$. If f is not irreducible, then we can factor f into irreducible components in time polynomial in $\log n$. If the factorisation in non-trivial, then we directly get a factorisation of n itself by substituting m for t! So from now on assume that f is irreducible over $\mathbb{Z}$.

Let $g(t) = a - bt$ with a, b coprime and $0 < b, |a| \le B$. The integers $a - bm$ are small compared to n (because $m \approx n^{1/d}$). It is now easy to sieve to pick up pairs a, b such that $a - bm$ is smooth: We fix b and sieve over a, then choose another b and sieve again, until we exhaust the choices of b. At the end of this procedure we have a square on the right-hand side of Equation (8.2).

Let's now turn our attention to the left-hand side of Equation (8.2). It would be easy if $\mathbb{Z}[\alpha]$ were a unique factorisation domain, since in that case we have a notion of "prime" and we can adapt the construction we know for smoothness. But in general, this is not the case.

Instead, consider the norm map $N : \mathbb{Q}(\alpha) \to \mathbb{Q}$. If $\alpha = \alpha_1, \alpha_2 \ldots, \alpha_d$ are the conjugates in $\mathbb{C}$ of α and $g \in \mathbb{Q}[t]$ then

$$N(g(\alpha)) = \prod_{i=1}^{d} g(\alpha_i).$$

The norm map is multiplicative, and it sends algebraic integers to $\mathbb{Z}$ — in particular, $N(\mathbb{Z}[\alpha]) \subseteq \mathbb{Z}$. From there, we have this simple result:

> If $\mathcal{S}$ is a set of coprime pairs (a, b) of integers, and $\prod_{(a,b)\in\mathcal{S}}(a - b\alpha)$ is a square in $\mathbb{Z}[\alpha]$, then $\prod_{(a,b)\in\mathcal{S}} N(a - b\alpha)$ is a square in $\mathbb{Z}$.

Note that we have, for rational numbers a and $b \neq 0$,

$$\begin{aligned}
N(a - b\alpha) &= \prod_{i=1}^{d}(a - b\alpha_i) \\
&= b^d \prod_{i=1}^{d}\left(\frac{a}{b} - \alpha_i\right) \\
&= b^d f\left(\frac{a}{b}\right) \\
&= a^d + c_{d-1}a^{d-1}b + \cdots + c_1 ab^{d-1} + c_0 b^d
\end{aligned}$$

Thus the norm of $a - b\alpha$ is a polynomial in a, b with integer coefficients.

We shall say that $a - b\alpha$ is smooth if and only if $N(a - b\alpha)$ is smooth. By using a sieve as previously, it should be clear by now that we can find a set $\mathcal{S}$ of coprime integers (a, b) such that $\prod_{(a,b)\in\mathcal{S}} N(a - b\alpha)$ is a square in $\mathbb{Z}$. Namely, we use a sieve to detect pairs a, b where $a - b\alpha$ is k-smooth, create exponent vectors from the prime factorisations of the corresponding norms, use linear algebra modulo 2 to create a square.

The fact that the norm is a square doesn't in itself guarantee that the original number was a square in $\mathbb{Z}[\alpha]$. To give just an example, $2 \pm i \in \mathbb{Z}[i]$ both have norm 5, so their product has norm 5^2. However, $(2 - i)(2 + i) = 5$ is not itself a square.

In general, a prime number $p \in \mathbb{Z}$ may factor into several prime ideals in $\mathbb{Z}[\alpha]$, each having norm a power of p.

8.3.4 Elliptic curve method

The methods described above have in common that their complexity is a funcion of the number's size. But there are other methods that work remarkably well even on large numbers, when such numbers have relatively small prime factors. The most well-known of such methods is Lenstra's elliptic curve factorisation algorithm (henceforth, ECM). Unlike general-purpose algorithms, the ECM's complexity is determined by the size of the largest prime factor, not by the size of the number to be factorised. At the time of writing, the best known algorithm for divisors not greatly exceeding 80 bits.

The method consists in the following steps:

1. On input a number n to be factorised, pick x_0, y_0, a from $\mathbb{Z}/(n)$, and compute $b = y_0^2 - x_0^3 - ax_0 \bmod n$. This amounts to choosing a random elliptic curve over $\mathbb{Z}/(n)$, of equation $y^2 = x^3 + ax + b$ and picking a random point $P = (x_0, y_0)$ on that curve.

2. Using the curve's addition law, we compute $[k]P$. If $[k]P$ and the neutral element become identical modulo some prime divisor $p|n$, then in projective coordinates the z-coordinate of $[k]P$ is divisible by p; this gives a way to find non-trivial factors of n by computing $\gcd(n, z_k)$ where z_k is the z-coordinate of $[k]P$.

The numbers k to try can be chosen in many ways, the simplest approach consisting in enumerating successive integers. If at some point the algorithm fails (i.e. $[k]P = O$ or no factor was found) we can pick another random curve and start over.

In general, the elliptic curve method operates in time

$$\exp\left(\left(\sqrt{2} + o(1)\right)\sqrt{\ln p \ln \ln p}\right)$$

where p is the smallest factor of n. Using well-chosen curves (i.e. not completely random) and leveraging batching, it is possible to further improve this algorithm's performance, both in terms of running time and success probability.

8.4 Exercice set

Exercise 8.1. Implement and analyse an algorithm that uses the Miller–Rabin algorithm to generate an ℓ-bit number, which is prime with probability at least $1 - 2^{\ell}$.

Exercise 8.2. A prime number p is a Sophie Germain prime if $2p+1$ is also prime. By using twice the Miller–Rabin algorithm, construct an algorithm that finds ℓ-bit Sophie Germain primes. Show that this algorithm runs in expected time $O(\ell^5 + k\ell^4)$ and has an error rate of $O(\ell^2/4^k)$.

Exercise 8.3. Implement the algorithm for irreducibility testing use it to find the list of irreducible polynomials of degree 5 in $\mathbb{F}_2$ (there are 14 of them).

Exercise 8.4. Implement the quadratic sieve algorithm and factor a 20-bit integer.

Exercise 8.5. We will make use in this exercise of the famous following theorem:

Theorem 53 (Birthday theorem) *Let $x_0, x_1, \ldots,$ be a sequence of integers sampled uniformly from the set $\{0, \ldots, p-1\}$. Let s be the smallest index such that $x_s = x_i$ for some $i < s$. Then the expected value of s is $O(\sqrt{p})$*

Proof: First note that $\Pr[s \geq j] = \prod_{i=0}^{j-1}(1 - i/p) \leq \prod_{i=0}^{j-1} \exp(-i/p) \leq \exp(-(j-1)^2/2p)$ for any $j \geq 1$. Then

$$\begin{aligned}
\mathbb{E}[s] &= \sum_{j=0} \Pr[s \geq j] = 1 + \sum_{j=1} \Pr[s \geq j] \\
&\leq 1 + \sum_{j=1} \exp(-(j-1)^2/2p) \\
&\leq 2 + \sqrt{2p} \int_0^\infty \exp(-x^2)\, \mathrm{d}x \\
&\leq 2 + \sqrt{2p} \int_0^\infty \exp(-x)\, \mathrm{d}x \\
&= 2 + \sqrt{2p} \\
&= O(\sqrt{p}).
\end{aligned}$$

□

1. Check the provided proof and explain the step at which an integral is introduced. Show that $\int \exp(-x^2)\, \mathrm{d}x = \sqrt{\pi}/2$ and improve the bound in the above theorem.

2. The birthday theorem shows that finding collisions is more likely than what one could think. We can use this idea for factoring integers.[13] Assume that $n = pq$ where p and q are unknown prime numbers and $p < q$.

   We may start by drawing a sequence $x_1, \ldots, x_k$ and check whether there exists any pair x_i, x_j such that $x_i - x_j$ divides n; better still, we consider pairs so that $\gcd(N, x_i - x_j) > 1$. Show that there are $p + q - 2$ possible values for $x_i - x_j$.

3. Show that with the above approach, one succeeds on average after $O(n^{1/4})$ elements x_i.

[13] And also to find collisions in hash functions, to test the limits of pseudo-random number generators, or to solve the discrete logarithm. It really is a useful theorem.

4. Rather than sampling and storing all these elements, we can keep only two values x_0 and x_1. If they don't satisfy $\gcd(n, x_1 - x_0) > 1$, then we update $x_0 \leftarrow x_1$ and $x_1 \leftarrow x_1^2 + a$ for some integer a. Apply this technique to factor $n = 15$ (use $a = 1$ and $a = 2$).

5. It may happen that the sequence generated by iterating $f : x \mapsto x^2 + a$ enters a cycle. If that happens, the algorithm will go on a loop forever. We can fix this by using a cycle detection technique: this gives the following algorithm, known as Pollard's ρ:

   - $x_0 = x_1 = 2$
   - Repeat:
     a) $x_0 \leftarrow f(x_0)$
     b) $x_1 \leftarrow f(f(x_1))$
     c) If $\leftarrow \gcd(n, x_1 - x_0)$ > 1 return $\gcd(n, x_1 - x_0)$
     d) If $x_0 = x_1$ return failure

   Show that if the sequence generated by iterating f repeats, then the algorithm returns failure.

6. Under the assumption that the sequence generated by iterating f is distributed uniformly at random, prove that Pollard's ρ algorithm outputs the smallest factor of n in expected time $O(\sqrt{p})$ and using $O(1)$ space.

Exercise 8.6. The power of randomised methods is remarkable. In this exercise we discuss the Schwartz–Zippel theorem, through the very concrete following question: how efficiently can we check that a polynomial P is the zero polynomial? We will prove the following result:

Theorem 54 (Schwartz–Zippel) *Let F be a field and $P \in F[x_1, \ldots, x_n]$, with $\deg P = d \geq 0$. Assume that $P \neq 0$. Let S be a finite subset of F. Let $r_1, \ldots, r_n$ be points sampled uniformly at random from S. Then*

$$\Pr[P(x_1, \ldots, x_n) = 0] \leq \frac{d}{|S|}.$$

1. Prove the case $n = 1$.

2. Now assume that the theorem holds for all polynomials in $n - 1$ variables. Write P as a polynomial in x_i:

$$P(x_1, \ldots, x_n) = \sum_{i=0}^{d} x_1^i P_i(x_2, \ldots, x_n).$$

   Since $P \neq 0$, there is some i such that $P_i \neq 0$. Take the largest such i, and show that $\deg P_i \leq d - i$. Show that

$$\Pr[P_i(r_2, \ldots, r_n) = 0] \leq \frac{d-i}{|S|}$$
$$\Pr[P(r_1, r_2, \ldots, r_n) = 0 | P_i(r_2, \ldots, r_n) \neq 0] \leq \frac{i}{|S|}.$$

3. Conclude the proof of the Schwartz–Zippel lemma.

4. Assuming that F, G are polynomial of degree d with coefficients in a finite field $\mathbb{F}_q$ such $q > d$. Show that $F = G \Leftrightarrow (F-G)(x) = 0 \forall x \in \mathbb{F}_q$. Construct a (probabilistic) polynomial time algorithm for testing polynomial equality in this case.

   As a counter-example, consider $x^2 - x \in \mathbb{F}_2[X]$.

Testing whether a polynomial evaluates to zero for every field element is a hard problem (known to be **coNP**-hard[14]). What we have shown is a randomised algorithm for solving this problem (and, by the latest question, the polynomial equality problem) in a special case.

> **Open question (2018).** Is there a sub-exponential deterministic algorithm to decide polynomial identity?

[14] Reduction: write a 3SAT formula as a polynomial over $\mathbb{F}_2$ such that the polynomial evaluates to zero everywhere if and only if the formula is unsatisfiable.

Chapter 9

Error-correcting codes

Communication at a distance is never a perfect process. A signal is usually corrupted on its way to the receiving end of a communication channel, the precise nature of this corruption depending on the channel's physical properties. The purpose of error-correcting codes is to send messages so that the receiving end can detect, and even maybe correct, incorrect parts of a message. Thus the physical nature of a channel can be abstracted out, and we can build protocols on top of this abstraction.

A simple, yet ineffective approach consists in repeating the message multiple times; algebraic tools provide much more effective codes, as well as important theorems about the properties of wide classes of solutions.

TODO

Figure 9.1: TODO: illustration.

Contents

9.1 Basic definitions

Definition 53 (Code) A code $C = (X, Y, \mathsf{Enc} : X \to Y, \mathsf{Dec} : Y \to X)$ is the data of two sets and two algorithms, respectively called the set of words, set of codewords, encoding algorithm, and decoding algorithm, satisfying the following correctness property:

$$\forall x \in X, \quad \mathsf{Dec}\left(\mathsf{Enc}(x)\right) = x.$$

Remark 61 Some authors use a more restrictive definition of a code, sometimes confusing Y with the code itself. This makes sense when working with codes in systematic form (see below), because encoding is straightforward. We adopt here a slightly more general setting, and will therefore call "code" the 4-tuple $C = (X, Y, \mathsf{Enc} : X \to Y, \mathsf{Dec} : Y \to X)$, refering to Y itself as the "codebook" or the set of codewords.

Example 38

- The trivial code on a set of words X is $1_X = (X, X, \mathsf{id}, \mathsf{id})$.
- If $C_1 = (\mathsf{Enc}_1, \mathsf{Dec}_1, X_1, Y_1)$ and $C_2 = (\mathsf{Enc}_2, \mathsf{Dec}_2, X_2, Y_2)$ are codes, and $Y_1 \simeq X_2$, then C_3 defined by

  $$\mathsf{Enc}_3 = \mathsf{Enc}_2 \circ \mathsf{Enc}_1, \qquad \mathsf{Dec}_3 = \mathsf{Dec}_1 \circ \mathsf{Dec}_2, \qquad X_3 = X_1, \qquad Y_3 = Y_2$$

  is a new code.
- Let k be an odd integer, $X = \{0, 1\}$, $Y = \{0, 1\}^k$, Enc being the function that takes 0 (resp. 1) to the binary string made of k zeros (resp. ones), and Dec being the majority function, which returns whichever digit is most represented in its input. This defines the k-repetition code R_k.

Remark 62 In general, the message to be transmitted will be written using a fixed number of symbols k from a fixed alphabet F, i.e., we will have $X \subseteq F^k$. Similarly, the codewords can be considered as words written from an alphabet E, i.e., $Y \subseteq E^n$. When $|E| = |F|$, as will often be the case, the ratio k/n is called the code's information rate.

Definition 54 (Hamming distance) Let $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ two elements from a set E^n. The Hamming distance between x and y is the number of indices i such that $x_i \neq y_i$. It is denoted $d_H(x, y)$.

Definition 55 (Minimal distance) Let $C = (X, Y, \ldots)$ be a code. The minimal distance of C is the integer

$$d = \min_{y_1 \neq y_2 \in Y} d_H\left(y_1, y_2\right).$$

Example 39

- The k-repetition code has minimal distance k (in fact the distance between any two codewords of this code is k).
- The following code has minimal distance $d = 3$:

  $$Y = \{\mathsf{010101}, \mathsf{101010}, \mathsf{111111}, \mathsf{000000}\}.$$

We can now address the question we set out to deal with, namely the problem of imperfect transmission of information. The situation can be modelled as follows: the sender wishes to emit some message $m \in X$; to do so, she uses a code C and actually sends $c \in Y$ over the channel; but the channel is imperfect and therefore the recipient receives $y' \in Y$.

Theorem 55 (Correction capacity) *Let d be the minimal distance of C, then up to t errors can be corrected, where*

$$t = \left\lfloor \frac{d-1}{2} \right\rfloor .$$

In that case C is called a t-error correcting code.

Proof: Let $B(y,\rho)$ denote the sphere of center y and radius ρ, i.e.,

$$B(y,\rho) = \left\{ y' \in F^n \mid d_H(y,y') \leq \rho \right\} .$$

By definition of d, for any $y \neq y' \in Y$, the spheres $B(y,t)$ and $B(y',t)$ are disjoint. Therefore it is not possible to have $y' \in F^n$ such that $d_H(y,y') < t$ and y, y' belong to separate spheres. $\square$

Remark 63 A code such that $F_n = \bigcup_{y \in Y} B(y,t)$ is called a perfect code. Needless to say, this does not happen often; in general, there are elements of F^n that cannot be safely attributed to an original codeword: we know that there was a mistake, but cannot correct it.

Example 40

- The k-repetition code has correcting capacity $t = (k-1)/2$.
- The following code can detect 2 errors and correct at most 1:

$$Y = \{010101, 101010, 111111, 000000\} .$$

9.2 Linear codes

9.2.1 Notations, generating and parity matrices

Definition 56 (Linear code) Let $p \in \mathfrak{P}$, $\ell \geq 1$, $q = p^\ell$. A (n,k,d)-linear code is a code of minimal distance d such that Y is a k-dimensional vector subspace of $\mathbb{F}_q^n$.

Remark 64 Since linear codes are very common, it is useful to recall standard names for the parameters:

- n is called the code's length
- k is called the code's dimension
- d is always the minimal distance.

In fact, linear codes are so common that some authors do not bother being explicit; they may refer to $[n,k,d]_q$ codes, which quite often corresponds to a (n,k,d)-linear code over $\mathbb{F}_q$.

Remark 65 The minimal distance of a linear code can be expressed as $d = \min_{0 \neq y \in Y} d_H(y,0)$.

Example 41 The code defined by

$$Y = \{\texttt{0000}, \texttt{1100}, \texttt{1010}, \texttt{1001}, \texttt{0110}, \texttt{0101}, \texttt{0011}, \texttt{1111}\}$$

is a linear code of length $n = 4$ and dimension $k = 3$ over $\mathbb{F}_2$, it has minimal distance $d = 2$.

Theorem 56 (Singleton) *Let C be a (n,k,d)-linear code. Then $d \leq n - k + 1$.*

Proof: Let (e_i) be the canonical basis of $\mathbb{F}_q^n$. Let $E = \langle e_1, \dots, e_{n-k+1} \rangle$. Since $\dim E + \dim Y > n$, $E \cap Y$ is non-empty and not reduced to 0; therefore there exists $y \in Y$ such that $d_H(y, 0) \leq n - k + 1$. □

Definition 57 (Maximum distance separable code) A (n, k, d)-linear code C that satisfies $d = n - k + 1$ is called a maximum distance separable, or MDS code.

Remark 66 A consequence of Singleton's theorem is that $d/n + k/n < 1 + 1/n$, which means that one cannot hope to have linear codes with high correction capacity (i.e., large d) and simultaneously high information rate (i.e., small n). In that regard MDS codes are the linear codes that strike the optimal balance.

Definition 58 (Generating and parity matrices) The generating matrix of C is the matrix G of a basis of Y. A parity matrix for G is a matrix H such that $GH^\top = 0$.

Remark 67 Thus for linear codes, encoding is performed by matrix multiplication, which is hopefully well known by the reader. We will therefore not describe the Enc algorithm in detail.

9.2.2 Decoding linear codes

Remark 68 Consider a linear code generated by G, and the corresponding parity matrix H. A message m is encoded as $c = mG$, and received with some error: $c' = mG + r$. Now by computing $cH^\top = mGH^\top + rH^\top = rH^\top \neq 0$, we already detect an error. The remainder, $s = rH^\top$ is called the syndrome. To correct the error (if possible), we have to find r such that $d_H(r, 0) \leq t$ and $rH^\top = s$. This approach is called syndrome decoding.

Example 42 Consider the following matrix:

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Note that H is a rank 3 matrix in $M_{3,7}(\mathbb{F}_2)$. It is the parity check matrix of the code $C = \ker H$, which is of distance 3, and therefore correction capacity $t = 1$. Syndrome decoding works as follows: assume that we received the message

$$c' = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{F}_2^7.$$

We know that there was an error during transmission, as $s = Hc' = \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \neq 0$. Furthermore, s turns out to be the third column of H, so that we have $Hc' = Hr$ for

$$r = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Thus the original codeword was

$$c = c' - r = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

The reader may (and should!) wonder how efficiently a technique such as syndrome decoding can be implemented. Concretely, the problem is the following: given a parity check matrix H, a syndrome s, and some integer w, decide whether there is a vector r such that $rH^\top = s$, of Hamming weight at most w.

Theorem 57 (Berlekamp–McEliece–van Tilborg) *The decision problem of decoding linear codes is* **NP***-complete.*

Proof: The proof is beyond the scope of this book; in short it reduces the three-dimensional matching problem (which is **NP**-complete) to the the decision problem of decoding linear codes. □

The above result (dating from 1978) had two interesting consequences: first, it made it necessary to look for special families of linear codes which are easier to decode! and second, it can be used to design code-based cryptosystems, some of which have regained interest in recent years.

9.2.3 Systematic form

Definition 59 (Systematic form) A (n, k, d)-linear code C is systematic if there exists a matrix B with k lines and $n-k$ rows, such that $G = (I_k|B)$ is a generating matrix for C. If such a matrix exist, it is unique, and the corresponding parity matrix is equal to $H = (-B^\top|I_{n-k})$. We refer to such G and H as matrices in systematic form.

Remark 69 One interest of codes in systematic form is that one can directly "read" the codewords, as the first k components are just the original message. The remaining $n-k$ components form the "redundancy".

Example 43 Consider the following generating matrix:

$$G = \begin{pmatrix} \overbrace{\boxed{\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}}}^{I_3} & \begin{matrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{matrix} \end{pmatrix}$$

in systematic form. To encode a message $m \in \mathbb{F}_2^3$, we simply compute mG (which is equivalent here to taking the row of G corresponding to m, when $m \neq 0$). Every codeword thus has length 5, and the code's minimal distance is $d = 2$.

Theorem 58 (Characterisation of systematic codes) *Let C be a linear code and $G = (g_{i,j})$ a generating matrix for C. Then C is systematic if and only if the matrix $\widetilde{G} = (g_{i,j})_{1 \le i,j \le k}$ is invertible.*

Proof: If C is systematic, and G in systematic form, then $\widetilde{G} = I_k$ is clearly invertible. If G is not in systematic form then it is related to the systematic form matrix by an invertible transformation, and is therefore invertible.

Conversely, if $\widetilde{G}$ is invertible, then using Gauss' reduction we bring G to systematic form while keeping G a generating matrix at each operation. □

Definition 60 (Equivalent codes) Two linear codes C_1 and C_2 are equivalent if there exists a permutation $\sigma \in \mathfrak{S}_n$ such that every $\sigma(Y_2) = Y_1$, where σ acts as $\sigma((x_1, \ldots, x_n)) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ on each element.

Theorem 59 (Equivalence with a systematic code) *Every code is equivalent to a systematic code.*

Proof: Let C be a (n, k, d)-linear code on $\mathbb{F}_q$, with generating matrix G. There exists a $k \times k$ submatrix $\widetilde{G}$, extracted from G, which is invertible, i.e., if g_i denote the columns of G, then the columns of $\widetilde{G}$ are g_{j_i} with $j_1 < \cdots < j_k$. Let $\sigma \in \mathfrak{S}_n$ be defined by

$$\sigma(i) = \begin{cases} j_i & \text{if } 1 \leq i \leq k \\ i & \text{if } k+1 \leq i \leq n \end{cases}$$

On the one hand this generates a code that is equivalent to C, with permutation σ. On the other hand the matrix $\widetilde{G}$ is invertible and therefore the code is systematic. □

Example 44 Consider the following generating matrix

$$G = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

This matrix defines a binary linear code of length 5, dimension 3, and distance 1; but it is not systematic as the 3×3 submatrix extracted from G is not invertible. But we can construct an equivalent code by reordering the columns from G, giving

$$G' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix}$$

We can then use Gauss' reduction to put G' in systematic form

$$G'' = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

In particular, the length and dimension are left unchanged, and the distance is still 1, as clearly visible from G''.

9.2.4 The binary Hamming code

Definition 61 Let $r > 0$, α be a generator of $\mathbb{F}_{2^r}^{\times}$ of order $n = 2^r - 1$, and consider the map $f : \mathbb{F}_2^n \to \mathbb{F}_{2^r}$ defined by

$$f(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \alpha_i.$$

Let $Y = \ker f$, which is a vector subspace of $\mathbb{F}_2^n$. This gives a (n, k, d)-linear code C known as the binary Hamming code of length $2^r - 1$.

Proposition 10 (Minimal distance of the binary Hamming code) *The binary Hamming code of length $2^r - 1$ has minimal distance $d = 3$.*

Proof: If there was $y \in Y$ satisfying $d_H(y, 0) = 1$, then we would have $\alpha^i = 0$ for some i, which cannot be; similarly $d_H(y, 0) = 1$ would imply $\alpha^{j-i} = 1$ for some $1 \leq i < j \leq n$. Therefore the minimal distance of C is $d \geq 3$.

Let G be a generating matrix for C and H the corresponding parity matrix. Every pair of columns c_i, c_j of H is distinct, and therefore H contains each non-zero r-column vectors with coefficients in $\mathbb{F}_2$. Thus adding any two columns of H results in a column that is already in H, i.e., $d \leq 3$. □

9.3 Cyclic codes

Definition 62 (Cyclic code) A code is cyclic if any rotation of a codeword is again a codeword.

Example 45 The k-repetition code is cyclic.

Proposition 11 (Characterisation of cyclic codes) *A (n, k, d)-linear code C is cyclic if and only if its image by the isomorphism $\psi : \mathbb{F}_q^n \simeq \mathbb{F}_q[X]/(X^n - 1)$ defined by*

$$\psi(x_1, \dots, x_n) = x_1 X^{n-1} + \cdots + x_{n-1} X + x_n$$

is an ideal.

Proof: A rotation corresponds to multiplication by X, so C is cyclic if and only if $\psi(Y)$ is stable under multiplication by X, and therefore by any element from $\mathbb{F}_q[X]$. It is, in other words, an ideal. □

Corollary 26 *The set of cyclic linear codes of length n is in bijection with monic polynomials that divide $X^n - 1$.*

Remark 70 This means that we can refer to a cyclic code C by this polynomial g_C, called the generating polynomial of C. Concretely, we can encode a message as follows: let $m = (x_1, \dots, x_k) \in \mathbb{F}_q^k$, and define

$$p_m = c_1 X^{n-1} + \cdots + x_k X^{n-k}.$$

Compute the Euclidean division of p_m by g_C, which gives a remainder p_r. The encoding is therefore $p_m + p_r$, which is in systematic form.

TODOfinish cyclic codes

9.4 The family of BCH codes

(TODO: explain BCH)

9.5 Exercise set.

Exercise 9.1. Credit card numbers are encoded using the Luhn code, which works as follows: a last digit is added, so that the sum of all digits (including the last one) modulo 10 gives zero. Is that a code? It is linear, cyclic? How many errors can it detect? How many can it correct?

Exercise 9.2. Let C be a linear code, with codeword matrix Y. We define

- The dual code $C^\dagger$ given by $Y^\dagger = \{y' \in \mathbb{F}_q^n \mid \forall y \in Y, y \cdot y' = 0\}$.
- The extended code C^+ given by $Y^+ = \{(x_1, \dots, x_{n+1}) \in \mathbb{F}_q^{n+1} \mid (x_1, \dots, x_n) \in Y, x_1 + \cdots + x_{n+1} = 0\}$.

Show that there are indeed codes. Show that $(C^\dagger)^\dagger$ is equivalent to C. For each code, compute the minimal distance, then give a generating and parity matrix. If a code C is self-dual, i.e., $C^\dagger = C$, show that Y is also the parity check matrix.

Exercise 9.3. Show that the Hamming code of length $7 = 2^3 - 1$ is perfect, has a correction capacity of 1, but is not MDS.

Exercise 9.4. The purpose of this exercise is to study a correcting code used for the French Minitel, which we will refer to as the Minitel code.

1. Let $P = X^7 + X^3 + 1$. Show that $\mathbb{F}_2[X]/(P) \simeq \mathbb{F}_{128}$, and that X is a generator of $\mathbb{F}_{128}^\times$.
2. To send a 15-byte message, i.e., 120 bits $a_0, a_1, \dots, a_{119} \in \mathbb{F}_2$, we write
$$c = a_0 X^{126} + \cdots + a_{119} X^7$$
which can also be written $c = a_{120} X^6 + \cdots + a_{126}$. The codeword sent is $a_0, a_1, \dots, a_{126}, a_{127}$, where a_{127} is a parity bit computed so that $a_0 + \cdots + a_{127} = 0$. (Thus the complete codeword fits in 16 bytes).
   a) How many errors can the parity bit detect? Correct?
   b) We receive $a'_0, \dots, a'_{126}, a'_{127}$ with a correct parity bit. Show that $a'_0 X^{126} + \cdots + a'_{126} = 0$.
   c) Assume that there is only one error, i.e., a single $a'_i \neq a_i$. Can one recover the original message?

Exercise 9.5. Show that the set of cyclic linear codes of length n is in bijection with the set
$$S_q = \{I \subseteq (\mathbb{Z}/(n))^\times \mid \forall u \in I, uq \in I\}.$$
.

Exercise 9.6. Consider the polynomial $P = x^{11} + x^{10} + x^6 + x^5 + x^4 + x^2 + 1$ in $\mathbb{F}_2[x]$.

- Show that P divides $x^{23} - 1$, and deduce that P generates a cyclic code. This code is known as the perfect binary Golay code G_{23}, which has deep connections to many mathematical questions.
- Show that G_{23} is perfect, i.e., that the spheres of radius three around code words form a partition of the vector space.

- Show that G_{23} is a $[23, 12, 7]_2$-linear code.
- The ternary Golay code G_{12} is generated in systematic form by thee matrix $[I_6|A]$ where

$$A = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 & 2 & 1 \\ 1 & 1 & 0 & 1 & 2 & 2 \\ 1 & 2 & 1 & 0 & 1 & 2 \\ 1 & 2 & 2 & 1 & 0 & 1 \\ 1 & 1 & 2 & 2 & 1 & 0 \end{pmatrix}$$

  Show that G_{12} is a linear $[12, 6, 6]_3$-code, that it is self-dual, and that the code obtained by removing a column from A is perfect, it is called G_{11} and ii $[11, 6, 5]_3$-linear.
- Show that adding a parity check bit to codewords in G_{11} gives G_{12}. Similarly, adding a parity check bit to G_{23} gives the code G_{24}. Show that G_{24} is self-dual.
- Exhibit a parity check matrix for G_{12} and G_{11}.

Exercice 9.7. The purpose of this exercise is to construct the Reed–Muller codes.[1] Let q be a prime power, and two integers $n > 0$ and $0 < d < q - 1$. A message in our context is a vector $m = (m_1, m_2, \ldots, m_k) \in \mathbb{F}_q^k$, where $k = \binom{n+d}{d}$.

1. Let $w_1, \ldots, w_k$ be a fixed set of points of $\mathbb{F}_q^n$. Show that it is always possible to construct a polynomial $P_m \in \mathbb{F}_q[X_1, \ldots, X_n]$ of degree at most d, such that for each $i = 1, \ldots, k$ we have $P_m(w_i) = m_i$.

   For a message m, the codeword is generated by evaluating the polynomial P_m at each point of $\mathbb{F}_q^n$. A codeword has therefore length q^n.
2. Assuming that the codeword is not corrupted during transmission, how would we proceed to recover m?
3. We do not assume any more that the codeword is pristine, let it be denoted by C', with the convention that $C'(x)$ is the x-th element of C'. (If no transmission error occured, then $C'(x) = P_m(x)$ for all x.) We do the following:

   a) Let v be a point in $\mathbb{F}_q^n$ sampled uniformly at random, and S be a subset of $\mathbb{F}_q^\times$ of size $d + 1$. Let $U = \{w + sv \mid w \in \{w_1, \ldots, w_n\}, s \in S\}$.

   b) Find the (unique) polynomial f of degree at most d such that for every $u \in U$, $f(u) = C'(u)$.

   c) Output $f(0)$.

   If a fraction δ of the codeword was corrupted, what is the probability that $C'(u) = P_m(u)$ for every $u \in U$? When that happens, what is the value returned by the above procedure?
4. Show that it is possible to detect (resp. correct) an error in a codeword without knowing the whole codeword. This feature makes Reed–Muller codes locally testable (resp. decodable).
5. Can we detect a decoding failure? What could we do if that happens?

[1] The attentive reader may recognise that an earlier exercise on Shamir's secret sharing follows exactly the same idea.

Chapter 10

Digital signature schemes

A digital signature should play the role of traditional signatures, i.e., authentify a document as being produced (or recognised as valid) by an identified person. Fundamentally, this requires this person to produce some information — some number — in a way that other people cannot. The first public solution was proposed in 1978 by Rivest, Shamir, and Adleman; with minor adjustments it is still used at the time of writing. In this chapter, we introduce a formalism to discuss digital signature schemes and their security, as well as some examples.

TODO

Figure 10.1: TODO: illustration.

Contents

10.1 Exercice set

TODO

Chapter 11

Problem sets

11.1 Additive polynomials

Let $p \in \mathfrak{P}$ and K be a ring of characteristic p.

1. Let $x, y \in K$. Show that
$$(x + y)^p = x^p + y^p$$
The map $\tau_K : K \to K$ defined by $x \mapsto x^p$. The above equation can be restated as
$$\tau_K(x + y) = \tau_K(x) + \tau_K(y)$$
for all $x, y \in K$.

2. Show that τ_K is a ring morphism. It is called the Frobenius endomorphism of K.

3. Let L be a ring of characteristic p, and $\alpha : K \to L$ be a ring morphism. Show that
$$\alpha \circ \tau_K = \tau_L \circ \alpha$$
which can be restated by saying that the following diagram of rings and ring morphisms commutes:
$$\begin{tikzcd} K \arrow[r, "\tau_K"] \arrow[d, "\alpha"'] \arrow[dr, phantom, "\circlearrowleft"] & K \arrow[d, "\alpha"] \\ L \arrow[r, "\tau_L"'] & L \end{tikzcd}$$
(We say that the Frobenius morphism is a natural transformation.)

4. Show that if K is a field, then τ_K is injective. We will henceforth assume that K is a field, i.e., $K \simeq \mathbb{F}_p$.

5. Show that for all $x \in K$, $\tau_K(x) = x$. Deduce that the p roots of the polynomial $X^p - X$ are the elements of K

6. A polynomial $P \in K[X]$ such that for all $x, y \in K$ we have $P(x + y) = P(x) + P(y)$ is called an additive polynomial. Show that τ_K defines an additive polynomial. Show that for all $n > 0$, $\tau_K^n = \tau_K \circ \cdots \circ \tau_K$ defines an additive polynomial.

7. Show that the set $T_K = \{\mathrm{id}, \tau_K, \tau_K^2, \ldots\}$, together with the usual addition and function composition (instead of usual multiplication) form a commutative ring (it is denoted $K\{\tau_K\}$).

8. Let $P \in K[X]$ of degree n having distinct roots $\{\omega_1, \dots, \omega_n\} = \Omega \subset K$. Prove that if P is additive, then Ω is an additive subgroup of K.

9. Assume $p \neq 2$, show that the polynomial $Q = X + (X^p - X)^2 k$ is additive in $\mathbb{F}_p$ for $k \geq 0$, but that its roots do not form an additive group.

The study of additive polynomials is related to that of linear codes. Indeed, there is an equivalence between the set of linear codes over $\mathbb{F}_{p^n}$ and the elements of the ring $\mathbb{F}_{p^n}\{\tau\}$ — in particular, there is a ring structure on linear codes; this ring can be studied in its own right, and the results (divisibility, etc.) can be transported back to codes. As an example, factorisation in the ring of codes shows that every k-dimensional binary code can be decomposed as a product of k one-dimensional subcodes (although not in a unique fashion).

11.2 Fermat–Wiles' theorem, case $n = 4$

In this problem we consider the equation

$$x^4 + y^4 = z^4.$$

As a preliminary, we are going to show that if a triangle has a right angle and integer sides, then its area cannot be a perfect square.

1. Let p, q be primes, with $p > q$, and let

$$\begin{aligned} a &= 2pq \\ b &= p^2 - q^2 \\ c &= p^2 + q^2. \end{aligned}$$

   Show that $a^2 + b^2 = c^2$.

2. Show that the area of a right triangle with integer sides is $A = pq(p + q)(p - q)$.

3. Show that if A is square, then there exist x, y, u, v pairwise coprime integers such that

$$\begin{aligned} p &= x^2 \\ q &= y^2 \\ p + q &= u^2 \\ p - q &= v^2 \end{aligned}$$

   And deduce that $2y^2 = (u + v)(u - v)$.

4. Show that (up to swapping u and v), there are integers r, s such that

$$\begin{aligned} u + v &= 2r^2 \\ u - v &= 4s^2 \end{aligned}$$

   and therefore $x^2 = r^4 + 4s^4$.

5. Let E be the set of integers c which correspond to the hypothenuse of a right triangle with integer sides whose area is a perfect square. Assume that E is not empty. Then it has a smallest element $c = p^2 + q^2$. Use the above to construct a strictly smaller solution and derive a contradiction.

We are now equipped to prove Fermat–Wiles' theorem for $n = 4$.

1. Show that without loss of generality, we may assume x, y, z are pairwise coprime, and x even.
2. Write the equation as $x^4 = (z^2 - y^2)(z^2 + y^2)$, and show that there is no solution by using the preliminary.
3. Show that this implies that to prove Fermat–Wiles' theorem in general, it suffices to show it for primes $p > 3$.

11.3 Fermat–Wiles' theorem, case $n = 3$

In this problem we consider the equation

$$x^3 + y^3 = z^3.$$

1. Show that without loss of generality, we may assume x, y, z pairwise coprime.
2. We assume z is even, and write $x = a + b$, $y = a - b$ with integers a, b. Show that

$$\frac{a}{4}(a^2 + 3b^2) = \left(\frac{z}{2}\right)^3$$

   and deduce from this that $a \equiv 0 \bmod 4$.
3. Assume that $3 \nmid z$.
   a) Show that there exist integers r, s such that

$$\begin{aligned} a &= 4r^3 \\ a^2 + 3b^2 &= s^3 \end{aligned}$$

   b) Show that there exists u, v coprime integers such that

$$\begin{aligned} a &= u(u + 3v)(u - 3v) \\ b &= 3v(u^2 - v^2) \end{aligned}$$

   c) Deduce from this that there are integers A, B, C such that

$$\begin{aligned} u &= 4A^3 \\ u + 3v &= B^3 \\ u - 3v &= C^3 \end{aligned}$$

   d) Show that this implies $(-2A)^3 + B^3 + C^3 = 0$, which is a solution to Fermat's equation, with strictly smaller integers than x, y, z.
4. Now assume $3 \mid z$.
   a) Show that

$$\frac{a}{36}\left(b^2 + 3\left(\frac{a}{3}\right)^2\right) = \left(\frac{z}{6}\right)^3.$$

   b) Show that the two left hand side multiplicands are therefore coprime cubes.

c) Show that this implies that there exist u, v coprime such that

$$\begin{aligned} a &= 36r^3 \\ b &= u(u+3v)(u-3v) \\ \frac{a}{3} &= 3v(u^2 - v^2) \end{aligned}$$

d) Show that $2v$, $v + u$, and $v - u$ are cubes, and construct a solution to Fermat's equation that is strictly smaller than x, y, z.

5. Conclude.

11.4 Cryptanalysis of the DVD encryption system

In the 1980s, when designing DVDs for movie contents, it was thought that some protection should be put to protect the contents. This gave rise to the Content Scrambling System (CSS), an inexpensive but very weak cipher, in line with the previous century's export legislation (namely, less than 40-bit security). It relies on the use of linear feedback shift registers (LSFRs), and easy-to-implement primitive (see Figure 11.1):

1. Let $s = (b_{n-1}, \ldots, b_0) \in (\mathbb{F}_2)^n$ be the 'initial state';
2. Repeat ℓ times:
   a) Output b_0;
   b) Compute $b \leftarrow b_{v_1} + b_{v_2} + \cdots + b_{v_d}$;
   c) Set $s \leftarrow (b, b_{n-1}, \ldots, b_1)$.

The parameters n, $V = \{v_1, \ldots, v_d\}$, ℓ, and the initial state s entirely determine the output from this algorithm. We therefore refer to it as $\mathrm{LSFR}_{n,V,\ell,s}$, and by convention treat the output as a number in $\{0, \ldots, 2^\ell - 1\}$.

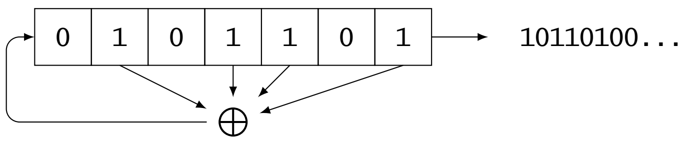


Figure 11.1: Illustration of an LSFR. Each clock cycle, a new bit is output.

1. Show that is the initial state is $s = (0, \ldots, 0)$ then the output is $(0, \ldots, 0)$.
2. Show that, given n consecutive bits of the output, it is possible to find all subsequent output bits.
3. The CSS algorithm works as follows:
   a) Let $A \in (\mathbb{F}_2)^{16}$ and $B \in (\mathbb{F}_2)^{24}$ be the initial states (usually represented as a single 40-bit seed);
   b) Let $s_1 \leftarrow 1\|A$ and $s_2 \leftarrow 1\|B$, where $\|$ denotes concatenation[a] and initialise $L_1 = \mathrm{LSFR}_{17,V_1,8,s_1}$, $L_2 = \mathrm{LSFR}_{25,V_2,8,s_2}$.
   c) Let $c \leftarrow 0$.
   d) Repeat forever

i. Let x be the output of L_1 and y be the output of L_2;
ii. Output $x + y + c \bmod 2^8$;[b]
iii. If $x + y > 255$ then $c \leftarrow 1$ else $c \leftarrow 0$.;

[a] Following Boneh and Shoup, we present here a variant, which is identical from a security standpoint, but slightly easier to work with. In the real CSS, instead of prepending a 1 to the initial seeds, one inserts the 1 in bit position 9 for the 17-bit LFSR and in bit position 22 for the 25-bit LFSR.

[b] The real CSS discards the first byte output by the 17-bit LFSR and the first two bytes output by the 25-bit LFSR. This is without effect for the rest of the discussion.

with $V_1 = \{0, 14\}$ and $V_2 = \{0, 3, 4, 12\}$. Implement this algorithm.

4. Guessing the seeds A, B naively requires of the order of 2^{40} trials; approximately how long would such a computation require?

5. Assume we know x_1, x_2, x_3, the first three bytes output by L_1. Let z_1, z_2, z_3 be the first three bytes output by the CSS algorithm. Show that one recovers the seed B by computing

$$\left(2^{17} + 2^{16} z_3 + 2^8 z_2 + z_1\right) - \left(2^{16} x_3 + 2^8 x_2 + x_1\right).$$

6. Since (x_1, x_2, x_3) is determined (but not uniquely!) by A, show that one recovers the full seed in at most 2^{16} operations. Approximately how long would such a computation require? Implement this algorithm.

7. Show that, as claimed, the simplifications made in our description of CSS are minor, by implementing the attack against the real CSS algorithm.

Index