Introduction to Abstract Algebra
"Groups First"

Bruno Benedetti
University of Miami

Fall 2025

Abstract

The main purpose of these notes is to understand what $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are, as well as their polynomial rings. The official textbook this year is Shahriri, *Algebra in Action*, AMS.

Contents

0 Preliminaries **2**
0.1 Natural numbers, induction, and primes . . . 2
0.2 Euclidean division, unique factorization, and the Euclidean algorithm . . . 5
0.3 Modular arithmetics . . . 8
0.4 Functions and Quotients. The sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{Z}_m$ as quotients . . . 11
0.5 Optional reading: From $\mathbb{Q}$ to $\mathbb{R}$ and $\mathbb{C}$: The needs of geometers . . . 18
0.6 Exercises . . . 25

1 Permutations and Matrices **28**
1.1 Permutations . . . 28
1.2 Matrices and determinants . . . 32
1.3 Exercises . . . 35

2 Abstract groups **36**
2.1 Definition, examples, and first properties of groups . . . 36
2.2 Subgroups and Lagrange's theorem . . . 38
2.3 Period and cyclic subgroups . . . 39
2.4 Group homomorphisms . . . 44
2.5 Exercises . . . 47

3 Normal subgroups, quotients, and Abelian groups **48**
3.1 Normal subgroups . . . 48
3.2 Quotients and the First Isomorphism Theorem . . . 49
3.3 Abelian Groups and the Chinese Remainders theorem . . . 52
3.4 Finite Abelian groups are products of cyclic groups . . . 56
3.5 Exercises . . . 58

4 C-rings, Fields, Domains, and Polynomials **60**
4.1 Commutative Rings 60
4.2 Invertible elements and Fields 61
4.3 Zerodivisors and Domains 62
4.4 Polynomials 65
4.5 Division of polynomials and cyclicity of U_p 68
4.6 Exercises 72

0 Preliminaries

In this section we briefly recall a few topics you probably know already. But even if you "know too much", it's good to agree on notation and terminology, so that we are all on the same page.

0.1 Natural numbers, induction, and primes

You are probably all familiar with the infinite set of natural numbers (also known as "nonnegative integers")

$$\mathbb{N} = \{0, 1, 2, 3, \ldots, n, n+1, \ldots\}$$

It is usually stated in textbooks that natural numbers "come from Nature", whence the name "natural". Leopold Kronecker[1] once stated "God created the integers; all else is the work of man.". This is not entirely true: To accept them, three important abstraction steps are necessary. These steps are non-trivial, as throughout the history of mankind, not all populations have achieved or accepted them:

- the notion of *cardinality*, i.e. the realization that two finite sets in bijection with one another have something in common; whence the *names* of numbers. This is not universal: Even today, the Pirahã people in Amazonas, Brazil, have no names for numbers, and have no linguistic way of expressing exact quantity, not even "one". [2]
- the notion of *zero*, as the cardinality of an "empty set". The ancient Greeks had no symbol for zero, for example; Mayas did have a symbol for zero around the year 36 BC, using it as placeholder in their base-20 numerical system; arithmetic operations with zero were first introduced by the Indian mathematician Brahmagupta[3], around 650 AD.
- the existence of an *infinite set*, that is, a set that can be in bijection with a proper subset of itself. The bijection in this case is the *successor* map, that is, the map that adds one to each element; so an equivalent way of formulating this principle is, "the belief that every number has a successor". Once again, this intuition is not universal, and in logic there is a movement of logicians from around 1900, called *(strict) finitists*, who rejected it.

Given two natural numbers a and b, we say that "a *divides* b" (or equivalently that "a *is a divisor of* b", or equivalently that "b *is a multiple of* a") if there exists a natural number k such that

$$b = k \cdot a.$$

Prime numbers are the natural numbers with exactly two distinct divisors (so 1 is not prime!):

$$2, 3, 5, 7, 11, 13, \ldots$$

Definition 1. Given two natural numbers a and b, their *greatest common divisor*, denoted by $\gcd(a, b)$ is the largest integer that divides both a and b. It exists as long as a, b are not both 0. Two natural numbers a, b are called *coprime* if $\gcd(a, b) = 1$.

Example 2. If p is prime, then it has only 1 and p as divisors. Thus for any natural number n

$$\gcd(p, n) = \begin{cases} p & \text{if } p \text{ divides } n, \\ 1 & \text{otherwise.} \end{cases}$$

[1] H. M. Weber's memorial article, Leopold Kronecker, in Jahresbericht der Deutschen Mathematiker-Vereinigung, Vol. 2 1891-92

[2] Frank et al., *Number as a cognitive technology: Evidence from Pirahã language and cognition*, Cognition 108 (2008), 819–824.

[3] Wallin, Nils-Bertil,"The History of Zero". YaleGlobal online, 19 November 2002

Induction is a standard technique to prove a statement for infinite subsets of $\mathbb{N}$. It is based on the fact that every natural number is obtained from 0 by adding 1 sufficiently many times. So if we show that a property P holds for zero and is maintained when we move from any number to its successor, then P holds also for one, for two, for three.... and eventually is shared by all natural numbers.

Formally, a proof by induction consists of two parts:

- ("Basis") We prove that the statement holds true for a specific integer n_0.
- ("Step") We prove that, if there exists a natural number n such that the statement holds for n, then the statement must hold also for $n+1$.

Once again, the validity of the statement for n_0 implies the validity for n_0+1, which in turn implies the validity for n_0+2, and so on: A domino effect, which eventually proves the statement for all integers $n \geq n_0$. If our basis was $n_0 = 0$, then we end up proving the statement for the whole of $\mathbb{N}$.

Example 3. Let us prove by induction that

$$\sum_{i=0}^{n+1} i = \binom{n+2}{2} \text{ for all } n \in \mathbb{N}.$$

- ("Basis") For $n = 0$, the formula above boils down to $0 + 1 = \binom{2}{2}$, which is correct. This brings good luck!
- ("Step") Let us assume that $\sum_{i=0}^{n+1} i = \binom{n+2}{2}$ holds true for some n. Then

$$\sum_{i=0}^{n+2} i = (n+2) + \sum_{i=0}^{n+1} i \overset{!}{=} (n+2) + \binom{n+2}{2} = \binom{n+2}{1} + \binom{n+2}{2} = \binom{n+3}{2}.$$

Non-Example 4. Here is a "fake proof" by induction that

$$\sum_{i=0}^{n+1} i = \frac{(2n+3)^2}{8} \text{ for all positive integers } n.$$

Let us assume that $\sum_{i=0}^{n+1} i = \frac{(2n+3)^2}{8}$ holds true for some n. Then

$$\sum_{i=0}^{n+2} i = (n+2) + \sum_{i=0}^{n+1} i \overset{!}{=} (n+2) + \frac{(2n+3)^2}{8} = \frac{4n^2+20n+25}{8} = \frac{(2n+5)^2}{8} = \frac{(2(n+2)+1)^2}{8}.$$

What did we do wrong? Induction consists of *two* parts, a step and a basis... The basis is not superfluous! We need a domino tile where our domino effect can start.

Remark 5. A common mistake is to memorize the induction step as follows, "Let us assume that the statement holds for all n; then let us prove it for $n+1$". This makes no sense: If we already know that the statement holds for all n, then we are done!, no need to think about $n+1$. That's not how induction works. What instead we are assuming is much less, namely, that the statement holds for *one* specific n; from there we want to be able to say the same thing about its successor, $n+1$.

We should pay special attention to whether our induction step is imposing extra conditions on n. If the induction step works only for $n \geq n_0$, then the basis for the induction should be its verification at n_0, and not at 0.

Example 6. Let us prove that $n^2 - 5n + 6 \geq 0$ for all integers n. Let's assume it for n; then

$$(n+1)^2 - 5(n+1) + 6 = (n^2 + 2n + 1) - 5n - 5 + 6 = (n^2 - 5n + 6) + 2n - 4 \geq 2n - 4.$$

Now to conclude we would like to say that $2n - 4 \geq 0$. But this is true only when $n \geq 2$. So we are not done yet; it is incorrect for us to "make the domino tiling start" at $n = 0$ because as far as we know, the validity at 0 might not imply the validity at 1. So we proceed as follows:

- First, we ask ourselves whether the claim holds true for $n = 2$, which is the correct induction basis. Since $2^2 - 10 + 6 = 0$, the answer is "yes". Together with the induction step, this does prove

  $$\text{"}n^2 - 5n + 6 \geq 0 \text{ for all integers } n \geq 2\text{."}$$

- This is not what we were asked to do, but almost: we are left with only finitely many cases to consider!, namely, $n = 0$ and $n = 1$. We can check them by hand: for $n = 0$ we have $0 - 0 + 6 > 0$, for $n = 1$ we have $1^2 - 5 + 6 = 0$. This concludes the proof.

Non-Example 7. Here is a "fake proof" (by induction on the number n of students) that

"all students will receive the same grade in the final".

For $n = 1$, we are considering a class consisting of only one student, so the claim is clear. Now suppose we have proved the claim for classes with n students. Let C be any class with $n + 1$ students. Let x, y be any two students enrolled in the class, and let z be any other student. Consider $S \setminus \{x\}$: this is a set of n students, so we can apply the inductive assumption: Everybody in $S \setminus \{x\}$ is going to get the same grade. In particular, y and z will get the same grade. Analogously, by the inductive assumption everybody in $S \setminus \{y\}$ will get the same grade, so in particular x and z. Summing up, x, y, z will all get the same grade. But then *any* two students x, y will get the same grade. What is wrong here?

(Hint: If in a proof you pick three different elements from a set, then you are implicitly assuming that the set has at least three elements. Note that if we have a class of 3 students, and any pair of them gets the same grade, then indeed they all get the same grade...)

There is a variant of induction which sometimes is easier to use than the one above. It is called **strong induction**, or sometimes "complete' or "generalized induction". Essentially, it is just normal induction plus "bookkeeping", i.e. keeping track of everything you have proven before. It consists of two parts:

- ("Basis") Prove a statement for a specific integer n_0.
- ("Step") Prove that, if there is a natural number n such that the statement holds for *every* natural number k such that $n_0 \leq k \leq n$, then the statement must hold also for $n + 1$.

Let us use the shortening $P(n)$for "the property holds for n". It is easy to see how generalized induction works: First of all, $P(n_0)$ implies $P(n_0 + 1)$. Now that we know $P(n_0)$ and $P(n_0 + 1)$, we can infer $P(n_0 + 2)$. But then we know $P(n_0)$ $P(n_0 + 1)$, and $P(n_0 + 2)$, which together imply $P(n_0 + 3)$. And so on: Another domino effect, which results in a proof of the statement for all integers $n \geq n_0$. The difference with classical induction is that instead of proving

$$P(n) \Rightarrow P(n + 1),$$

where $P(n)$ stands for "the property holds for n", we keep track at each step of what we have proven already and show

$$[P(n) \text{ and } P(n-1) \text{ and } P(n-2) \ldots \text{ and } P_{n_0}] \Rightarrow P(n+1).$$

Proposition 8. *Any integer $n \geq 2$ can be written as product of primes.*

Proof. The statement is true for $n = 2$, because "$2 = 2$" writes 2 as products of primes. Now let n be an integer, and suppose the claim has already been proven for any integer in $\{2, 3, \ldots, n-1\}$. If n is prime, then

$$n = n$$

is a valid way to write down n as a product of primes, and we are done. If n is composite, then

$$n = n_1 \cdot n_2,$$

with $2 \leq n_i < n$ (for $i = 1, 2$). By strong induction, both n_i can be written as product of primes. But then so can n. □

Corollary 9 (Euclid)**.** *There is no largest prime.*

Proof. Let $p_1, \ldots, p_r$ be the complete list of the first r primes. Set $n \stackrel{\text{def}}{=} 1 + p_1 p_2 \cdots p_r$. Let p be any prime factor of n; the existence of p is guaranteed by Proposition 8. Were p belonging to $\{p_1, \ldots, p_r\}$, then

$$1 \;=\; p_1 \cdots p_r - n$$

would be a difference of two multiples of p, so 1 itself would be a multiple of p, a contradiction. Thus p does not belong to $\{p_1, \ldots, p_r\}$. In particular, p is a prime larger than all of $p_1, \ldots, p_r$. This proves that there is no largest prime. □

Remark 10. In some textbooks, the theorem above is often misquoted as follows: *given the set of the first k primes, their product plus one is a much larger prime.* This is a wrong argument:

$$1 + (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13) = 30031$$

is not prime, for example, because it is divisible by 59.

Remark 11. The argument of Corollary 9 can be adapted to prove a quantitative statement: Namely, that *the number of primes not larger than n is at least* $\log_2 \log_2 n$. In fact, with modern techniques we can estimate the growth of prime numbers better: in 1896, Hadamard and de-la-Vallee-Poussin independently proved that the number of primes $\leq n$ grows like $\frac{n}{\log_2 n}$; an elementary proof was fund in 1948 by Selberg and Erdös, again independently. This statement goes under the name of **PNT** (Prime Number Theorem).

0.2 Euclidean division, unique factorization, and the Euclidean algorithm

Theorem 12 (Euclidean division)**.** *Let n, d be natural numbers, $d \neq 0$. There exist natural numbers (q, r) such that*

(1) $n = qd + r$,

(2) $0 \leq r < d$.

In addition, the pair (q, r) is uniquely determined by (n, d).

Notation. The number q is called "quotient of the division of n by d"; the number r is called "remainder of the division of n by d". Sometimes d is called "divisor". This explains why one chooses the letters q, r, d.

Proof. Let us prove existence first. The claim is clear for $n < d$ (by choosing $q \stackrel{\text{def}}{=} 0$ and $r \stackrel{\text{def}}{=} n$). The claim is also clear for $d = 1$ (by choosing $q \stackrel{\text{def}}{=} n$ and $r \stackrel{\text{def}}{=} 0$). So the interesting case is

$$n \geq d \geq 2.$$

We proceed by strong induction on the first component of the pair (n, d) (the basis case $n = 0$ being already covered by the $n < d$ case of the discussion above). Consider $n_1 = n - d$. Clearly $n_1 \in \mathbb{N}$ (we are in the case $n \geq d$) and $n_1 < n$. By strong induction, the existence part of the theorem holds for the pair (n_1, d): So we can find natural numbers q_1, r_1 such that

$$n_1 = q_1 d + r_1 \text{ and } 0 \leq r_1 < d.$$

But then

$$n = n_1 + d = (q_1 + 1)d + r_1, \text{ with } 0 \leq r_1 < d$$

is the desired "division of n by d".

As for uniqueness, say $n = qd + r = q'd + r'$, with $0 \leq r < d$ and $0 \leq r' < d$. Now:

- if $q = q'$, then $r = n - qd = n - q'd = r'$, so $(q, r) = (q', r')$ and we are done.
- if $q > q'$, then $q \geq q' + 1$. Multiplying by d we get $qd \geq q'd + d$. Hence

  $$q'd + r' = n = qd + r \geq q'd + d + r,$$

  whence $r' \geq d + r$, a contradiction because $r \geq 0$ and $r' < d$.
- symmetrically, if $q' > q$ one gets $r \geq r' + d$, contradicting $r' \geq 0$ and $r < d$. □

Here is a famous result by Euclid, with a proof that uses strong induction three times. We will see a much simpler proof later.

Lemma 13 (Euclid)**.** *Let a and b be natural numbers. If a prime number p divides ab, it divides either a or b.*

Proof. [4] We proceed by strong induction on the minimum of the pair $\{a, b\}$. Up to relabeling, we can assume $a \leq b$, so that a is the smallest of the pair. If $a = 0$, or $a = 1$, then the claim is clear. So we assume $a \geq 2$ and distinguish two cases: either a is prime, or not.

- Suppose a is prime. Let p be a prime that does not divide a but divides ab for some b. Via Theorem 12, write $p = aq + r$ with $0 \leq r < a$. Since $ab = pc$ for some integer c, we have that

  $$ab = pc = (aq + r)c = acq + rc.$$

  This implies that $rc = a(b - cq)$, so the prime a divides rc. Since $r < a$, by strong induction the theorem holds for the pair $\{r, c\}$; that is, when a prime divides rc, it divides either r or c. But the prime a divides rc and does not divide r, because $r < a$. Hence, a divides c. Writing $c = ad$ for some integer d, we get

  $$ab = pc = pad,$$

  and canceling a we get $b = pd$. So p divides b.
- Suppose a is not prime. Then $a = d_1 \cdot d_2$, with both $d_1, d_2 < a$. Let p be a prime that does not divide a, and divides ab for some b. Then p divides neither d_1 nor d_2 (or else it would divide a), but

  $$p \text{ divides } d_1 d_2 \, b.$$

[4] Proof due to Barry Cipra, math.stackexchange.com/questions/1581173/proof-of-euclids-lemma, 2015

By strong induction (since $d_1 < a$) the statement of the theorem holds for the pair $\{d_1, d_2 b\}$: so either p divides d_1 (which is false), or p divides $d_2 b$. Hence,

$$p \text{ divides } d_2\, b.$$

By strong induction (since $d_2 < a$) the statement of the theorem holds for the pair $\{d_2, b\}$: so either p divides d_2 (which is false), or p divides $d_2 b$. Hence, p divides b. □

Lemma 14. *Let $a_1, \ldots, a_n$ be natural numbers. If a prime number p divides their product, then p divides (at least) one of $\{a_1, \ldots, a_n\}$.*

Proof. The case $n = 2$ is Lemma 13. By induction, suppose p divides $a_1 \cdot \ldots \cdot a_n \cdot a_{n+1}$. Call $b \stackrel{\text{def}}{=} a_1 \cdot \ldots \cdot a_n$. Since p divides $b \cdot a_{n+1}$, by Lemma 13 either p divides a_{n+1}, or p divides b, in which case by inductive assumption p divides one of $\{a_1, \ldots, a_n\}$. □

Theorem 15 (Unique Factorization)**.** *Any integer $n \geq 2$ can be decomposed as product of weakly-increasing primes, and such decomposition is unique.*

Proof. We already know that n can be written as product of primes by Proposition 8; so up to reordering them in weakly-increasing order, the existence of such decomposition is clear. The hard part is to prove uniqueness. Suppose

$$p_1 \cdot p_2 \cdot \ \ldots \ \cdot p_r = n = q_1 \cdot q_2 \cdot \ \ldots \ \cdot q_s,$$

with p_i, q_j primes, listed so that

$$p_1 \leq p_2 \leq \ \ldots \ \leq p_r \quad \text{and} \quad q_1 \leq q_2 \leq \ \ldots \ \leq q_s.$$

Since p_1 divides the product of the q_j's, by Lemma 14 it must divide at least one of the q_j's. Because they are both primes, this actually means that p_1 is *equal* to one of the q_j's. Since q_1 is the smallest of the q_j's, this means that

$$p_1 \geq q_1.$$

Symmetrically, q_1 divides the product of the p_i's, so it must divide one of them by Lemma 14. By primality, q_1 is equal to one of the p_i's, so in particular $p_1 \leq q_1$. Thus $p_1 = q_1$. But then we can cancel p_1 and q_1, and proceed recursively. Because

$$p_2 \cdot \ \ldots \ \cdot p_r = q_2 \cdot \ \ldots \ \cdot q_s,$$

we get that $p_2 = q_2$; and so on. It follows that $r = s$ and $p_i = q_i$ for each i. □

The unique factorization (Theorem 15), also known as "fundamental theorem of arithmetics", provides a way to find the gcd of two natural numbers a and b: We can decompose a and b into primes, and then collect together all common factors.

Corollary 16. *Suppose $a = p_1^{a_1} \cdot \ \ldots \ \cdot p_h^{a_h} \cdot p_{h+1}^{a_{h+1}} \cdot \ldots \cdot p_r^{a_r}$ and $b = p_1^{b_1} \cdot \ \ldots \ \cdot p_h^{b_h} \cdot q_{h+1}^{b_{h+1}} \cdot \ldots \cdot q_m^{b_m}$ are decompositions into distinct primes, so that $\{p_{h+1}, \ldots, p_r\} \cap \{q_{h+1}, \ldots, q_m\} = \emptyset$. (Here each a_i, b_i and c_i is a positive integer.) Then,*

$$\gcd(a, b) = p_1^{\min(a_1, b_1)} \cdot \ p_2^{\min(a_2, b_2)} \ \ldots \ \cdot p_h^{\min(a_h, b_h)}.$$

Example 17. Since $168 = 2^3 \cdot 3 \cdot 7$ and $60 = 2^2 \cdot 3 \cdot 5$, then

$$\gcd(168, 60) = 2^2 \cdot 3 = 12.$$

This method seems quick, but it isn't. The problem is that we have hidden the difficulty under the carpet: Given 168, how can you find its prime factors quickly? In general, factoring requires a huge amount of computational time. Many cryptography systems (e.g. RSA) that keep your emails and credit cards secure, are based on the fact that a product of two distinct primes uniquely determines the two primes, but it takes really long to figure them out from the product if you do not have extra information. It turns out however that we can write down an algorithm to find the gcd that is much quicker than factoring.

Lemma 18. *Let n, m be positive integers. Let $n = qm + r$, with $0 \leq r < m$. Then*

$$\gcd(n, m) = \gcd(m, r).$$

Proof. Set $d_1 \stackrel{\text{def}}{=} \gcd(n, m)$ and $d_2 \stackrel{\text{def}}{=} \gcd(m, r)$. Since d_1 divides both n and m, it divides also $r = n - qm$; so it's a common divisor of m and r. So $d_1 \leq d_2$. On the other hand, d_2 divides m and r, so it divides also $n = qm + r$. So it's a common divisor of n and m. So $d_2 \leq d_1$. □

Algorithm 19 (Algorithm to compute the GCD)**.** INPUT: a, b positive integers, with $b < a$.

def $\gcd(a, b)$:
    Do the Euclidean division $a = qb + r$.
    if $r = 0$:
        return b.
    else:
        return $\gcd(b, r)$.

The algorithm is recursive. Termination is ensured by the fact that at each iteration, the remainder decreases. Eventually, it will get to zero: but if the remainder of the division of x by y is zero, it means that $\gcd(x, y) = y$.

Example 20. Let us compute $\gcd(168, 60)$ using the Euclidean algorithm.

$$\begin{cases} 168 &= 2 \cdot 60 + 48 \\ 60 &= 1 \cdot 48 + 12 \\ 48 &= 4 \cdot 12 + 0. \end{cases}$$

So the $\gcd(168, 60) = \gcd(60, 48) = \gcd(48, 12) = 12$. Note that we did not have to factor 168.

0.3 Modular arithmetics

Definition 21. Let $m \geq 2$ be an integer. Let

$$\mathbb{Z}_m \stackrel{\text{def}}{=} \{0, 1, \ldots, m - 1\}.$$

Modular addition and *modular multiplication* are the two operations on $\mathbb{Z}_m$ defined as follows:

$$\begin{aligned} a \oplus b &\stackrel{\text{def}}{=} \text{the remainder of the division of } a + b \text{ by } m \\ a \odot b &\stackrel{\text{def}}{=} \text{the remainder of the division of } ab \text{ by } m. \end{aligned}$$

Modular operations behave in a very similar manner to usual arithmetic operations; later in the course, once we introduce quotients rings, we will understand why.

Lemma 22. *Both $\oplus$ and $\odot$ are associative: That is, for each a, b, c,*

$$(a \oplus b) \oplus c = a \oplus (b \oplus c) \quad \text{and } (a \odot b) \odot c = a \odot (b \odot c).$$

Proof. We prove it only for $\oplus$; the $\odot$ case is analogous and left as exercise. By definition, $a \oplus b$ is the remainder r_1 of the Euclidean division

$$(a+b) = q_1 m + r_1, \text{ with } 0 \leq r_1 < m.$$

So $(a \oplus b) \oplus c$ is the remainder r_2 of the Euclidean division

$$r_1 + c = q_2 m + r_2, \text{ with } 0 \leq r_2 < m.$$

So $r_2 = r_1 + c - q_2 m = (a + b - q_1 m) + c - q_2 m = (a + b + c) - (q_1 + q_2)m$; hence,

$$(a + b + c) = (q_1 + q_2)m + r_2,$$

and since $r_2 < m$, the expression above is a Euclidean division. In other words, r_2 is the remainder of the division of $a + b + c$ by $(q_1 + q_2)$.

On the other hand, $b \oplus c$ is by definition the remainder r_3 of the Euclidean division

$$(b + c) = q_3 m + r_3, \text{ with } 0 \leq r_3 < m.$$

So $a \oplus (b \oplus c)$ is the remainder r_4 of the Euclidean division

$$a + r_3 = q_4 m + r_4, \text{ with } 0 \leq r_4 < m.$$

But then $r_4 = a + r_3 - q_4 m = a + (b + c - q_3 m) - q_4 m = (a + b + c) - (q_3 + q_4)m$; and as above, we get that

$$(a + b + c) = (q_3 + q_4)m + r_4, \text{ with } 0 \leq r_4 < m$$

is also a Euclidean division. By the uniqueness of the Euclidean division, it follows that $r_2 = r_4$, so $(a \oplus b) \oplus c = a \oplus (b \oplus c)$. □

Lemma 23. *The operation $\odot$ distributes $\oplus$: That is, for each a, b, c,*

$$(a \oplus b) \odot c = (a \odot c) \oplus (b \odot c).$$

Proof. As above, $a \oplus b \stackrel{\text{def}}{=} r_1$, where

$$(a + b) = q_1 m + r_1, \text{ with } 0 \leq r_1 < m.$$

In turn, $(a \oplus b) \odot c \stackrel{\text{def}}{=} r_2$, where

$$r_1 c = q_2 m + r_2, \text{ with } 0 \leq r_2 < m.$$

In particular, $r_2 = r_1 c - q_2 m = (a + b - q_1 m)c - q_2 m = (ac + bc) - (cq_1 + q_2)m$. Since $r_2 < m$, we see that r_2 is the remainder of the Euclidean division of $(ac + bc)$ by m.

Now let $r_3 = a \odot c$ and $r_4 = b \odot c$ be the remainders of the Euclidean division by m of ac and bc, respectively. Clearly, $r_3 + r_4$ will be the remainder of the Euclidean division by m of $(ac + bc)$. But then $r_3 + r_4 = r_2$. □

Remark 24. An important difference between modular arithmetics and usual arithmetics, is that in modular arithmetics sums and products of nonzero integers can be zero. For example, in $\mathbb{Z}_6$ we have

$$\begin{aligned} 1 \oplus 5 &= 0. \\ 2 \odot 3 &= 0. \end{aligned}$$

Proposition 25. *Let $m \geq 2$ be a natural number. The set $\mathbb{Z}_m$ with the operation of* ***modular addition*** *satisfies the following four properties:*
(a) Closure. *If a, b are in $\mathbb{Z}_m$, so is $a \oplus b$.*
(b) Associativity. *If a, b, c are in $\mathbb{Z}_m$, $a \oplus (b \oplus c) = (a \oplus b) \oplus c$.*
(c) Identity. *There is a unique element 0 in $\mathbb{Z}_m$ such that for every a in $\mathbb{Z}_m$, $a \oplus 0 = a = 0 \oplus a$.*
(d) Inverses. *For every a in $\mathbb{Z}_m$, there exists a unique b in $\mathbb{Z}_m$ with the property that $a \oplus b = 0 = b \oplus a$.*

A natural question is: Which elements of $\mathbb{Z}_m$ are "invertible with respect to modular multiplication"? For example, in $\mathbb{Z}_{15}$ the equation

$$2 \odot x = 1$$

has a solution, namely, $x = 8$; whereas it is easy to see (by checking all fifteen cases) that

$$3 \odot x = 1$$

has no solution. Going further,

$$4 \odot x = 1$$

has a solution, namely, $x = 8$; whereas

$$5 \odot x = 1$$

has again no solution. How does it work?

Proposition 26. *Let $m \geq 2$. Let $a \in \mathbb{Z}_m$.*
The equation $a \odot x = 1$ has solutions $x \in \mathbb{Z}_m$ if and only if $\gcd(x, m) = 1$.

Proof. Both implications follow from Bezout's theorem. In fact:

- Suppose $a \odot x = 1$ has a solution $x \in \mathbb{Z}_m$. This means that in $\mathbb{Z}$, the Euclidean division of ax by m has remainder 1. Thus there is a natural number q such that

  $$ax = qm + 1.$$

  In particular, the equation $ax - my = 1$ has solutions in $\mathbb{N}$, so by Bezout's theorem $\gcd(a, m) = 1$.
- Conversely, suppose $\gcd(a, m) = 1$. Then by Bezout's theorem $ax - my = 1$ has integer solutions. Since $ax \geq 1$ and $m > 1$, from $ax - 1 = my$ we see that the 'integer solution' y cannot be negative, so it is a natural number. Thus $ax = my + 1$ is an identity in $\mathbb{N}$. Since $0 < 1 < 2 \leq m$, such identity coincides with the Euclidean division of ax by m. Thus $a \odot x = 1$ in $\mathbb{Z}_m$. □

Recall that two natural numbers u, v are called *coprime* if $\gcd(u, v) = 1$.

Proposition 27. *Let $m \geq 2$ be a natural number. The set U_m of numbers in $\mathbb{Z}_m$ coprime with m, with the operation of* ***modular multiplication****, satisfies the following four properties:*
(a) Closure. *If a, b are in U_m, so is $a \odot b$.*
(b) Associativity. *If a, b, c are in U_m, $a \oplus (b \oplus c) = (a \oplus b) \oplus c$.*
(c) Identity. *There is a unique element 1 in U_m such that for every a in U_m, $a \odot 1 = a = 1 \odot a$.*
(d) Inverses. *For every a in U_m, there exists a unique b in U_m with the property that $a \odot b = 1 = b \odot a$.*

Proof. Placing some factorization of a next to some factorization of b yields some factorization of ab. By the Unique factorization theorem, in the previous sentence we may replace each "some" by "the". Thus if a and m have no common prime factors, and b and m have no common prime factors, then ab and m have no common prime factor. This shows the first property. The second property is true for all a, b, c in $\mathbb{Z}_m$ (whether coprime with m or not), by Lemma 22. For the third property, existence is obvious. To show uniqueness, suppose there is a z such that for all elements a of U_m, $a \odot z = a = z \odot a$; then in particular $1 \odot z = 1$, but at the same time $1 \odot z = z$; hence $1 = z$. As for the last property, existence is by Proposition 26. To show uniqueness, suppose there is a w such that $a \odot w = 1 = w \odot a$. Then

$$w = w \odot 1 = w \odot (a \odot b) = (w \odot a) \odot b = 1 \odot b = b. \qquad \square$$

0.4 Functions and Quotients. The sets $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{Z}_m$ as quotients

Let X, Y be any two sets. Recall that their Cartesian product is defined by

$$X \times Y \stackrel{\text{def}}{=} \{(x, y) \text{ such that } x \in X, y \in Y\}.$$

Definition 28. A *function* $f : X \longrightarrow Y$ consists of two sets X, Y and a subset $F \subseteq X \times Y$, such that for each element $x \in X$ there is always exactly one element y of Y for which $(x, y) \in F$. Usually we denote this y by $f(x)$, and we say it is the *image of* x *(under* f*)*. We also call X (resp. Y) the *domain* (resp. the *codomain*) of the function. The *image of the set* X is the set $\operatorname{Im} X \stackrel{\text{def}}{=} \{f(x) \text{ such that } x \in X\}$.

Functions are often described by a formula that tells us how to find $f(x)$ given x. For example: Given any set X, the *identity function on* X (denoted by id_X) is the function whose output is always identical to the input. In this case, our notation to describe the function is

$$\begin{aligned} id_X : X &\longrightarrow X \\ x &\longmapsto x. \end{aligned}$$

Sometimes one does not have an explicit formula available, but one can still give a method to associate x with its image: for example,

$$\begin{aligned} f : \mathbb{N} &\longrightarrow \mathbb{N} \\ x &\longmapsto \text{the } x\text{-th prime number}. \end{aligned}$$

If no general pattern is available, we can always express f by specifying all its values:

$$\begin{aligned} f : \{0, 1, 2\} &\longrightarrow \{0, 1, 2\} \\ 0 &\longmapsto 1 \\ 1 &\longmapsto 0 \\ 2 &\longmapsto 2. \end{aligned}$$

Definition 29. A function $f : X \to Y$ is *injective* if for each $x \neq x'$ one has $f(x) \neq f(x')$.

We assume familiarity with logic and quantifiers ($\forall$, $\exists$) and logical equivalence (contrapositives, etc.) For example, it should be clear that an equivalent way to define injectivity is:

$$\forall x, x' \in X, \quad f(x) = f(x') \Rightarrow x = x'.$$

Injectivity depends not only on the "formula", but also on the domain involved. For example, the function "first letter of" is injective on the set { Alba, Bruno }, but not injective on the set { Alba, Alice, Bruno }.

Definition 30. A function $f : X \to Y$ is *surjective* if the image of X coincides with the codomain Y; or in other words, if for each $y \in Y$ there is some $x \in X$ (not necessarily unique) such that $y = f(x)$.

Surjectivity depends not only on the "formula" for f, but also on the domain and the codomain. For example, let E be the set of even natural numbers:

$$\begin{array}{rcl} f : \mathbb{N} & \longrightarrow & \mathbb{N} \\ x & \longmapsto & 2x \end{array} \text{ is not surjective,} \qquad \begin{array}{rcl} f : \mathbb{N} & \longrightarrow & E \\ x & \longmapsto & 2x \end{array} \text{ is,} \qquad \begin{array}{rcl} f : \mathbb{N} \setminus \{0\} & \longrightarrow & E \\ x & \longmapsto & 2x \end{array} \text{ is not.}$$

Definition 31. A function $f : X \to Y$ is *bijective* if it is both injective and surjective. That is, if for each $y \in Y$ there exists exactly one $x \in X$ such that $y = f(x)$.

Given a function $f : X \to Y$ and a function $g : Y \to Z$, their *composite* is the function

$$\begin{array}{rcl} g \circ f : \quad X & \longrightarrow & Z \\ x & \longmapsto & g(\, f(x)\,). \end{array}$$

Proposition 32. *Let $f : X \to Y$ be a function between two non-empty sets.*
(1) *f is surjective $\iff$ there exists $g : Y \to X$ (called "right inverse") such that $f \circ g = id_Y$.*
(2) *f is injective $\iff$ there exists $g : Y \to X$ (called "left inverse") such that $g \circ f = id_X$.*
(3) *f is bijective $\iff$ there exists $g : Y \to X$ (called "inverse") such that $g \circ f = id_X$ and $f \circ g = id_Y$.*

Remark 33. Before starting with the proof, note that two functions are equal when they have same domain, same codomain, and they yield same outputs when given same inputs. So to verify an equality of functions like $g \circ f = id_Y$, both going from Y to Y, we'll need to check that $g \circ f(y) = id_Y(y)$ for all $y \in Y$.

Proof of Proposition 32.
(1), "⇒". Define

$$\begin{array}{rcl} g : Y & \longrightarrow & X \\ y & \longmapsto & \text{some } x \text{ such that } f(x) = y. \end{array}$$

(If there is more than one x such that $f(x) = y$, we simply choose one.) Then by construction, $f \circ g(y) = f(x) = y$ for all $y \in Y$. Hence $f \circ g = id_Y$.
(1), "⇐". For each $y \in Y$, we know that $id_Y(y) = f \circ g(y)$, so $y = f(g(y))$, which means $y \in \operatorname{Im} f$.
(2), "⇒". Choose a point x_0 of X. Define

$$\begin{array}{rcl} g : Y & \longrightarrow & X \\ y & \longmapsto & \begin{cases} x_0, & \text{if } y \notin \operatorname{Im} f \\ \text{the unique } x \text{ such that } f(x) = y, & \text{if } y \in \operatorname{Im} f. \end{cases} \end{array}$$

Then for all x in X, $g \circ f(x) = g(f(x)) = x$. So $g \circ f = id_X$.
(2), "⇐". Suppose $f(x) = f(x')$. Applying g, and remembering that $g \circ f = id_X$, we get

$$x = id_X(x) = g \circ f(x) = g(f(x)) = g(f(x')) = g \circ f(x') = id_X(x') = x'.$$

(3), "⇒". This does not follow immediately from items (1) and (2), because a priori it could be that the two g's (right inverse and left inverse) are different. However, if f is bijective we can simply define

$$\begin{array}{rcl} g : Y & \longrightarrow & X \\ y & \longmapsto & \text{the unique } x \text{ such that } f(x) = y. \end{array}$$

and it is easy to see that it does the trick.
(3), "⇐". This follows from (1) and (2). (Why?) □

Relations and quotients

Definition 34. Let X be an arbitrary, non-empty set. An *equivalence relation on* X is a subset R of $X \times X$ that satisfies the following properties:
REL1: $(x, x) \in R$ for all x. ("reflexivity")
REL2: If $(x, y) \in R$, then $(y, x) \in R$. ("symmetry")
REL2: If $(x, y) \in R$ and $(y, z) \in R$, then $(x, z) \in R$. ("transitivity")

Definition 35. Let R be equivalence relation on a set X, instead of $(x, y) \in R$ we shall write $x \sim y$, and read it "x is in a relation with y". The *equivalence class* of an element x of X is

$$\overline{x} \stackrel{\text{def}}{=} \{y \in X \text{ such that } y \sim x\} \stackrel{\text{def}}{=} \{y \in X \text{ such that } (x, y) \in R\}.$$

Example 36. On any non-empty set X, one can always put two "extreme" equivalence relation: the first one is

$$R_0 = \{(x, y) \text{ such that } x = y\}.$$

Under this, any element of X is in a relation only with himself. So the equivalence classes are as small as possible: They contain one element each.

The other extreme is

$$R_1 = X \times X.$$

Under this, any element of X is in a relation with everyone! So there is only one giant equivalence class containing all elements.

Example 37. Let X be the set of students in your Algebra class. If we define

$$R = \{(x, y) \text{ such that } x, y \text{ are born in the same year}\}$$

this is an equivalence relation. You are in a relation with anybody who is born the same year as you. If you view the student names as files, you can think of the equivalence classes as folders, labeled by birthyear.

Example 38. Let $\mathbb{N}$ be the set of integers. let us define

$$R = \{(a, b) \text{ such that } a - b \text{ is even}\}.$$

This is an equivalence relation, because (REL1) $a - a$ is even, (REL2) if $a - b$ is even so is $b - a$, and (REL3) if $a - b$ and $b - c$ are even, so is their sum $a - c$. This equivalence relation is called *congruence mod 2*.

Non-Example 39. The empty relation $R = \emptyset$ is not an equivalence relation: It satisfies (REL2) and (REL3), but not (REL1).

Non-Example 40. Let $X = \mathbb{N}$. The relation

$$R = \{(a, b) \text{ such that } |a - b| < 5\}$$

is not an equivalence relation: It satisfies (REL1) and (REL2), but not (REL3).

For a real-life analogy, "being close to" is not a relation of equivalence: if it takes less than 5 minutes to go from a to b, and it takes also less than 5 minutes to go from b to c, not necessarily it takes less than 5 minutes to go from a to c! It could be that distances add up.

Non-Example 41. Let $X = \mathbb{N}$. The relation

$$R = \{(a, b) \text{ such that } a \leq b\}$$

is not an equivalence relation: It satisfies (REL1) and (REL3), but not (REL2).

Definition 42. Let R be an equivalence relation on a set X. The *quotient* ${}^{X}\!/_{\sim}$ is the set of all equivalence classes in X. In other words,

$${}^{X}\!/_{\sim} = \{\overline{x} \text{ such that } x \in X\}.$$

By definition, two elements of X are equal in the quotient (i.e. $\overline{x} = \overline{x'}$ in ${}^{X}\!/_{\sim}$) if and only if they are in a relation with one another (i.e. $x \sim x'$).

Very often in mathematics, we have to define functions or operations on quotients. Here is a general trick for that:

1. To define a function $f : {}^{X}\!/_{\sim} \to Y$, we can simply define a function $F : X \to Y$, and then to check "compatibility with the quotient", i.e. check that

$$x \sim x' \Longrightarrow F(x) = F(x').$$

2. To define an internal operation

$$o : {}^{X}\!/_{\sim} \times {}^{X}\!/_{\sim} \longrightarrow {}^{X}\!/_{\sim}$$

we may simply define an operation

$$O : X \times X \longrightarrow X,$$

and then check that O is compatible with the quotient, i.e.

$$x \sim x', y \sim y' \Longrightarrow O(x, y) \sim O(x', y').$$

The set $\mathbb{Z}$ as quotient and Bezout's theorem

Example 43 ($\mathbb{Z}$ as quotient)**.** The set $\mathbb{Z}$ of integers can be defined as follows: on $X \stackrel{\text{def}}{=} \mathbb{N} \times \mathbb{N}$ we introduce the equivalence relation

$$(a, b) \sim (a', b') \stackrel{\text{def}}{\Longleftrightarrow} a + b' = a' + b.$$

Then $\mathbb{Z} \stackrel{\text{def}}{=} {}^{X}\!/_{\sim}$. By convention, we denote $\overline{(a, 0)}$ simply by 'a' and $\overline{(0, a)}$ simply by '$-a$'.

The hint is: you should think of $\overline{(a, b)}$ as what you have always written as '$a - b$'. On the set $\mathbb{Z}$, we can define addition componentwise

$$\overline{(a, b)} + \overline{(c, d)} = \overline{(a + c, b + d)}$$

and multiplication as

$$\overline{(a, b)} \cdot \overline{(c, d)} = \overline{(ac + bd, ad + bc)}.$$

Are these operations legitimate? if $X = \mathbb{N} \times \mathbb{N}$, it is easy to see that the componentwise addition from $X \times X$ to X is compatible with the quotient: if $(a, b) \sim (a', b')$ and $(c, d) \sim (c', d')$, which means that $a + b' = a' + b$ and $c + d' = c' + d$, then $a + c + b' + d' = a' + c' + b + d$, which means that $(a + c, b + d) \sim (a' + c', b' + d')$. But what about multiplication?, if $(a, b) \sim (a', b')$

and $(c,d) \sim (c',d')$, is it true that $(ac+bd, ad+bc) \sim (a'c'+b'd', a'd'+b'c')$? To check this, we assume $a+b' = a'+b$ and $c+d' = c'+d$, and we want to show

$$ac+bd+a'd'+b'c' = ad+bc+a'c'+b'd'.$$

The trick is to add to both sides of the equality above the quantity $ac'+bd'+a'd+b'c$, and prove that on both sides you get as result $2(a'+b)(c'+d)$. We leave this as exercise.

Now that we have $\mathbb{Z}$ available, we can extend some of the statements we had for $\mathbb{N}$ to $\mathbb{Z}$, and simplify some of the theorems:

Definition 44. Given a and b in $\mathbb{Z}$, we say that "*a divides b*" (or equivalently that "*a is a divisor of b*", or equivalently that "*b is a multiple of a*") if there exists an integer k such that

$$b = k \cdot a.$$

The integers 1 and -1 are called *units*.

Definition 45. The *absolute value* $|n|$ of an integer n is defined to be n itself if $n \geq 0$, and $-n$ if $n < 0$. In other words, going back to the original definition of $\mathbb{Z}$ as quotient, we define

$$|\overline{(a,b)}| \stackrel{\text{def}}{=} \max\{a-b, b-a\}.$$

Theorem 46 (Euclidean division)**.** *For any pair (z,d) of integers, $d \neq 0$, there exists a unique pair of integers (q,r) such that $z = qd+r$ and $0 \leq r < |d|$.*

Proof. The claim is clear if z is a multiple of d, in which case $r = 0$. So without loss we may assume that z is *not* a multiple of d. We distinguish four cases:

- If $z \geq 0$ and $d \geq 0$, Theorem 12 allows us to conclude.
- If $z \geq 0$ and $d < 0$, we know how to divide z by $-d$, so $z = q(-d)+r$ with $q \in \mathbb{N}$ and $0 < r < |d|$. But then $z = (-q)d+r$ is the desired division.
- If $z < 0$ and $d \geq 0$, we know how to divide $-z$ by d, so $-z = qd+r$ with $q \in \mathbb{N}$ and $0 < r < |d|$. But then
$$z = -qd-r = (-q-1)d+(d-r)$$
is the desired division, with remainder $0 < d-r < |d|$.
- If $z < 0$ and $d < 0$, we know how to divide $-z$ by $-d$, so $-z = q(-d)+r$ with $q \in \mathbb{N}$ and $0 < r < |d|$. But then
$$z = qd-r = (q+1)d+(-d-r) = (q+1)d+(|d|-r)$$
is the desired division, with remainder $0 < |d|-r < |d|$. □

Theorem 47 (Unique Factorization for Integers)**.** *Any integer different than $0, 1, -1$ can be decomposed as a unit times a product of weakly-increasing positive primes, and such decomposition is unique.*

The next Theorem tells us that $\gcd(a,b)$ is the smallest positive integer that can be written as an "integer linear combination" of a and b.

Theorem 48 (Bezout)**.** *Let a, b be integers, not both zero. Then $\gcd(a,b)$ is the smallest natural number $k \neq 0$ for which the equation $k = ax+by$ has solutions in $\mathbb{Z}$.*

Proof. Set

$$M \stackrel{\text{def}}{=} \{xa + yb \,:\, x, y \in \mathbb{Z}\}.$$

Note that a, b are both in M (by choosing $x = 1$ and $y = 0$, or the other way around). So M contains at least one nonzero integer. Also, if z is in M, so is $-z$, because if $z = xa + yb$ then $-z = (-x)a + (-y)b$. Thus M contains at least one nonzero natural number. Now let d be the **smallest nonzero natural number** in M, and let x_d, y_d be some integers satisfying

$$d = ax_d + by_d. \tag{1}$$

We claim that d *divides any element of* M. In fact, choose an arbitrary $m \in M$ and write it as

$$m = ax_m + by_m. \tag{2}$$

By the Euclidean division of m by d, we have $m = qd + r$, with $0 \leq r < d$. So $r = m - qd$. Plugging in equations (1) and (2) we get

$$r = m - qd = ax_m + by_m - q(ax_d + by_d) = (x_m - qx_d)a + (y_m - qy_d)b.$$

So r is in M. But then r must be 0, or else it would be a nonzero natural number in M smaller than d, contradicting how d was chosen. So m is a multiple of d. This holds for all integers m in M. In particular, since a and b are both in M, d divides both a and b. It remains to show that d is the *greatest* common divisor of a and b. This is easy: Any other natural number c dividing both a and b clearly also divides $ax_d + by_d = d$; which implies $c \leq d$. □

Corollary 49. *The equation* $1 = ax + by$ *has integer solutions if and only if* $\gcd(a, b) = 1$.

Proof. This is the case $d = 1$ of Theorem 48. □

Remark 50. This yields a much shorter proof of Euclid's Lemma 13. In fact, let p be a prime that divides a product ab of two integers. If p does not divide a, then $\gcd(a, p) = 1$, so by Bezout's corollary we can find integers x, y such that $1 = ax + py$. Multiplying this expression by b, we obtain

$$b = (ab)x + bpy,$$

which writes b as the sum of two "integer multiples" of p. So by collecting a factor p from both, we can write b as p times some integer, which for sign reasons cannot be of the form "$(-1)n$". So b is p times 1 times a natural number.

The set $\mathbb{Q}$ as quotient

Example 51 ($\mathbb{Q}$ as quotient)**.** The set $\mathbb{Q}$ can also be defined as follows: on $Y \stackrel{\text{def}}{=} \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ we introduce the equivalence relation

$$(a, b) \sim (a', b') \stackrel{\text{def}}{\iff} ab' = a'b.$$

Then $\mathbb{Q} \stackrel{\text{def}}{=} Y/_{\sim}$. By convention, we denote $\overline{(a, b)}$ by '$\frac{a}{b}$'.

In other words, two fractions $\frac{a}{b}, \frac{a'}{b'}$ are considered identical if $ab' = a'b$. For example, $\frac{1}{2}$, $\frac{-1}{-2}$, and $\frac{3}{6}$ are the same. So if you want each rational number to be represented by precisely one pair (a, b), perhaps you would prefer to write something like

$$\mathbb{Q} = \{0\} \cup \left\{ \frac{a}{b} \text{ such that } a, b \in \mathbb{Z},\ b > 0,\ a \neq 0,\ \text{and } \gcd(a, b) = 1 \right\}.$$

Addition and multiplication are defined by

$$\frac{a}{b} + \frac{c}{d} \stackrel{\text{def}}{=} \frac{ad+cb}{bd} \qquad \text{and} \qquad \frac{a}{b}\,\frac{c}{d} \stackrel{\text{def}}{=} \frac{ac}{bd}.$$

Note that both operations are compatible with the quotient. In other words, if $\frac{a}{b} = \frac{a'}{b'}$ and if $\frac{c}{d} = \frac{c'}{d'}$, then by definition $ab' = a'b$ and $cd' = c'd$. So

$$(ad+cb)b'd' = ab'dd' + cd'bb' = a'bdd' + c'dbb' = (a'd' + b'c')bd \qquad \text{and} \qquad acb'd' = a'c'bd,$$

which imply, respectively, that

$$\frac{a}{b} + \frac{c}{d} \stackrel{\text{def}}{=} \frac{ad+cb}{bd} = \frac{a'd'+c'b'}{b'd'} \stackrel{\text{def}}{=} \frac{a'}{b'} + \frac{c'}{d'} \qquad \text{and} \qquad \frac{a}{b}\,\frac{c}{d} \stackrel{\text{def}}{=} \frac{ac}{bd} = \frac{a'c'}{b'd'} \stackrel{\text{def}}{=} \frac{a'}{b'}\,\frac{c'}{d'}.$$

We all know that $\mathbb{Z}$ can be viewed as a subset of $\mathbb{Q}$, thanks to the identification

$$\begin{aligned} \iota : \quad & \mathbb{Z} \longrightarrow \mathbb{Q} \\ & z \longmapsto \tfrac{z}{1}. \end{aligned}$$

This map ι is injective, but not surjective, of course: elements like $\frac{1}{2}$ are not in the image.

The set $\mathbb{Z}_m$ as quotient

Definition 52 (Congruence mod m)**.** Fix an integer $m \geq 2$. Let a, b be two integers. We say that "a is congruent to b modulo m", and write

$$a \equiv b \mod m,$$

if $a-b$ is a multiple of m. For example, -39 is congruent to 9 modulo 12, because $-39-9 = -48$.

Proposition 53. *Fix an integer $m \geq 2$. For any integer z, there is a unique $x \in \{0, 1, \ldots, m-1\}$ congruent to z modulo m.*

Proof. The claim is clear by Theorem 46: If we write

$$z = qm + r \text{ with } q \in \mathbb{Z} \text{ and } 0 \leq r < m,$$

then z is congruent to the remainder r; and because of uniqueness of Euclidean division, z cannot be congruent to any other x in $\{0, 1, \ldots, m-1\}$. □

Congruence behaves well with respect to products and sums:

Lemma 54. *If $a \equiv a' \mod m$ and $b \equiv b' \mod m$, then*

$$ab \equiv a'b' \mod m \qquad \text{and} \qquad a+b \equiv a'+b' \mod m.$$

Proof. By assumption, there are integers c, d such that $a - a' = cm$ and $b - b' = dm$. Then

$$ab - a'b' = ab - a'b + a'b - a'b' = b(a-a') + a'(b-b') = bcm + a'dm$$

is a multiple of m, so $ab \equiv a'b'$. Similarly,

$$(a+b) - (a'+b') = (a-a') + (b-b') = cm + dm$$

is a multiple of m, so $a + b \equiv a' + b'$. □

Corollary 55. *Congruence modulo m is an equivalence relation on $\mathbb{Z}$. The quotient of $\mathbb{Z}$ is precisely $\mathbb{Z}_m$.*

As an exercise, you may verify that the modular addition and multiplication on $\mathbb{Z}_m$ can be defined simply via the usual addition and multiplication on $\mathbb{Z}$, by checking that the latter operations are compatible with the quotient.

Notation. From now on, when working with $\mathbb{Z}_m$, we will simply write down ab instead of $a \odot b$, and $a + b$ instead of $a \oplus b$.

0.5 Optional reading: From $\mathbb{Q}$ to $\mathbb{R}$ and $\mathbb{C}$: The needs of geometers

Shortly after Euclid, Pythagoras made an amazing discovery: He found out that $\mathbb{Q}$ is not enough to describe elementary geometry. For example, the diagonal of the square whose edge has length 1, cannot be measured exactly within $\mathbb{Q}$.

Lemma 56 (Pythagoras)**.** *For any prime p, the equation $x^2 = p$ has no solutions in $\mathbb{Q}$.*

Proof. By contradiction, suppose we could write uniquely

$$p = \left(\frac{a}{b}\right)^2, \text{ with } a, b \in \mathbb{Z},\ b > 0,\ a \neq 0, \text{ and } \gcd(a, b) = 1.$$

Clearing denominators, $pb^2 = a^2$. So p divides a^2. By Euclid's Lemma 13, this means that p divides a. So write $a = pk$, with k in $\mathbb{N}$. Plugging in, we get

$$pb^2 = (pk)^2 = p^2k^2.$$

Canceling a p, we get $b^2 = pk^2$. But then p divides b^2 and again by Lemma 13, p divides b. Hence p is a common factor of a and b. A contradiction, we assumed $\gcd(a, b) = 1$. □

In fact, a stronger statement is true:

Theorem 57 (Dedekind, 1858)**.** *For any natural number n, if $x^2 = n$ has no solution in $\mathbb{N}$, then it has also no solution in $\mathbb{Q}$.*

Proof. By contradiction, suppose there is a natural number $m \geq 1$ such that $x^2 = m$ has no solutions in $\mathbb{N}$, but some solution $\frac{a}{b}$ in $\mathbb{Q}$. Without loss, we can assume that $a > 0$, $b > 0$, and b is smallest possible (which basically is the same as assuming $\gcd(a, b) = 1$). Let ℓ be the largest natural number such that $\ell^2 \leq m$. Since $x^2 = m$ has no natural solution, we actually have $\ell^2 < m$. Clearly $\ell < \frac{a}{b}$ (because otherwise we would have $\ell \geq \frac{a}{b}$ and passing to the squares, $\ell^2 \geq m$, a contradiction). In other words, $0 < \frac{a}{b} - \ell$. Also $\frac{a}{b} - \ell < 1$ (or else we would have $\frac{a}{b} \geq \ell + 1$ and passing to the squares, $m \geq (\ell + 1)^2$, contradicting the way ℓ was chosen). Hence,

$$0 < \frac{a}{b} - \ell < 1. \tag{3}$$

Now consider

$$b' \stackrel{\text{def}}{=} b\left(\frac{a}{b} - \ell\right).$$

Equation (3) multiplied by b tells us that $0 < b' < b$. On the other hand, being equal to $a - b\ell$, this b' is an integer, and thus a natural number. Finally,

$$b' \cdot \frac{a}{b} = b\left(\frac{a}{b} - \ell\right) \cdot \frac{a}{b} = b\left(\frac{a}{b}\right)^2 - \ell b\frac{a}{b} = bm - \ell a,$$

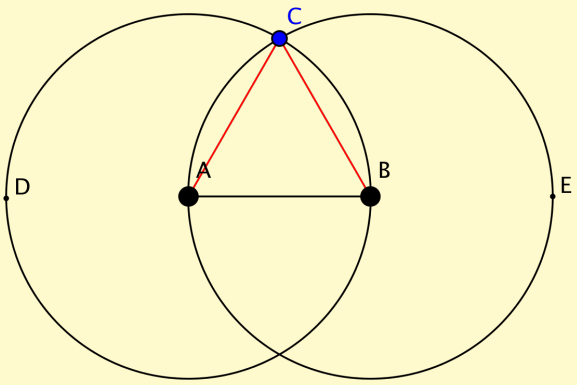
which proves that also $b' \cdot \frac{a}{b}$ is a (positive) integer. Call this integer a'. But then we can write

$$\frac{a}{b} = \frac{a'}{b'}$$

with $0 < b' < b$, a contradiction with the way we chose b. □

Consider now the first theorem of Euclid's *Elements*, perhaps the oldest global treatise of mathematics, published around 300 BC. We have highlighted a sentence in the original proof.

Proposition 1. *How to construct an equilateral triangle on a given segment.*



"It is required to construct an equilateral triangle on the segment AB. Describe the circle BCD with center A and radius AB. Again describe the circle ACE with center B and radius BA. Draw segments CA and CB from **the point C at which the circles cut one another** to the points A and B. Now, since the point A is the center of the circle CDB, therefore AC equals AB. Again, since the point B is the center of the circle CAE, therefore BC equals BA. But AC was proved equal to AB, so each of the segments AC and BC equals AB. Since things which equal the same thing also equal one another, AC also equals BC. Therefore the three segments AC, AB, and BC equal one another. Therefore the triangle ABC is equilateral, and it has been constructed on the given straight segment AB." □

Euclid assumed as intuitive that the two circles should intersect. But suppose for a moment that we lived in the plane $\mathbb{Q} \times \mathbb{Q}$. It would look just like the usual Cartesian plane, except that we would only see points with both coordinates rational. A "circle with center A and radius r" would still be definable as the collection of points in $\mathbb{Q} \times \mathbb{Q}$ at distance r from A. Let us place Cartesian coordinates with the origin in A, and suppose B has coordinates $(1, 0)$. By Pythagoras' theorem, C and D should have coordinates $\left(\frac{1}{2}, y\right)$ and $\left(\frac{1}{2}, -y\right)$, respectively, where y is a solution of

$$y^2 = \frac{3}{4}.$$

Which is a problem, because then C and D would not be in $\mathbb{Q} \times \mathbb{Q}$ (cf. Lemma 56). So in the "rational plane" $\mathbb{Q} \times \mathbb{Q}$, already the first theorem of Euclid's book would be nonsense!

We also point out that this "missing number" (which, after defining the square root, we will call $\frac{\sqrt{3}}{2}$) is the "least upper bound" of the set

$$\left\{ x \in \mathbb{Q} \text{ such that } x^2 < \frac{3}{4} \right\},$$

which consists entirely of rational numbers. So we have also just found out that the least upper bound of a family of elements of $\mathbb{Q}$ need not be in $\mathbb{Q}$! This has consequence on another problem of doing geometry in $\mathbb{Q} \times \mathbb{Q}$, namely, the computation of curve lengths. In his two essays *On the sphere and the Cylinder* and *Measurement of the Circle*, dating back to the third century b.C.,

Archimedes showed that in any circle the ratio between perimeter and diameter is a number (called π) between $3 + \frac{10}{71}$ and $3 + \frac{1}{7}$. Archimedes obtained these bounds by comparing the perimeter of the inscribed regular polygon with 96 edges (the lower bound) and the perimeter of the circumscribed regular polygon with 96 edges (the upper bound). The same method, applied to regular polygons with a higher number of edges, lead to sharper bounds. Quoting Peano[5]:

> "The postulates that were stated by Archimedes in *On the sphere and the cylinder* are equivalent to the following definitions:
>
> - the length of a curvilinear plane convex arc is the common value of the least upper bound of the length of the polygonal inscribed arcs and the greatest lower bound of the circumscribed ones;
> - the area of a convex surface is the common value of the least upper bound of the length of the polygonal inscribed convex surfaces, and the greatest lower bound of the area of the circumscribed ones;
> - the length of a curvilinear arc is the least upper bound of the length of the polygonal inscribed arcs."

What did Archimedes assume as implicit? Well, the belief that such 'limit numbers' *must exist*. But as we saw above, there is no guarantee that the least upper bound of a sequence in $\mathbb{Q}$ is itself in $\mathbb{Q}$! And in fact, one can prove that π is not in $\mathbb{Q}$.

Long story short, our rational plane is somewhat "incomplete": In order to do geometry, we need a larger set than $\mathbb{Q}$. This larger set, called $\mathbb{R}$, is best defined in a topology course; but below we sketch a construction that should give you an idea. A preliminary notation: For any rational number q, we define $|q| \stackrel{\text{def}}{=} \max\{q, -q\}$.

Lemma 58 (Triangular inequality)**.** *For all rational numbers a, b, c, one has*

$$|a + b| \leq |a| + |b|.$$

Proof. From the definition of $|q|$, we have

$$a + b \leq |a| + b \leq |a| + |b|, \text{ and}$$

$$-a - b \leq |a| - b \leq |a| + |b|$$

So $\max\{a + b, -a - b\} \leq |a| + |b|$. Which can be rewritten as $|a + b| \leq |a| + |b|$. □

Definition 59. Let X be a set. A *sequence in X* is a function $a : \mathbb{N} \to X$. For brevity, we denote the image $a(n)$ by a_n.

Definition 60. A *sequence in $\mathbb{Q}$* is called

- *convergent*, if there exists $\ell \in \mathbb{Q}$ (called "limit" of the sequence) such that

$$\forall k \in \mathbb{N}\ \exists M \in \mathbb{N} \text{ such that } \forall n \geq M \text{ we have } |a_n - \ell| < \frac{1}{k+1},$$

  a condition which is usually abbreviated by '$\lim_{n\to\infty} a_n = \ell$';
- *Cauchy*, if

$$\forall k \in \mathbb{N}\ \exists M \in \mathbb{N} \text{ such that } \forall n, m \geq M \text{ we have } |a_n - a_m| < \frac{1}{k+1}.$$

[5] Giuseppe Peano, *Sulla definizione dell'area di una superficie*, Rendiconti dell'Accademia dei Lincei, 1890, 54–57; translated in A. Papadopoulos, *Metric Spaces Convexity and Nonpositive Curvature*, EMS 2005, p. 31.

- *bounded*, if there exists B in $\mathbb{N}$ such that for all n, $|a_n| \leq B$.

Example 61. A constant sequence is bounded, Cauchy, and convergent.

Non-Example 62. The sequence $a_n = n$ is not bounded. We will see in the next Proposition that because of this, it is neither Cauchy, nor convergent.

Proposition 63. *All convergent sequences are Cauchy and all Cauchy sequences are bounded.*

Proof. Using the triangular inequality ($|a_n - a_m| \leq |a_n - \ell| + |a_m - \ell|$), it is easy to see that every convergent sequence is Cauchy. To see that Cauchy sequences are bounded, let us take a Cauchy sequence $(a_n)_{n\in\mathbb{N}}$. If we choose $k = 0$ and $m = M$ in the definition, we get that there exists M such that for all $n \geq M$, one has $|a_n - a_M| < 1$. In particular, for all $n \geq M$, by the triangular inequality one has

$$|a_n| \leq |(a_n - a_M) + a_M| \leq |a_n - a_M| + |a_M| < 1 + |a_M|.$$

Now since they are finitely many, set

$$B \stackrel{\text{def}}{=} \max\{|a_0|, |a_1|, \ldots, |a_{M-1}|, 1 + |a_M|\}.$$

By construction, $|a_n| \leq B$ for all $n \in \mathbb{N}$. □

Example 64. The sequence $a_n = (-1)^n$ is obviously bounded, but it is neither Cauchy nor convergent. In fact, for any M, we are going to find two larger indices $n, m \geq M$ such that $a_n = 1$ and $a_m = -1$, so that $|a_n - a_m| = 2$.

Example 65. There are also sequences that are Cauchy, but not convergent. Let a_n be recursively defined as

$$a_0 \stackrel{\text{def}}{=} 1, \quad z_n \stackrel{\text{def}}{=} max\left\{z \in \{0,1,2,3,4,5,6,7,8,9\} : (a_{n-1} + 10^{-n}z)^2 \leq 2\right\}, \quad a_n \stackrel{\text{def}}{=} a_{n-1} + 10^{-n}z.$$

So $a_0 = 1$, $a_1 = 1.4$, $a_2 = 1.41$, $a_3 = 1.414$, and so on. (Intuitively, a_n is just "the truncation at at the n-th digit of the decimal representation of $\sqrt{2}$", although we cannot define it this way because we have not defined the symbol "$\sqrt{2}$" yet.) This a_n will satisfy the Cauchy property by choosing, for all k, $M = k$: In fact, for all n, m greater than M one clearly has $|a_n - a_m| < 10^{-M} < \frac{1}{M+1}$. However, Pythagoras' discovery that $\mathbb{Q}$ contains no element whose square is 2 implies that this sequence a_n is not convergent in $\mathbb{Q}$.

The idea is to artificially "expand" $\mathbb{Q}$ by inserting all the limits of all Cauchy sequences.

Definition 66. Let $\mathbb{R}$ be the set of all Cauchy sequences in $\mathbb{Q}$, with the following identification: We consider two sequences a_n and d_n identical if their difference $(a - d)_n \stackrel{\text{def}}{=} a_n - d_n$ converges to the constantly-zero sequence.

There is a natural way to view $\mathbb{Q}$ as subset of $\mathbb{R}$, via the map that associates to any rational number q the *constant* sequence, $(q, q, q, q, \ldots)$. However, there are much more elements in $\mathbb{R}$ than those coming from $\mathbb{Q}$. This was proven by Cantor using what is known today as *diagonal argument*. The starting point of this argument is that every real number x can be represented by means of a *decimal representation*:

$$x = a + \sum_{i=1}^{\infty} b_i \cdot 10^{-i}, \text{ with } a \in \mathbb{Z}, b_i \in \{0, 1, \ldots, 9\}.$$

This representation is not always unique. For example, 2.399999... is identical to 2.4 (because the difference tends to zero). However, this is the only thing that can go wrong: If we simply throw out all decimal representations that are eventually always nine, then every real number admits a unique decimal representation.

Theorem 67 (Cantor)**.** $\mathbb{R}$ *is not countable.*

Sketch of proof. It suffices to show that even the interval $(0, 1)$ is not countable. By contradiction, suppose we could list all elements of $(0, 1)$, as

$$x_1, x_2, \ldots, x_n, x_{n+1}, \ldots$$

By what we said above, every element x_i has a decimal representation, which consists of the integer 0 followed by a sequence of digits in $\{0, \ldots, 9\}$. (Remember we have thrown out representations that end with a nine periodic.) Now construct an element y of $\mathbb{R}$ as follows: for the first decimal digit of y, choose either 0 or 1, making sure that your choice does not coincide with the first decimal digit of x_1. In particular, the number y we are writing down will be different from x_1: They differ in the first decimal place. For the second digit of y, again, choose 0 or 1, making sure not to agree with the second decimal digit of x_2. In particular, $y \neq x_2$. And so on: For the i-th digit of y, choose either 0 or 1, disagreeing with the i-th digit of x_i. In the end, by construction you will have produced an element of $\mathbb{R}$ that is different from all x_i. A contradiction: The x_i were supposed to be a complete list of all elements of $\mathbb{R}$. □

Algebraic properties of $\mathbb{R}$

It turns out that many of the nice properties of $\mathbb{Q}$ are inherited by this "expanded set" $\mathbb{R}$. For example, let (a_n) and (b_n) be two sequences in $\mathbb{Q}$. We can define their *sum* as the sequence (c_n) such that $c_n = a_n + b_n$ for all n. Easy exercise: *The sum of two Cauchy sequences is Cauchy.* Note that in $\mathbb{R}$, we are identifying Cauchy sequences whose differences converges to zero, but this identification is compatible with the way we have just defined sums of sequences, in the sense that if $a - d$ converges to zero, then also $(a + b) - (d + b)$ converges to zero, and so in $\mathbb{R}$ the two sequences $(a + b)$ and $(d + b)$ are actually considered the same. This way the *sum* of two real numbers is well-defined.

Similarly, we can define a *product* of two sequences the sequence (c_n) such that $c_n = a_n \cdot b_n$ for all n. Not-so-easy exercise: *The product of any two Cauchy sequences is Cauchy.* Here is the proof. If a, b are Cauchy sequence, then they are both bounded, so there are constants A, B in $\mathbb{Q}$ such that for all n one has $|a_n| \leq A$ and $|b_n| \leq B$. So if we set $C \stackrel{\text{def}}{=} \max\{A, B\}$, we simultaneously have $|a_n| \leq C$ and $|b_n| \leq C$ for all n. Moreover, by definition of "Cauchy sequence", for all k we can find an M' such that for all $n, m \geq M'$, we have

$$|a_n - a_m| < \frac{1}{2C(k+1)},$$

and an M'' such that for all $n, m \geq M''$, we have

$$|b_n - b_m| < \frac{1}{2C(k+1)}.$$

So if we set $M = \max\{M', M''\}$, for all $n, m \geq M$ we simultaneously have

$$|a_n - a_m| < \frac{1}{2C(k+1)} \qquad \text{and } |b_n - b_m| < \frac{1}{2C(k+1)}.$$

Now comes the crucial trick:

$$\begin{aligned} |a_n b_n - a_m b_m| &\leq |a_n b_n - a_n b_m + a_n b_m - a_m b_m| \\ &\leq |a_n b_n - a_n b_m| + |a_n b_m - a_m b_m| \\ &= |a_n| \cdot |b_n - b_m| + |b_m| \cdot |a_n - a_m| \\ &\leq C \cdot |b_n - b_m| + C \cdot |a_n - a_m| \\ &< C \cdot \tfrac{1}{2C(k+1)} + C \cdot \tfrac{1}{2C(k+1)} = \tfrac{1}{k+1}. \end{aligned}$$

This shows that $a \cdot b$ is Cauchy. It remains to check whether it is true or not that if $a - d$ converges to zero, then also $(a \cdot b) - (d \cdot b)$ converges to zero. For this, note that since b is bounded, there exists B such that $|b_n| \leq B$ for all n. Then since $a - d$ converges to zero, there exists M' such that for all $n \geq M'$ one has

$$|a_n - d_n| \leq \frac{1}{(k+1)B}.$$

In particular, for all $n \geq M'$ one has then

$$|a_n b_n - d_n b_n| = |(a_n - d_n)b_n| = |a_n - d_n| \cdot |b_n| \leq |a_n - d_n| \cdot B < \frac{1}{(k+1)B} \cdot B,$$

which shows that $(a \cdot b) - (d \cdot b)$ converges to zero.

This very long foreword has now come to a conclusion: **The operations of sum and product extend from $\mathbb{Q}$ to $\mathbb{R}$.** Not only this is true, but all the main properties of sums and products are maintained in the expansion.

Theorem 68. *$\mathbb{R}$ is endowed with two internal operations $+$ and $\cdot$ that satisfy the axioms:*

F1 *The operation $+$ is* associative. *That is, for all x, y, z in $\mathbb{R}$, $x + (y + z) = (x + y) + z$.*

F2 *The operation $+$ is* commutative. *That is, for all x, y, z in $\mathbb{R}$, $x + y = y + x$.*

F3 *The operation $+$ has a (unique)* neutral element. *That is, there exists an element z in $\mathbb{R}$ such that for all x in $\mathbb{F}$, $x + z = x$. This z is the constantly-zero sequence, which we denote simply by "0".*

F4 *Every element has a unique* additive inverse. *That is, for all x in $\mathbb{R}$ there exists exactly one element y in $\mathbb{F}$ such that $x + y = 0$. From now on we denote such element by "$-x$".*

F5 *The operation $\cdot$ is* associative. *That is, for all x, y, z in $\mathbb{R}$, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.*

F6 *The operation $\cdot$ is* commutative. *That is, for all x, y, z in $\mathbb{R}$, $x \cdot y = y \cdot x$.*

F7 *The operation $\cdot$ has a (unique)* neutral element. *That is, there exists a (unique) element $z \neq 0$ in $\mathbb{R}$ such that for all x in $\mathbb{R}$, $xz = x$. Since this z is the constantly-one sequence, from now on we denote such neutral element by "1".*

F8 *Every element except 0 has a (unique)* multiplicative inverse. *That is, for all $x \neq 0$ in $\mathbb{R}$ there exists exactly one element y in $\mathbb{R}$ such that $xy = 1$. From now on we denote such element by "x^{-1}".*

F9 *The operation $\cdot$* distributes *$+$: for all x, y, z in $\mathbb{R}$, $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.*

Proof. We only prove two, leaving the others to a very, very patient student.

COMMUTATIVITY. Every element of $\mathbb{R}$ is a Cauchy sequence in $\mathbb{Q}$. From the way we defined the sum of sequences (namely, "pointwise"),

$$(a + b)_n = a_n + b_n = b_n + a_n = (b + a)_n.$$

EXISTENCE OF MULTIPLICATIVE INVERSE. Let a be a Cauchy sequence different than 0 in $\mathbb{R}$. Because it is Cauchy, for any k there is an M' such that for all $n, m \geq M'$ we have $|a_n - a_m| < \frac{1}{2(k+1)}$. Now, remember that in $\mathbb{R}$ two sequences are identified if their difference converges to 0. So, "$a \neq 0$ in $\mathbb{R}$" really means that a does not converge to zero. That means that there exists k such that, for all M (and in particular for $M = M'$), there is at least one index $n \geq M$ such that $|a_n| > \frac{1}{k+1}$. By the triangular inequality, then,

$$\frac{1}{k+1} < |a_n| = |a_n - a_m + a_m| \leq |a_n - a_m| + |a_m| < \frac{1}{2(k+1)} + |a_m|.$$

So there exists k and there exists M' such that, for all $m \geq M'$, we have $|a_m| > \frac{1}{k+1}$. Thus we can define a new sequence b as follows:

$$|b_m| = \begin{cases} 1 & \text{if } m < M', \\ (a_m)^{-1} & \text{if } m \geq M'. \end{cases}$$

What we have just said implies that b is bounded: For all m, since $|a_m| > \frac{1}{k+1}$, we have $|b_m| < k+1$. We need some more effort to show that b is Cauchy. To see this, since a and b are bounded, there is a C such that for all n both $|a_n| < C$ and $|b_n| < C$ are true. In other words, for all n both $1 < \frac{C}{|a_n|}$ and $1 < \frac{C}{|b_n|}$ are true. But then for $n, m \geq M'$ (the value above),

$$\begin{aligned} |b_n - b_m| = 1 \cdot |b_n - b_m| \cdot 1 &< \tfrac{C}{|b_n|} \cdot |b_n - b_m| \cdot \tfrac{C}{|b_m|} \\ &= C^2 \left| \tfrac{b_n - b_m}{b_n \cdot b_m} \right| \\ &= C^2 \left| \tfrac{1}{b_m} - \tfrac{1}{b_n} \right| \\ &= C^2 \left| a_m - a_n \right|. \end{aligned}$$

Now a is Cauchy, so for each k we can find an $M'' \geq M'$ such that for all $m, n \geq M''$, we have

$$|a_m - a_n| < \frac{1}{C^2 k + 1}.$$

And so t for all $m, n \geq M''$, we can show that $|b_n - b_m| < \frac{1}{C^2 k+1}$, as desired. □

There is, however, a final bonus. We can define square roots of all *positive* real numbers. Let's go back to Example 65. For any rational number $q \geq 1$, let us define

$$a_0 \stackrel{\text{def}}{=} 1, \quad z_n \stackrel{\text{def}}{=} max\left\{z \in \{0,1,2,3,4,5,6,7,8,9\} \,:\, (a_{n-1} + 10^{-n} z)^2 \leq q\right\}, \quad a_n \stackrel{\text{def}}{=} a_{n-1} + 10^{-n} z.$$

This is a Cauchy increasing sequence and so in $\mathbb{R}$ it has a limit. Call it ℓ. By excluding the other two cases $\ell^2 < q$ and $\ell^2 > q$, it is not difficult to see that it must be $\ell^2 = q$. Thus for any rational number $q \geq 1$, there is an $\ell > 0$ in $\mathbb{R}$ that satisfies $\ell^2 = q$. But then for any rational number $0 < q < 1$, there is also $\ell' > 0$ in $\mathbb{R}$ that satisfies $(\ell')^2 = q$: Namely, $\ell' \stackrel{\text{def}}{=} \frac{1}{\ell}$, where ℓ is the positive real number that satisfies $\ell^2 = \frac{1}{q}$.

This works not just for all positive $q \in \mathbb{Q}$, but also for all positive $q \in \mathbb{R}$. In fact, set

$$\begin{array}{rccc} \sqrt{} : & \mathbb{R}_{\geq 0} & \longrightarrow & \mathbb{R}_{\geq 0} \\ & 0 & \longmapsto & 0 \\ & x > 0 & \longmapsto & \text{the unique } \ell > 0 \text{ such that } \ell^2 = x. \end{array}$$

It is easy to see that this function is increasing, i.e. if $0 < x < y$, then $\sqrt{x} < \sqrt{y}$.

Complex numbers

Once $\mathbb{R}$ is defined, it is somewhat disturbing that equations of the type $x^2 = r$ still have no solutions over $\mathbb{R}$ if r is negative. To fix this, it is easy to define the set of complex numbers

$$\mathbb{C} \stackrel{\text{def}}{=} \{a + bi \text{ such that } a, b \in \mathbb{R}\}.$$

Recall that i is short for "imaginary unit", so $i^2 = -1$. Addition in $\mathbb{C}$ is defined as

$$(a + bi) + (c + di) \stackrel{\text{def}}{=} (a + c) + (b + d)i,$$

whereas the formula for multiplying is

$$(a + bi) \cdot (c + di) \stackrel{\text{def}}{=} (ac - bd) + (ad + bc)i.$$

Lemma 69. $\mathbb{C}$ *contains "all square roots": That is, For any* z *in* $\mathbb{C}$*, there exists* $\delta \in \mathbb{C}$ *such that* $\delta^2 = z$*.*

Proof. Let $z = a + ib$. If $b = 0$ then $z \in \mathbb{R}$ and the claim is easy: When $a \geq 0$, δ is a real number satisfying $x^2 = a$, and when $a < 0$, δ is i times a real number satisfying $x^2 = -a$. So let us assume $b \neq 0$. We look for an element $\delta = x + iy$ such that

$$a + ib = (x + iy)(x + iy) = x^2 - y^2 + i(2xy).$$

In other words, given real numbers a and b, we need to solve over $\mathbb{R}$ the system

$$\begin{cases} x^2 - y^2 &= a \\ 2xy &= b. \end{cases}$$

Substituting $y = \frac{b}{2x}$ into the first equation, we obtain $[x^2 - \frac{b^2}{4x^2} = a$, or in other words, $4x^4 - b^2 = 4ax^2$. Setting $t \stackrel{\text{def}}{=} x^2$ and imposing $t > 0$, this gives rise to a quadratic equation,

$$4t^2 - 4at - b^2 = 0,$$

which has a positive solution, namely, $t = \frac{2a + \sqrt{4a^2 + 4b^2}}{4}$. (It is positive because $\sqrt{4a^2 + 4b^2} \geq \sqrt{4a^2} = |2a|$.) Once we found t, we immediately derive $x = \sqrt{t}$ and $y = \frac{b}{2x}$. □

Lemma 70. *For any* $(a, b) \neq (0, 0)$ *in* $\mathbb{R}^2$*, the element* $a + ib$ *has a multiplicative inverse in* $\mathbb{C}$*.*

Proof. If $(a, b) \neq (0, 0)$, then $r \stackrel{\text{def}}{=} a^2 + b^2$ is a nonzero real number, and

$$(a + ib)\left(\frac{a}{r} - \frac{b}{r}i\right) = \left(\frac{a^2}{r} + \frac{b^2}{r}\right) + i\left(\frac{ab}{r} - \frac{ab}{r}\right) = \frac{a^2 + b^2}{r} + i \cdot 0 = 1.$$ □

0.6 Exercises

0. Recalling that $\binom{n}{k} \stackrel{\text{def}}{=} \frac{n!}{k!(n-k)!}$, prove that for any integers $n \geq k \geq 1$ one has

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1}.$$

1. Use induction on n to prove Newton's formula:

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

   Hint: You may use exercise 0 and the following "reindexing trick":

$$\sum_{k=0}^{n} \binom{n}{k} a^{k+1} \, b^{n-k} = \sum_{k=1}^{n+1} \binom{n}{k-1} a^k \, b^{n+1-k}.$$

2. Let n be a positive integer. Prove that for any $k \in \{0, \ldots, n\}$, one has

$$\binom{n}{k} < 2^n.$$

3. Use induction to prove the *pidgeonhole principle*: any map from a set with $n+1$ objects to a set with n objects, is not injective.

4. Use induction to prove the *generalized pidgeonhole principle*: If more than kn objects are placed into n boxes, then at least one box must contain more than k objects. (The case $k = 1$ is the pidgeonhole principle.)

5. Prove that
$$1^3 + 2^3 + 3^3 + \ldots + n^3 = \frac{1}{4}n^2(n+1)^2.$$

6. Prove that $n! > 3^n$ for n large.

7. Prove that for all $i \in \mathbb{N}$ and for all integers $n \geq 1$
$$\sum_{k=1}^{n} \binom{i+k-1}{i} = \binom{n+i}{i+1}.$$

8. Compute the $\gcd(528, 303)$ using the Euclidean algorithm.

9. Prove the "Euclidean division for $\mathbb{Z}$" (given two integers a, b, with $b \neq 0$, there exists a unique pair of integers (r, q) such that $a = bq + r$, and $0 \leq r < |b|$") by induction on $|a|$.

10. Find all integer solutions of the equation $2x + 3y = 15$.

11. Let a, b be positive integers such that $\gcd(a, b) = 1$. Let
$$\mathcal{H}_0^-(a, b) \stackrel{\text{def}}{=} \{(x, y) \in \mathbb{Z} \times \mathbb{Z} \text{ such that } ax - by = 0\}$$
be the set of integer solutions of $ax - by = 0$. Prove that
$$\mathcal{H}_0^-(a, b) = \{(bk, ak) \text{ such that } k \in \mathbb{Z}\}.$$

12. Find all integer solutions of $6x - 9y = 15$. Can you find all *positive* integer solutions?

13. Think of a number. Square it. Divide the result by 3. Why is the remainder always 0 or 1, but never 2? Justify your answer.

14. What is the last digit of 3^{2001}? Hint: work in $\mathbb{Z}_{10}$.

15. What are the last two digits of 913250946798^6? Hint: work in $\mathbb{Z}_{100}$. What about the last three digits? Where would you work?

16. Compute $12345678^{23456789} \bmod 3$.

17. Suppose that a number x can be written in the decimal representation as "*abcabc*", with a, b, c decimal digits. (For example, $x = 285285$.) Show that x is always a multiple of 13.

18. Let n be any integer ≥ 2. Prove that
$$n \text{ is prime } \iff \text{ for all integers } k \in \{1, \ldots, n-1\}, \ \binom{n}{k} \text{ is a multiple of } n.$$

19. (Freshman's Dream) Use the previous result to show that for any prime number p, for any integers x, y,
$$(x + y)^p \equiv x^p + y^p \mod p.$$

20. (Fermat's Little Theorem, 1640) Use the Freshman's Dream and induction on a to show the following: if p is any prime number, then for any $a \in \mathbb{N}$ one has

$$a^p \equiv a \mod p.$$

21. Find the remainder of the division of 11^{118} by 59. (Hint: use Fermat's little theorem.)

22. Write the number 73 on a piece of paper, fold it up, and give it to an unsuspecting friend. Ask your friend to write his/her birth year twice in a calculator. (E.g., I would write 19821982.) Then ask your friend if the number is divisible by any chance by 137; ask him/her to verify with the calculator. Then say, "please divide the result by your birthyear". Ask your friend to unwrap the paper: the calculator and the piece of paper will magically tell the same number, 73! Can you spoil the magic and explain the trick?

23. Under what conditions is a six-digit number whose decimal digits are *abcabc* divisible by $7, 9, 11$ and 13? (For example, 135135 is divisible by all of them.)

24. For any positive integers a and b, prove that $\gcd(a, b) \cdot \mathrm{lcm}(a, b) = ab$.

1 Permutations and Matrices

1.1 Permutations

Let n be any positive integer. Let $[n] \stackrel{\text{def}}{=} \{1, \ldots, n\}$. Let

$$\mathcal{S}_n \stackrel{\text{def}}{=} \{\sigma : [n] \longrightarrow [n] \text{ bijective}\}.$$

The elements of $\mathcal{S}_n$ are called *permutations*.

Lemma 71. *The composition of injective (resp. surjective) maps is injective (resp. surjective). In particular, the composition of any two permutation is a permutation.*

Proof. Let $f : X \to Y$ and $g : Y \to Z$ be arbitrary functions.

- if f, g are injective and $gf(x) = gf(x')$, then by the injectivity of g one has $f(x) = f(x')$, which by the injectivity of f implies $x = x'$;
- if f, g are surjective and $z \in Z$, pick any y in Y such that $g(y) = z$, then pick any x in X such that $f(x) = y$: by construction, $gf(x) = g(y) = z$. □

Note that if σ and τ are both invertible, then the inverse of $\sigma\tau$ is the function $\tau^{-1}\sigma^{-1}$.

Proposition 72. *The set $\mathcal{S}_n$ with the operation of composition satisfies the following properties:*
(a) Closure. *If σ, τ are in $\mathcal{S}_n$, so is $\sigma\tau$.*
(b) Associativity. *If ρ, σ, τ are in $\mathcal{S}_n$, $\rho(\sigma\tau) = (\rho\sigma)\tau$.*
(c) Identity. *There is a unique function 1 (namely, the identity on $\mathcal{S}_n$) such that for every σ in $\mathcal{S}_n$, $\sigma 1 = \sigma = 1\sigma$.*
(d) Inverses. *For every σ in $\mathcal{S}_n$, there exists a unique τ in $\mathcal{S}_n$ with the property that $\sigma\tau = 1 = \tau\sigma$.*

There are three types of notation to write down the same permutation:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}, \quad \sigma = (12)(456), \quad \text{and } \sigma = (12)(45)(56).$$

(The second notation writes σ as product of disjoint cycles; the third, as product of non-disjoint flips; we will explain them in a few minutes.) The first notation is called *two-line notation*. The rule behind it is, σ maps each elements of the first row into the element of the second row immediately below. (In this case, the first row is ordered, but it does not have to be: What matters is that below each i sits $\sigma(i)$.) For example, $\sigma(3) = 3$. To compose two functions, we write them on top of one another, remembering that when we write $\tau \circ \sigma$ the first permutation applied is σ, so σ should be on top. The two-line notation of $\tau \circ \sigma$ is then obtained by looking at only the first and the last row, ignoring all intermediate ones. For example, if

$$\tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \end{pmatrix} \quad \text{and} \quad \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix}, \quad \text{then}$$

$$\tau \circ \sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \\ 2 & 1 & 4 & 5 & 6 & 3. \end{pmatrix} \quad \text{and} \quad \sigma \circ \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 3 & 5 & 6 \\ 2 & 1 & 5 & 3 & 6 & 4. \end{pmatrix}$$

En passant, notice that $\tau \circ \sigma$ and $\sigma \circ \tau$ are different, so the operation is not commutative.

Definition 73. Let $2 \leq k \leq n$ be integers. A *cycle (of length k)* in a permutation $\sigma \in \mathcal{S}_n$ is a k-tuple

$$(a_1, a_2, \ldots, a_k),$$

such that $a_i < a_{i+1}$, $\sigma(a_k) = a_1$ and $\sigma(a_i) = a_{i+1}$ for all $i \in \{1, \ldots, k-1\}$. Cycles of length two are called *flips* (or *transpositions*).
Any cycle g (of length k) is naturally associated to a permutation $\gamma \in \mathcal{S}_n$, as follows: $\gamma = \sigma$ on the elements of the cycle, and $\gamma = id$ otherwise.

Example 74. In the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 1 & 3 & 5 & 6 & 4 \end{pmatrix},$$

there is a cycle of length 3, namely, $g = (4, 5, 6)$. Its associated permutation is

$$\gamma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 2 & 3 & 5 & 6 & 4 \end{pmatrix}.$$

Theorem 75. *Every permutation different than the identity can be written as product of disjoint cycles, in a unique way (up to changing the order of the cycles).*

Proof. Let a_1 be the smallest integer such that $\sigma(a_1) \neq a_1$. Let t_1 be the smallest integer such that $\sigma^{t_1}(a_1) = a_1$. Then the first cycle is

$$\left(a_1, \sigma(a_1), \sigma^2(a_1), \ldots, \sigma^{t_1-1}(a_1)\right).$$

Now let a_2 be the smallest integer that does not belong to the cycle above, and satisfies $\sigma(a_2) \neq a_2$. Let t_2 be the smallest integer such that $\sigma^{t_2}(a_2) = a_2$. The second cycle is

$$\left(a_2, \sigma(a_2), \sigma^2(a_2), \ldots, \sigma^{t_2-1}b(a_2)\right).$$

And so on. We sketch the algorithm with the help of an example. Suppose

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 8 & 4 & 6 & 5 & 7 & 9 & 2 & 1 \end{pmatrix}.$$

To find the first cycle, we start with 1 and apply iteratively σ, until we get back to 1. So

$$\sigma(1) = 3, \quad \sigma(3) = 4, \quad \sigma(4) = 6, \quad \sigma(6) = 7, \quad \sigma(7) = 9, \quad \sigma(9) = 1.$$

So the first cycle is $(1, 3, 4, 6, 7, 9)$. Now let us consider the smallest integer not contained in this cycle, and apply σ repeatedly, until we get back to such integer. In our case, we re-start with 2:

$$\sigma(2) = 8, \quad \sigma(8) = 2.$$

So $(2, 8)$ is the second cycle. By construction, it is disjoint from the first cycle, because σ is injective. Now the smallest integer that belongs to neither of the previous cycles is 5. Since $\sigma(5) = 5$, we are done. Our final result is

$$\sigma = (1, 3, 4, 6, 7, 9)(2, 8).$$

Now, technically what we found is just a *list* of disjoint cycles. But if we interpreted every cycle as its associated permutation in $\mathcal{S}_n$, the list can actually be interpreted a product of permutations. More precisely, if $\gamma, \gamma_1, \gamma_2$ are the permutations of $\mathcal{S}_n$ associated respectively to σ, to $(1, 3, 4, 6, 7, 9)$, and to $(2, 8)$, then it is clear that

$$\gamma = \gamma_1 \circ \gamma_2.$$

For this reason, we speak of "product of cycles". Note that disjoint cycles commute:

$$\gamma = \gamma_1 \circ \gamma_2 = \gamma_2 \circ \gamma_1.$$

To complete our "proof by example", we claim that up to commuting the disjoint cycles, this decomposition is unique. This is easy: Let

$$\gamma = \eta_1 \circ \ldots \circ \eta_k$$

be another decomposition into disjoint cycles. Without loss of generality, suppose 1 appears in η_1. Since $\sigma(1) = 3, \sigma(3) = 4$, etc., it is clear that η_1 must be the cycle $(1, 3, 4, 6, 7, 9)$. Similarly, suppose η_2 is the cycle containing 2: Then $\eta_2 = (2, 8)$. Since $\eta(5)$ must be 5, we conclude that $\gamma_i = \eta_i$ for all i. □

Lemma 76. *Every cycle of length k can be written as product of $k-1$ non-disjoint flips (not necessarily in a unique way).*

Proof. If $k \geq 2$, we claim that

$$(a_1, a_2, \ldots, a_k) = (a_1, a_2)(a_2, a_3) \cdots (a_{k-2}, a_{k-1})(a_{k-1}, a_k).$$

By this we mean that if γ is the permutation of $\mathcal{S}_n$ associated to $(a_1, \ldots, a_k)$, and γ_i is the permutation of $\mathcal{S}_n$ associated to (a_i, a_{i+1}), then

$$\gamma = \gamma_1 \circ \gamma_2 \circ \ldots \circ \gamma_{k-1}.$$

As a warm up, let us check this first for the element a_1. By definition, $\gamma(a_1) = a_2$. On the other hand, γ_i swaps a_i with a_{i+1}, so it has no effect on a_1 if $i \geq 2$. Formally,

$$\gamma_i(a_1) = \begin{cases} a_1 & \text{if } i \geq 2 \\ a_2 & \text{if } i = 1. \end{cases}$$

So

$$\gamma_1 \circ \gamma_2 \circ \ldots \circ \gamma_{k-1}(a_1) = \gamma_1 \circ \gamma_2 \circ \ldots \circ \gamma_{k-2}(a_1) = \ldots = \gamma_1(a_1) = a_2,$$

as desired. Now let us check the effect on the generic element a_j, with $j < k$. Clearly $\gamma(a_j) = a_{j+1}$, with the exception of a_k, for which $\gamma(a_k) = a_1$. On the other hand,

$$\gamma_i(a_j) = \begin{cases} a_j & \text{if } i \geq j+1 \\ a_{j+1} & \text{if } i = j \\ a_{j-1} & \text{if } i = j-1 \\ a_j & \text{if } i \leq j-2. \end{cases}$$

So if $j < k$, we have

$$\gamma_1 \circ \ldots \circ \gamma_{k-1}(a_j) = \ldots = \gamma_1 \circ \ldots \circ \gamma_j(a_j) = \gamma_1 \circ \ldots \circ \gamma_{j-1}(a_{j+1} = \ldots = \gamma_1(a_{j+1}) = a_{j+1}.$$

For a_k instead we have

$$\gamma_1 \circ \ldots \circ \gamma_{k-1}(a_k) = \gamma_1 \circ \ldots \circ \gamma_{k-2}(a_{k-1}) = \gamma_1 \circ \ldots \circ \gamma_{k-3}(a_{k-2}) = \ldots = \gamma_1(a_2) = a_1. \qquad \square$$

Example 77. Let us verify that $(1,3,4,6,7,9) = (1,3)(3,4)(4,6)(6,7)(7,9)$. In fact, the right hand side is given by the first and the last row of the matrix

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 2 & 3 & 4 & 5 & 6 & 9 & 8 & 7 \\ 1 & 2 & 3 & 4 & 5 & 7 & 9 & 8 & 6 \\ 1 & 2 & 3 & 6 & 5 & 7 & 9 & 8 & 4 \\ 1 & 2 & 4 & 6 & 5 & 7 & 9 & 8 & 3 \\ 3 & 2 & 4 & 6 & 5 & 7 & 9 & 8 & 1 \end{pmatrix}$$

and the left hand side is precisely

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 2 & 4 & 6 & 5 & 7 & 9 & 8 & 1 \end{pmatrix}$$

Definition 78. A permutation $\sigma \in \mathcal{S}_n$ is called *even* if it can be written as the product of an even number of flips, and *odd* if it can be written as the product of an odd number of flips.

Example 79. A k-cycle is an even permutation if k is odd, and an odd permutation if k is even. In fact, any k-cycle is the product of $k-1$ flips.

Remark 80. A priori, it could be that a permutation is both even and odd. (We will actually prove later that this cannot happen.) Note that independently from this result, we can still say that the composition of two even permutation is an even permutation.

Lemma 81. *If a permutation is even (resp. odd), so is its inverse*

Proof. Every flip is the inverse of itself. So if $\sigma = \gamma_1\gamma_2\cdots\gamma_{\ell-1}\gamma_\ell$, with γ_i flips, then

$$\sigma^{-1} = (\gamma_1\,\gamma_2\,\cdots\,\gamma_{\ell-1}\,\gamma_\ell)^{-1} = (\gamma_\ell)^{-1}(\gamma_{\ell-1})^{-1}\cdots(\gamma_2)^{-1}(\gamma_1)^{-1} = \gamma_\ell\,\gamma_{\ell-1}\,\cdots\,\gamma_2\,\gamma_1. \qquad \square$$

Theorem 82. *No permutation of $\mathcal{S}_n$ is both even and odd.*

Proof. By contradiction, suppose σ is both even and odd. By the previous Lemma, so is σ^{-1}. By composing them, we obtain in particular that the identity is odd. Let ℓ be the smallest odd number such that the identity can be written as a product of ℓ flips. Obviously $\ell > 1$. Write

$$id = \gamma_1\gamma_2\cdots\gamma_\ell$$

with γ_i flips, and suppose $\gamma_\ell = (a,b)$. Consider $\gamma_{\ell-1}$ and the effect it might possibly have on a and b: In $\gamma_{\ell-1}$, either both a and b could be transposed; or just a; or just b; or neither of them.

- If $\gamma_{\ell-1} = (a,b)$, then it is the inverse of γ_ℓ, so we could write the identity more succinctly as $\gamma_1\gamma_2\cdots\gamma_{\ell-2}$. A contradiction with how we chose ℓ.
- If $\gamma_{ell-1} = (a,d)$, with $d \neq b$, then

$$\gamma_{\ell-1}\gamma_\ell = (a,d)(a,b) = (a,b,d) = (a,b)(b,d).$$

- If $\gamma_{ell-1} = (b,d)$, with $d \neq a$, then

$$\gamma_{\ell-1}\gamma_\ell = (b,d)(a,b) = (a,d,b) = (a,d)(b,d).$$

- If $\gamma_{ell-1} = (c,d)$, with $\{a,b\} \cap \{c,d\} = \emptyset$, then

$$\gamma_{\ell-1}\gamma_\ell = (c,d)(a,b) = (a,b)(c,d).$$

So the first case is impossible, and in the other three cases, we can move the flip that moves a to the left (leaving as "last flip" a flip that does not touch a). Now we repeat the argument for $\gamma_{\ell-1}$ and $\gamma_{\ell-2}$. And so on, until the only flip that moves a is γ_1. But this is a contradiction: If among all flips $\gamma_1, \ldots, \gamma_\ell$, only γ_1 moves a, how can their composition be the identity? Their composition will move a as well! □

Proposition 83. *For $n \geq 2$, the set $\mathcal{S}_n$ has exactly $n!$ elements, whereas the set*

$$A_n \stackrel{\text{def}}{=} \{\textit{even permutations}\}$$

has exactly $\frac{n!}{2}$ elements.

Proof. Consider the following function between sets

$$\begin{array}{rccc} \psi: & A_n & \longrightarrow & (\mathcal{S}_n \setminus A_n) \\ & \sigma & \longmapsto & \sigma \circ (1,2). \end{array}$$

This function is well-defined and bijective, the inverse being

$$\begin{array}{rccc} \phi: & (\mathcal{S}_n \setminus A_n) & \longrightarrow & A_n \\ & \tau & \longmapsto & \tau \circ (1,2). \end{array}$$

This proves that A_n and its complement within $\mathcal{S}_n$ have the same number of elements. On the other hand, $\mathcal{S}_n$ has $n!$ elements, because to write down a bijection $\sigma : [n] \to [n]$ we have n choices for $\sigma(1)$, $n-1$ choices for $\sigma(2)$, and so on. Hence, A_n has $\frac{n!}{2}$ elements. □

1.2 Matrices and determinants

Throughout this subsection, let $\mathbb{K}$ be one the sets $\mathbb{Q}, \mathbb{R}$, $\mathbb{C}$ or $\mathbb{Z}_p$, with p prime.

Definition 84. An $m \times n$ *matrix* A is an array of numbers in $\mathbb{K}$ called *entries*, arranged in rows (numbered from top to bottom) and columns (numbered from left to right). The entry in row i and column j is denoted by $a_{i,j}$. Thus a generic $m \times n$ matrix looks like

$$\begin{pmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,n} \\ \vdots & & \ddots & \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{pmatrix}$$

Definition 85. Given an $\ell \times m$ matrix A and an $m \times n$ matrix B , their *(row-by-column) product* AB is the $\ell \times n$ matrix C with entries defined by

$$c_{i,j} = \sum_{k=1}^{m} a_{i,k} b_{k,j}.$$

Note that $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}$, whereas $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b & a \\ d & c \end{pmatrix}$. So this "product" is very much not commutative: It could be that all entries of AB are different than the corresponding entries of BA. However, the row-by-column product is associative, in the sense that $A(BC) = (AB)C$ for any three matrices of sizes such that these two expressions make sense. Here is a proof: if A is an $m \times a$ matrix, B an $a \times b$ matrix, C a $c \times n$ matrix, then

$$\begin{array}{ll} A(BC)_{i,j} & = \sum_{k=1}^{a} a_{i,k} BC_{k,j} = \sum_{k=1}^{a} a_{i,k} \sum_{h=1}^{b} b_{k,h} c_{h,j} = \sum_{k=1}^{a} \sum_{h=1}^{b} a_{i,k} b_{k,h} c_{h,j} \quad = \\ & = \sum_{h=1}^{b} (\sum_{k=1}^{a} a_{i,k} b_{k,h}) c_{h,j} = \sum_{h=1}^{b} AB_{i,h} c_{h,j} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad = (AB)C_{i,j}. \end{array}$$

Moreover, for any $m \in \mathbb{N}$, let us call I_m the $m \times m$ matrix with ones on the main diagonal and zeroes elsewhere (i.e. $I_{k,k} = 1$ and $I_{j,k} = 0$ if $j \neq k$.) It is easy to see that for all $\ell \times m$ matrices A and for all $m \times n$ matrices B one has $AI_m = A$ and $I_m B = B$.

Definition 86. A matrix is called *square* if $m = n$, i.e., if it has the same number of rows and columns. The *diagonal elements* of an $n \times n$ matrix are $\{a_{i,i}$ such that $1 \leq i \leq n\}$. A square matrix is *upper triangular* if $a_{ij} = 0$ for all $i > j$. A square matrix is *lower triangular* if $a_{ij} = 0$ for all $i < j$. A square matrix is *diagonal* if it is both upper triangular and lower triangular. The *inverse* of an $n \times n$ matrix A is a matrix B such that $AB = BA = I_n$, if any such B exists.

Definition 87 (Determinant)**.** The *determinant* of a square matrix A is defined by

$$\det A \stackrel{\text{def}}{=} \sum_{\sigma \in \mathcal{S}_n} \operatorname{sgn}(\sigma)\, a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)},$$

where $\operatorname{sgn} \sigma$ is $+1$ if σ is even and -1 if σ is odd.

Proposition 88. *If A is a diagonal matrix, $\det A$ is the product of the diagonal elements. In particular, $\det I_n = 1$.*

Proof. If σ is any permutation different than the identity, then there exists an i such that $\sigma(i) \neq i$, and thus $a_{i,\sigma(i)} = 0$, by definition of diagonal matrix. This means that the product $a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdots a_{n,\sigma(n)}$ is zero, because one of its factors is zero. This is true for every $\sigma \neq id$: Thus the only nonzero summand in the definition of determinant of A is the one corresponding to the identity permutation, namely, $a_{1,1} \cdot a_{2,2} \cdots a_{n,n}$. □

The main result of the theory of determinants is the following theorem, whose proof occupies half of our Linear Algebra course at UM, and so is omitted here for reasons of time.

Theorem 89 (Cauchy–Binet)**.** *For any two $n \times n$ matrices A, B,*

$$\det(AB) = \det A \cdot \det B = \det(BA).$$

Moreover, $\det A \neq 0$ if and only if A has an inverse; if and only if A has a right (or left) inverse.

Corollary 90. *If $AB = I_n$, then both A and B are invertible, and $A^{-1} = B$.*

Proof. See the Exercises □

Given Cauchy–Binet's theorem, of particular interest are two sets:

Definition 91. The set GL(n, $\mathbb{K}$) (called "general linear group") consists of all invertible $n \times n$ matrices with entries in $\mathbb{K}$. The set SL(n, $\mathbb{K}$) (called "special linear group") consists of all $n \times n$ matrices with entries in $\mathbb{K}$ and determinant equal to 1.

Proposition 92. *$G = GL(n, \mathbb{K})$ with row-by-column multiplication satisfies:*
(a) Closure. *If A, B are in G, so is AB.*
(b) Associativity. *If A, B, C are in G, $A(BC) = (AB)C$.*
(c) Identity. *There is a unique matrix I (namely, the identity matrix I_n) such that for every A in G, $AI = A = IA$.*
(d) Inverses. *For every A in G, there exists a unique B in G with the property that $AB = I_n = BA$.*
The same is true if we replace G with $G' = SL(n, \mathbb{K})$.

Proposition 93 (cf. Rotman[6]). *Given n positive integers $a_1, \ldots, a_n$,*

$$\gcd(a_1, \ldots, a_n) = 1 \iff \begin{array}{l} \exists \text{ an } n \times n \text{ matrix } A \text{ with entries in } \mathbb{Z} \text{ whose} \\ \text{first row is } a_1, \ldots, a_n, \text{ and whose determinant is } 1. \end{array}$$

Proof.

"$\Leftarrow$" Set $d \stackrel{\text{def}}{=} \gcd(a_1, \ldots, a_n)$. By definition, the determinant of A is a sum of $n!$ products of integers, the first of which integers is a multiple of d. So $\det A$ is also a multiple of d. Since $\det A = 1$, we conclude $d = 1$.

"$\Rightarrow$" By induction on n. For $n = 2$ this is essentially Bezout's theorem (Theorem 48): If $\gcd(a_1, a_2) = 1$, we can find integers x_1, x_2 such that $a_1 x_1 + a_2 x_2 = 1$, and so the matrix

$$A \stackrel{\text{def}}{=} \begin{pmatrix} a_1 & a_2 \\ -x_2 & x_1 \end{pmatrix}$$

has determinant 1, as desired. For larger n, set

$$a \stackrel{\text{def}}{=} \gcd(a_1, \ldots, a_{n-1}) \qquad \text{and} \quad b_i \stackrel{\text{def}}{=} \frac{a_i}{a} \quad (\text{for } i = 1, \ldots, n-1).$$

By construction, $\gcd(b_1, \ldots, b_{n-1}) = 1$. So we can apply the inductive assumption and find an $(n-1) \times (n-1)$ matrix B with entries in $\mathbb{Z}$ whose first row is $b_1, \ldots, b_{n-1}$ and whose determinant is 1. Let C be the submatrix formed by the lower $n-2$ rows of B. Note that C is not square (it has $n-2$ rows and $n-1$ columns), but by definition

$$\det \begin{pmatrix} b_1 & \cdots & b_{n-1} \\ & C & \end{pmatrix} = \det B = 1.$$

Now, it is easy to see that $\gcd(a, a_n) = \gcd(a_1, \ldots, a_n) = 1$. So by Bezout's Theorem 48, we can find integers s and t such that

$$a_n t + a s = 1.$$

With these integers and with the aforementioned $(n-2) \times (n-1)$ matrix C, let us create the new $n \times n$ matrix

$$A \stackrel{\text{def}}{=} \begin{pmatrix} ab_1 & \cdots & ab_{n-1} & a_n \\ & C & & 0 \\ -tb_1 & \cdots & -tb_{n-1} & s \end{pmatrix}.$$

Let us see that A is the desired matrix. Indeed, A has integer entries and first row equal to $(a_1, \ldots, a_{n-1}, a_n)$. It remains to show that $\det A = 1$. This is easy if you know a couple of tricks to compute determinants:

$$\begin{aligned} \det A &= (-1)^{n+1} a_n \det \begin{pmatrix} & C & \\ -tb_1 & \cdots & -tb_{n-1} \end{pmatrix} + (-1)^{2n} s \cdot \det \begin{pmatrix} ab_1 & \cdots & ab_{n-1} \\ & C & \end{pmatrix} = \\ &= (-1)^{(n+1)} (-1)^{n-2} a_n \cdot \det \begin{pmatrix} -tb_1 & \cdots & -tb_{n-1} \\ & C & \end{pmatrix} + s \cdot a \cdot \det \begin{pmatrix} b_1 & \cdots & b_{n-1} \\ & C & \end{pmatrix} = \\ &= (-1)^{2n-1} a_n \cdot (-t) \cdot \det \begin{pmatrix} b_1 & \cdots & b_{n-1} \\ & C & \end{pmatrix} + s \cdot a \cdot \det \begin{pmatrix} b_1 & \cdots & b_{n-1} \\ & C & \end{pmatrix} = \\ &= (-1)^{2n} a_n \cdot t \cdot \det B + s \cdot a \cdot \det B = a_n t + a s = 1. \quad \square \end{aligned}$$

[6] Rotman, *Introduction to the Theory of Groups*, Springer, 1995, Theorem VI.4, p. 488

1.3 Exercises

1. Show that if A is upper triangular (or lower triangular), then $\det A$ is the product of the diagonal elements. Hint: If σ is a permutation different than the identity, is it true that there exist an i with $\sigma(i) > i$ and a j such that $\sigma(j) < j$?

2. Let $f : \mathbb{Z} \longrightarrow \mathbb{Z}$ be the function that maps z to $z + 1$ if z is even, and to $z - 1$ if z is odd. Is f a bijection? If so, what is f^2? What is f^3?

3. Let A, B, C be sets. Let $f : A \to B$ and $g : B \to C$ be functions such that gf is a bijection. Must f be a bijection? Must g?

4. Let A, B be $n \times n$ matrices. Prove that if $AB = I_n$, then automatically $BA = I_n$ (i.e. A has an inverse, and it coincides with B).

5. Prove that the determinant of any square matrix coincides with that of its transpose.
   Hint: Every permutation is of the form τ^{-1}, for some τ permutation; and τ is even if and only if τ^{-1} is even. Moreover, the list $(a_{1,\tau^{-1}(1)}, a_{2,\tau^{-1}(2)}, \ldots, a_{n,\tau^{-1}(n)})$ is just a reshuffling of the list $(a_{\tau(1),1}, a_{\tau(2),2}, \ldots, a_{\tau(n),n})$...

6. Prove that for each $n \geq 3$, every even permutation of $\mathcal{S}_n$ can be written as product of 3-cycles.
   Hint: First show that the product of any two flips is either the identity, or a 3-cycle, or a product of two 3-cycles. To this end, it may be useful to compute $(a, b, c)(c, d, a)$.

7. Represent the following as product of disjoint cycles:

$$(1267)(34562)(68) \qquad (123456)(1357)(163) \qquad (14)(15)(16)(17)$$

2 Abstract groups

2.1 Definition, examples, and first properties of groups

Definition 94. A **group** consists of a set G endowed with an operation $(x, y) \mapsto x \star y$ that satisfies the following axioms:

(Closure) For all x, y in G, the element $x \star y$ is in A.

(Associativity) For all x, y, z in G, $x \star (y \star z) = (x \star y) \star z$. So we may leave out brackets.

(Identity) There exists a (necessarily unique[7]) *neutral element* e in G such that for all x in G, $x \star e = x = e \star x$.

(Inverses) For every x in G, there exists a (necessarily unique[8]) *inverse* y in G such that $x \star y = e = y \star x$.

Example 95 (Bijections and Permutations)**.** Given an arbitrary set A, the set

$$G = \{f : A \longrightarrow A \text{ bijective}\}$$

is a group with respect to composition; the neutral element is the identity function. When A is finite, the bijections are called *permutations*, as we saw. The inverse of f is denoted by f^{-1}.

Example 96 $((\mathbb{Z}, +))$**.** $\mathbb{Z}$ is a group with respect to addition, with 0 as neutral element. The inverse of n is denoted by $-n$.

Example 97 $((\mathbb{Z}_m, +)$ and $(U_m, \cdot))$**.** For any positive integer m, $\mathbb{Z}_m$ is a group with respect to addition modulo m. The (additive) inverse of 0 is 0, whereas the (additive) inverse of any other r is denoted by $m - r$. Instead $\mathbb{Z}_m^* \stackrel{\text{def}}{=} \mathbb{Z}_m \setminus \{0\}$ is a group with respect to multiplication (modulo m) if and only if m is prime. If m is not prime, however, it is still true that the integers in $\{1, \ldots, m-1\}$ coprime with m form a group with respect to multiplication modulo m. This is usually called *multiplicative group of integers modulo* m, denoted by $(U_m, \cdot)$. The multiplicative inverse of an r in U_m is typically denoted by r^{-1}. Note that for any prime p, U_p is simply $\mathbb{Z}_p$ minus $\{0\}$. The neutral element of $(\mathbb{Z}_p, +)$ is 0; the neutral element of $(U_p, \cdot)$ is 1.

Example 98 $((\mathbb{Q}, +)$ and $(\mathbb{Q}^*, \cdot))$**.** The set of rational numbers

$$\mathbb{Q} \stackrel{\text{def}}{=} \left\{ \frac{a}{b} \text{ such that } a, b \in \mathbb{Z} \text{ and } b \neq 0 \right\}$$

is a group with respect to the addition defined by

$$\frac{a}{b} + \frac{c}{d} \stackrel{\text{def}}{=} \frac{ad + cb}{bd}.$$

The neutral element is $\frac{0}{1}$, or simply 0; the 'additive inverse' of $\frac{a}{b}$ is $\frac{-a}{b}$. This group is denoted by $(\mathbb{Q}, +)$.
At the same time, if $\mathbb{Q}^* \stackrel{\text{def}}{=} Q \setminus \{0\}$, then also $(\mathbb{Q}^*, \cdot)$ is a group, where $\cdot$ is the usual multiplication defined by

$$\frac{a}{b} \frac{c}{d} \stackrel{\text{def}}{=} \frac{ac}{bd}.$$

The neutral element of $(\mathbb{Q}^*, \cdot)$ is of course $\frac{1}{1}$, or simply 1; the 'multiplicative inverse' of $\frac{a}{b}$, with $a \neq 0$, is $\frac{b}{a}$.

[7] Were there two neutral elements e_1 and e_2, we would have $e_1 \star e_2 = e_1$ (because e_2 is neutral) yet also $e_1 \star e_2 = e_2$ (because e_1 is neutral), so $e_1 = e_2$.

[8] Were there two inverses y_1 and y_2 for the same element x, we would have $y_1 \star x \star y_2 = e \star y_2 = y_2$ (because y_1 is inverse) yet also $y_1 \star x \star y_2 = y_1 \star e = y_1$ (because y_2 is inverse), so $y_1 = y_2$.

Example 99 ($(\mathbb{R},+)$ and $(\mathbb{R}^*,\cdot)$)**.** $(\mathbb{R},+)$ is a group with 0 as neutral element. Moreover, $\mathbb{R}^* \stackrel{\text{def}}{=} \mathbb{R}\setminus\{0\}$ is a group with respect to multiplication, with neutral element 1.

Example 100 ($(\mathbb{C},+)$ and $(\mathbb{C}^*,\cdot)$)**.** The set of complex numbers is defined by

$$\mathbb{C} \stackrel{\text{def}}{=} \{a+bi \text{ such that } a,b\in\mathbb{R}\}.$$

Recall that i is short for "imaginary unit" and $i^2=-1$; addition in $\mathbb{C}$ is defined as

$$(a+bi)+(c+di) \stackrel{\text{def}}{=} (a+c)+(b+d)i,$$

whereas the formula for multiplying is

$$(a+bi)\cdot(c+di) \stackrel{\text{def}}{=} (ac-bd)+(ad+bc)i.$$

Then $(\mathbb{C},+)$ is a group, with $0+0i$ as neutral element. In this group, the inverse of $a+bi$ is $-a-bi$.
At the same time, $\mathbb{C}^* \stackrel{\text{def}}{=} \mathbb{C}\setminus\{0\}$ is a group with respect to multiplication, with neutral element $1=1+0i$. In this group, the inverse of $a+bi$ is $\frac{1}{a^2+b^2}(a-bi)$.

Example 101. Let $\mathbb{F}$ be one of $\mathbb{Q},\mathbb{R},\mathbb{C},\mathbb{Z}_p$ with p prime. The sets $GL_n(\mathbb{F})$ (respectively, $SL_n(\mathbb{F})$) of $n\times n$ matrices with entries in $\mathbb{F}$ and determinant nonzero (respectively, one) is a group with respect to $\times$, the operation of row-by-column multiplication.

Example 102 (Quaternions)**.** Consider on $\mathbb{R}^4\setminus\{(0,0,0,0)\}$ the following operation, introduced by Hamilton in 1843:

$$\begin{aligned}&(a,b,c,d)\star(a',b',c',d') \stackrel{\text{def}}{=}\\ &\quad\stackrel{\text{def}}{=} (aa'-bb'-cc'-dd',\ \ ab'+a'b+cd'-c'd,\ \ ac'+a'c-bd'+b'd,\ \ ad'+a'd+bc'-b'c).\end{aligned}$$

With respect to these operations, $\mathbb{R}^4\setminus\{(0,0,0,0)\}$ becomes a group, called the *quaternions*. It is easy to see that the neutral element is $(1,0,0,0)$ and the inverse of an element (a,b,c,d) is

$$(a,b,c,d)^{-1} = \frac{1}{a^2+b^2+c^2+d^2}(a,-b,-c,-d).$$

Note that $(0,1,0,0)\star(0,0,1,0)=(0,0,0,1)$, whereas $(0,0,1,0)\star(0,1,0,0)=(0,0,0,-1)$.

Example 103. Let $(G,\otimes),(H,\cdot)$ be groups. Then the cartesian product

$$G\times H \stackrel{\text{def}}{=} \{(g,h) \text{ such that } g\in G, h\in H\}$$

is a group with respect to the "entrywise" operation

$$(g_1,h_1)\star(g_2,h_2) \stackrel{\text{def}}{=} (g_1\otimes g_2, h_1\cdot h_2).$$

In fact, the neutral element is just the pair (e_G,e_H) of the respective neutral elements; and the inverse of the pair (g,h) is simply the pair

$$(\text{inverse of } g \text{ in } (G,\otimes),\ \ \text{inverse of } h \text{ in } (H,\cdot)).$$

Important notation change. We can't go on like this. We have to make a choice for the notation of the inverse of an element. Here there are two philosophies. Should we think of $\star$ more as an addition, and so denote the inverse of x by $-x$? Or should we think of $\star$ more as a multiplication, and so denote the inverse of x by x^{-1}? Both choices are perfectly reasonable. Here is the verdict. Since for abstract groups we are not requiring commutativity, and the above examples of non-commutative groups were matrices, quaternions, permutations (for which the notation is basically multiplicative), then for abstract groups we are going to choose the *multiplicative notation*, and simply drop the symbol $\star$ as implicit, exactly like we do with multiplication. In other words, we will write ab instead of $a \star b$ and speak of ab as the "product" of a and b. Consistently, the inverse of x will be denoted by x^{-1}. Some authors, again for consistence, denote the neutral element by 1; but we prefer to keep the notation e. We will also often write "let G be a group" instead of "let $(G, \cdot)$ be a group".

Remark 104. Instead of $a \star b^{-1}$ you might be tempted to write $\frac{a}{b}$. Don't do it!, because our operation might not be commutative – so if you write $\frac{a}{b}$, it's not clear whether you meant $a \star b^{-1}$ or $b^{-1} \star a$.

Proposition 105 (Cancellation)**.** *Let G be a group. For any $a, b, c \in G$, if $ab = ac$ then $b = c$.*

Proof. "Left-multiply" by a^{-1}. □

Proposition 106 ('Inverse of product')**.** *Let G be a group. For any $a, b \in G$, one has*

$$(ab)^{-1} = b^{-1}a^{-1}.$$

Proof. Since the inverse is unique, we only need to check that $(ab)(b^{-1}a^{-1}) = e = (b^{-1}a^{-1})(ab)$. This follows from associativity: for example,

$$(ab)(b^{-1}a^{-1}) = (a(bb^{-1}))a^{-1} = (ae)a^{-1} = aa^{-1} = e.$$ □

2.2 Subgroups and Lagrange's theorem

Definition 107. Let G be a group. A *subgroup* of G is a subset $H \subseteq G$ that is a group with respect to the same operation.

Proposition 108. *Let G be a group. $H \subseteq G$ is a subgroup $\iff$ H satisfies*

(SG1) for each a, b in H, the element ab^{-1} is in H.

Proof. "$\Rightarrow$" This is easy: if $a, b \in H$ group, then b^{-1} is in H, so ab^{-1} is in H.

"$\Leftarrow$" Applying (SG1) to $b = a$ we get that the neutral element $e = aa^{-1}$ is in H. But then for each b in H we get that eb^{-1} is in H, again by (SG1). So H contains the inverse of any of its elements. Finally, we should check that the operation is internal. Let x, y be arbitrary elements of H. We have just proven that $y^{-1} \in H$. Applying (SG1) to $a = x$ and $b = y^{-1}$, we get that $x((y^{-1})^{-1})$ is in H. In other words, $xy \in H$. □

Example 109. $(\mathbb{Z}, +)$ is subgroup of $(\mathbb{Q}, +)$, which is subgroup of $(\mathbb{R}, +)$, which is subgroup of $(\mathbb{C}, +)$. Similarly, $(\{-1, 1\}, \cdot)$ is subgroup of $(\mathbb{Q}^*, \cdot)$, which is subgroup of $(\mathbb{R}^*, \cdot)$, which is subgroup of $(\mathbb{C}^*, \cdot)$.

Example 110. If $(H_j)_{j \in I}$ is a family of subgroups of a group G, then their intersection is also a subgroup of G. Let's check via Proposition 108: Let $a, b \in \bigcap_{j \in I} I_j$. For each j, both a, b belong to the subgroup I_j. Hence ab^{-1} is in I_j. Since this holds for all j, ab^{-1} is in $\bigcap_{j \in I} I_j$.

Example 111. For $n \geq 2$, A_n is a subgroup of $\mathcal{S}_n$. In fact, if σ and τ are even permutations, so is σ^{-1}. But then also $\sigma^{-1}\tau$ is even. By Proposition 108, A_n is a subgroup.

Example 112. $SL_n(\mathbb{F})$ is a subgroup of $GL_n(\mathbb{F})$.

Lagrange's theorem

Lemma 113. *Let H be a subgroup of a group G. For any $a, b \in G$, there is a bijection between H and any of the following four sets:*

- *the set $aH \stackrel{\text{def}}{=} \{ah \text{ such that } h \in H\}$;*
- *the set $bH \stackrel{\text{def}}{=} \{bh \text{ such that } h \in H\}$;*
- *the set $Ha \stackrel{\text{def}}{=} \{ha \text{ such that } h \in H\}$;*
- *the set $Hb \stackrel{\text{def}}{=} \{hb \text{ such that } h \in H\}$.*

Proof. Fix a, b in G. The two functions

$$\psi : \begin{array}{ccc} aH & \longrightarrow & bH \\ x & \longmapsto & ba^{-1}x \end{array} \quad \text{and} \quad \phi : \begin{array}{ccc} bH & \longrightarrow & aH \\ x & \longmapsto & ab^{-1}x \end{array}$$

are well-defined and inverse of one another. This is true for any a, b in G; so in particular, it is true if we choose $a = e$. But $eH \stackrel{\text{def}}{=} \{eh \text{ such that } h \in H\} = H$. Thus there is a bijection between any two of H, aH, and bH. In a completely analogous way, one constructs a bijection between Ha and Hb. This holds for any a, b in G, so in particular for $a = e$; but $He = H$. Thus there is a bijection between any two of H, Ha, and Hb. □

Theorem 114 (Lagrange)**.** *Let G be a finite group with g elements. If H is a subgroup of G with h elements, then h divides g.*

Proof. If $H = G$ there is nothing to show. Otherwise, pick an element a_1 not in H. Clearly, $a_1 = a_1 e$ is in $a_1 H$. If $G = H \cup a_1 H$ stop; otherwise, pick an element a_2 not in $H \cup a_1 H$. Clearly $a_2 \in a_2 H$. And so on. We claim that $H, a_1 H, a_2 H \ldots$ are all disjoint. Let us prove the claim by contradiction. Set $a_0 \stackrel{\text{def}}{=} e$. Suppose there is an x in $a_i H \cap a_j H$, for some $i < j$. So there exist $h_i, h_j \in H$ such that $a_i h_i = x = a_j h_j$. If we set $h \stackrel{\text{def}}{=} h_i h_j^{-1}$, then $h \in H$ and

$$a_j = x h_j^{-1} = a_i h_i h_j^{-1} = a_i h \in a_i H.$$

A contradiction, a_j was chosen outside $H \cup \ldots \cup a_{j-1} H$. So the claim is proven. Since G is finite, the discovery of disjoint classes inside it eventually ends, and we can write

$$G = H \cup\ a_1 H \cup \ldots \cup\ a_t H.$$

But by Lemma 113, these $t + 1$ disjoint sets all have the same number of elements, namely, h. So $n = (t + 1)h$. Hence, h divides n. □

Remark 115. For any subgroup $H \subseteq G$, the sets of the type aH, where a ranges over all elements of G, are called *left cosets of H*; since a might be chosen in H, the subgroup H is one of them. As we saw from Lagrange's theorem, it can happen that $aH = a'H$ for $a \neq a'$. When G is finite, the total number of left cosets of H is exactly $\frac{|G|}{|H|}$. Similarly, the sets of the type Ha are called *right cosets of H*, and with a completely analogous argument, one can show that when G is finite, the total number of right cosets is exactly $\frac{|G|}{|H|}$.

2.3 Period and cyclic subgroups

Definition 116. Let a be an element of a group G. Let $z \in \mathbb{Z}$. We define

$$a^z \stackrel{\text{def}}{=} \begin{cases} aa \cdots a \ \text{(z times)}, & \text{if } z > 0, \\ e & \text{if } z = 0, \\ a^{-1} \cdots a^{-1} \ \text{(-z times)}, & \text{if } z < 0. \end{cases}$$

It is clear that the inverse of a^n is $(a^{-1})^n$, which by definition is a^{-n}.

Proposition 117 (Power properties)**.** *Let G be a group. Let a be any element of G. For any integers z, w one has $a^z a^w = a^{z+w}$ and $(a^z)^w = a^{zw}$.*

Proof. Left as exercise. (Hint: Do the case $w > 0$ and $z > 0$ first. Then do all other cases.) □

Definition 118 (period/order)**.** Let G be a group. Let a be an element of G. The *period of a*, also known in many textbooks as the *order of a*, is

$$\pi(a) \stackrel{\text{def}}{=} \begin{cases} +\infty & \text{if all powers of } a \text{ are distinct,} \\ t & \text{if } t \text{ is the smallest positive integer for which } a^t = e. \end{cases}$$

Remark 119. The two cases above are mutually exclusive, and cover all possibilities: If the powers of a are not all distinct, then $a^z = a^w$ for some $z > w$, whence using cancellation (Proposition 105) we get that $a^{z-w} = e$. So the set of integers $n > 0$ for which $a^n = e$ is non empty. Conversely, if $a^t = e$, then not all powers of a are distincts, since also $a^0 = e$.

Example 120. Consider $a = (1, 2, 3, 4)$ in $\mathcal{S}_5$. Since $a \neq e$, $a^2 \neq e$, $a^3 \neq e$, but $a^4 = e$, then $\pi(a) = 4$. More generally, the period of any k-cycle is k.

Example 121. In any group G, the neutral element e is the only element of period 1.

Example 122. In the group $(\mathbb{Q}^*, \cdot)$, except for ± 1, all elements have infinite period.

Lemma 123. *Let a be an element of a group G. Let $k \in \mathbb{N}$. Then*

$$a^k = e \iff k \text{ is a multiple of } \pi(a).$$

Proof. "⇐" Easy: if $k = m\pi(a)$, then $a^k = (a^{\pi(a)})^m = e^m = e$.
"⇒" Let us perform a Euclidean division $k = q\pi(a) + r$ with $0 \leq r < \pi(a)$.
If $r = 0$ then k is a multiple of $\pi(a)$ and we are done. If $r > 0$, we have

$$e = a^k = a^{q\pi(a)+r} = (a^{\pi(a)})^q a^r = e^q a^r = a^r,$$

a contradiction with the definition of period: r is smaller than $\pi(a)$. □

Proposition 124. *If an element a of a group G has period m, then for all $k \in \mathbb{N}$ the element a^k has period $\frac{m}{\gcd(m,k)}$.*

Proof. Exercise. Hint: Set $m' \stackrel{\text{def}}{=} \frac{m}{\gcd(d,k)}$ and $k' \stackrel{\text{def}}{=} \frac{k}{\gcd(m,k)}$, and show that $\gcd(m', k') = 1$ and that $km' = mk'$. Use this to show $(a^k)^{m'} = e$. Now let t be any integer such that $(a^k)^t = a^{kt} = e$; you want to show that t is a multiple of m'. But since $a^{kt} = e$, by Lemma 123 kt is a multiple of m: so write the identity $kt = mq$, for some $q \in \mathbb{N}$, and divide this identity by $\gcd(m, k)$. You get $k't = m'q$. But k' has no common divisor with m'... □

Remark 125. We started this Section very fast, thanks also to the multiplicative notation. However, we are often going to apply these results to groups where the operation is a sum. It could be confusing to do the translation, so here is some guidance. In a group like $(\mathbb{Z}_m, +)$:

- the "positive powers" a^n of an element a are obtained by operating a with itself its times, so they are what we would call the *multiples* of a; and in general, the powers a^z, with $z \in \mathbb{Z}$, of an a in $\mathbb{Z}_m$, are simply its integer multiples $\{za : z \in \mathbb{Z}\}$;
- the "power properties" simply become "for all a in $\mathbb{Z}_m$ and for all z, w in $\mathbb{Z}$, $za + wa = (z + w)a$ and $z(wa) = (zw)a$";

- the "period of a" is the smallest positive integer t for which $ta = 0$ (or $+\infty$, if there is no such integer); in general, $ka = 0$ if and only if k is a multiple of the period of a;
- If an element a has period m, and $k \in \mathbb{N}$, then ka has period $\frac{m}{\gcd(m,k)}$.

The same type of "translation" holds for groups like $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$ and $(\mathbb{C}, +)$.

Definition 126 ($< a >$)**.** Let X be a subset of a group G. We denote by $< X >$ the smallest subgroup of G containing X (or equivalently, the intersection of all subgroups containing X). If X consists of a single element a, we write $< a >$ instead of $< \{a\} >$.

Lemma 127. *For any group G and for any element $a \in G$,*

$$< a >= \{a^z \text{ such that } z \in \mathbb{Z}\}$$

is a subgroup of G with exactly $\pi(a)$ elements. If $\pi(a)$ is finite,

$$< a >= \{e, a, a^2, a^3, \dots, a^{\pi(a)-1}\}.$$

Proof. For the first equality: if a^z and a^w are two elements of $\{a^z \text{ such that } z \in \mathbb{Z}\}$, so is $a^z(a^w)^{-1} = a^z a^{-w} = a^{z-w}$. Thus by Proposition 108, $\{a^z \text{ such that } z \in \mathbb{Z}\}$ is a subgroup. It contains $a^1 = a$. Also, any subgroup of G contaning a must also contains all its powers. Hence, $\{a^z \text{ such that } z \in \mathbb{Z}\}$ is the smallest subgroup containing a, which is what we denoted by $< a >$. If $\pi(a)$ is infinite, then all powers of a are distinct, so $< a >$ has infinitely many elements. Now suppose $\pi(a)$ is finite, and let us prove the second identity. The inclusion $\supseteq$ is obvious; $\subseteq$ follows from the fact that if $z = q \cdot \pi(a) + r$ (Euclidean division), then $a^z = (a^{\pi(a)})^q \cdot a^r = a^r$. □

Remark 128. In general, $< a, b > \supsetneq \{a^z\, b^t \text{ such that } z, t \in \mathbb{Z}\}$, because the left-hand side contains also elements of the type $ab^{-1}a^5a^{-3}b^7$, which we do not know how to rearrange. This problem would be solved if we knew in advance that a and b *commute* (that is, $ab = ba$): Then we could rewrite $ab^{-1}a^5b^{-7}a^{-3}b^5 = a^3b^{-3}$ and we would have $< a, b >= \{a^z\, b^t \text{ such that } z, t \in \mathbb{Z}\}$. In the Exercises you are asked to prove the following fact: if in a group G there are elements $a_1, \dots, a_n$ any two of which commute, i.e. $a_i a_j = a_j a_i$ for all i, j, then

$$< a_1, \dots, a_n >= \{a_1^{z_1}\, a_2^{z_2} \cdots a_n^{z_n} \text{ such that } z_i \in \mathbb{Z}\}.$$

Proposition 129. *Let G be a finite group with n elements. Then for all $a \in G$, one has $a^n = e$.*

Proof. For each a, the subgroup $< a >$ has cardinality $\pi(a)$ by Lemma 127. By Lagrange's theorem 114 the integer $\pi(a)$ divides n. But then by Lemma 123 one has $a^n = e$. □

Theorem 130 (Fermat's little theorem, 1640)**.** *Let p be any prime number. In $\mathbb{Z}$, for any a one has $a^p \equiv a \mod p$. Moreover, for any $b \in \mathbb{Z}$ such that $\gcd(b, p) = 1$, one has $b^{p-1} \equiv 1$.*

Proof. Let us prove the second statement first. Let b be an integer that is not a multiple of p. If we divide b by p, and write $b = qp + r$, with $0 \leq r \leq p - 1$, it is clear that r cannot be a multiple of p (or else also b would be). So we can view r as an element of $U_p = \mathbb{Z}_p \setminus \{0\}$, which has $p - 1$ elements. By Proposition 129, we know that $r^{p-1} = 1$ in U_p. Translating it for $\mathbb{Z}$, this means that $r^{p-1} \equiv 1 \mod p$. Since $b \equiv r \mod p$, by Lemma 54 we conclude that

$$b^{p-1} \equiv 1 \mod p. \tag{4}$$

This proves the second statement. Moreover, if we multiply Equation (4) by b, by Lemma 54 we get that $b^p \equiv b \mod p$. So it only remains to show that if a is a multiple of p, then $a^p \equiv a \mod p$. But this is obvious: If p divides a, it divides any power of it, so the statement left to prove boils down to $0 \equiv 0 \mod p$. □

Remark 131. Fermat's little theorem can also be proven by induction on a: See the exercises at the end of Chapter 0.

Cyclic subgroups

Definition 132 (Cyclic groups)**.** A group G is called *cyclic* if there exists a in G such that $G =< a >$. In this case we say that a is *a generator for* G.

Lemma 133. *Let G be a finite group with n elements. G is cyclic if and only if it contains at least one element of period n.*

Proof. By Lemma 127, for any a in G, the subgroup $< a >$ has $\pi(a)$ elements. □

Definition 134 (Finitely generated groups)**.** We say that a group G is *finitely generated*, if $G =< X >$ for some finite subset X of G.

Obviously, all cyclic groups are finitely generated. All finite groups are also finitely generated, since one could choose $X = G$.

Example 135. $\mathbb{Z}_4$ is cyclic: a generator is 1. (Another possible generator is 3. Instead, 2 won't do, because the smallest subgroup containing 2 is $\{0, 2\}$, not the whole $\mathbb{Z}_4$).
Similarly, U_5 is cyclic. A possible generator is 2. (Another possible generator is 3. Instead, 4 won't do, because the smallest subgroup containing 4 is $\{1, 4\}$.) Note that $(\mathbb{Z}_4, +)$ and $(U_5, \cdot)$ are very similar!

Non-Example 136. $\mathbb{Z}_2 \times \mathbb{Z}_2$ is finitely generated, but not cyclic. The smallest subgroup containing $e = (0, 0)$ is $\{e\}$; moreover, for any $x \neq e$ in $\mathbb{Z}_2 \times \mathbb{Z}_2$, one has $x + x = e$, so the smallest subgroup containing x is $\{e, x\}$. Similarly, $U_8 = \{1, 3, 5, 7\}$ is not cyclic. The smallest subgroup containing 1 is $\{1\}$; moreover, for any $x \neq 1$ in $\mathbb{Z}_8$, one has $x \cdot x = 1$, so the smallest subgroup containing x is $\{1, x\}$. Note that $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$ and $(U_8, \cdot)$ are very similar!

Proposition 137. *Every subgroup of $(\mathbb{Z}, +)$ is cyclic, of the form $< m >$ for some $m \in \mathbb{N}$.*

Proof. Let S be a subgroup of $(\mathbb{Z}, +)$. If $S = \{0\}$ then $S =< 0 >$. Otherwise, let m be the smallest positive integer in S. Clearly $< m > \subseteq S$. Let us prove the opposite inclusion: For any s is in S, the Euclidean division $x = q \cdot m + r$, with $0 \leq r < m$, tells us that $r = x - qm$ is in S, because in $(\mathbb{Z}, +)$ this is how we denote the operation $x \cdot (m^q)^{-1}$. Thus if $r > 0$ we have a contradiction with how m was chosen. So $r = 0$, which means that $x = qm$. Thus $S =< m >$. □

Proposition 138. *$(\mathbb{Q}, +)$ and $(\mathbb{Q}^*, \cdot)$ are neither cyclic nor finitely generated.*

Proof. See the Exercises. □

Next comes a partial converse to Lagrange's theorem. The dream would be to prove that "if an integer m divides the size of a group G, then G has a subgroup with m elements"; but this dream is impossible to reach, because there are counterexamples!

Non-Example 139. The set A_4 of even permutations of four elements has size 12. Yet one can see by inspection that it has no subgroup of size 6.

However, it turns out that a converse statement is true (with an extra bonus!, cyclicity) when the divisor of the size of G is prime. The next proof is taken from James H. McKay, *Another Proof of Cauchy's Theorem*, American Mathematical Monthly 66 (1959), page 119.

Theorem 140 (Cauchy)**.** *If a prime p divides the size of a finite group G, then G has a cyclic subgroup with p elements.*

Proof. We need to show that some element of G has period p. We are going for something stronger, namely, that the number of period-p elements is of the form $kp-1$ for some positive integer k. In particular, this number is at least $p-1$, which is positive. Say G has n elements. Look at the set

$$S \stackrel{\text{def}}{=} \{(x_1, x_2, \ldots, x_p) \text{ such that } x_1 \cdot x_2 \cdot \ldots \cdot x_p = e\}.$$

Our first claim is that if $(x_1, x_2, \ldots, x_p)$ is in S, so is $(x_p, x_1, x_2, \ldots, x_{p-1})$. In fact, if

$$x_1 \cdot x_2 \cdot \ldots \cdot x_p = e,$$

left-multiplying by x_p and right-multiplying by x_p^{-1} we get

$$x_p x_1 \cdot x_2 \cdot \ldots \cdot x_{p-1} = e.$$

So the first claim is proven. Note that re-applying the first claim over and over, once we know that $(x_p, x_1, x_2, \ldots, x_{p-1})$ is in S, then also $(x_{p-1}, x_p, x_1, \ldots, x_{p-2})$ is in S; but then automatically also $(x_{p-2}, x_{p-1}, x_p, x_1, \ldots, x_{p-3})$ is in S; and so on. Eventually, what we have proven is that S is closed under "cyclic shifting" of the components of its points.
Next, we compute the size of S. This is easy: Once we freely choose $x_1, \ldots, x_{p-1}$ in $\{1, \ldots, n\}$, there is a unique x_p such that $x_1 \cdot x_2 \cdot \ldots \cdot x_p = e$; so the set S has cardinality n^{p-1}.
Next, we partition S into equivalence classes, as follows. If all components of a p-tuple are equal, its equivalence class shall consist of only 1 element. If instead we have a $(x_1, \ldots, x_p)$ with $x_i \neq x_j$ for some i, j, then the equivalence class of $(x_1, \ldots, x_p)$ shall consist of the p (different!) elements

$$(x_1, x_2, \ldots, x_{p-1}, x_p),\ (x_2, \ldots, x_{p-1}, x_p, x_1),\ \ldots, (x_p, x_1, x_2, \ldots, x_{p-1}).$$

So let a be the number of equivalence classes with only one member, and b be the number of equivalence classes with p members. Since

$$n^{p-1} = a \cdot 1 + b \cdot p,$$

and n is a multiple of p, we obtain that a is also a multiple of p. But by definition, a counts exactly the elements x such that $(x, x, \ldots, x)$ is in S, which means that $x^p = e$. So we can conclude that the equation $x^p = e$ has a number of solutions in G that is a multiple of p. Write this number as kp. One of the solutions is the neutral element, because $e^p = e$. So there are exactly $kp-1$ elements y different than the identity, such that $y^p = e$. The period of any such y must divide p, which is a prime number. So the period of any such y is exactly p. □

Remark 141. There is a theory that allows you to say more. It's called *Sylow theory*, after a Norwegian high school teacher called Ludwig Sylow. In 1872, he proved three important theorems on this topic. The first Sylow theorem says "If the size of a finite group G is a multiple of p^m, with p a prime $m \in \mathbb{N}$, then G has a subgroup with p^m elements". Any such subgroup is called "a p-subgroup", and is not necessarily cyclic: For example, $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ has no cyclic subgroups with 4 elements. The Second and Third Sylow theorems concern the number of distinct p-subgroups. They imply for example that "if m is the *largest* integer such that p^m divides the size of a group G, with p a prime, then G has a *unique* subgroup with p^m elements". Again $G = \mathbb{Z}_2 \times \mathbb{Z}_2$ shows that this "maximal p-subgroup" is not cyclic in general.

2.4 Group homomorphisms

Definition 142. A function $f : G \longrightarrow H$ between two groups $(G, \star)$ and $(H, \cdot)$ is called *group homomorphism* if

$$f(a \star b) = f(a) \cdot f(b).$$

A bijective group homomorphism is called *group isomorphism*. A group G is *isomorphic* to a group H if there exists a group isomorphism from G to H.

Example 143. The inclusion $\iota : \mathbb{Z} \to \mathbb{Q}$, $\iota(z) = z$ is an injective (not surjective) group homomorphism.

Example 144. The map $f(0,0) = 1$, $f(0,1) = 3$, $f(1,0) = 5$, $f(1,1) = 7$ is an isomorphism between $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$ and $(U_8, \cdot)$.

Proposition 145. *The inverse of a bijective group homomorphism is a (bijective) group homomorphism. In particular, being isomorphic is an equivalence relation on the class of all groups.*

Proof. Let $f : G \longrightarrow H$ be an isomorphism between two groups $(G, \star)$ and $(H, \cdot)$. Let f^{-1} be the inverse function. Let h_1, h_2 be elements of H. Since f is a group homomorphism,

$$h_1 \cdot h_2 = (ff^{-1}(h_1)) \ \cdot \ (ff^{-1}(h_2)) \overset{!}{=} f\left(f^{-1}(h_1) \star f^{-1}(h_2)\right).$$

Applying f^{-1} to the previous equality, we obtain that $f^{-1}(h_1 \cdot h_2) = f^{-1}(h_1) \star f^{-1}(h_2)$. □

Proposition 146. *Any group homomorphism maps the identity to the identity.*
Moreover, the inverse of the image of an element is the image of its inverse.

Proof. Let $f : G \longrightarrow H$ be a group homomorphism from a group $(G, \star)$ to a group $(H, \cdot)$. By the cancellation property (Prop. 105), from $f(e_G) \cdot f(e_G) = f(e_G) = e_H \cdot f(e_G)$ one gets $f(e_G) = e_H$. Similarly from $f(a) \cdot f(a^{-1}) = f(e_G) = e_H = f(a) \cdot [f(a)]^{-1}$ one gets $f(a^{-1}) = [f(a)]^{-1}$. □

Proposition 147. *The image of any group homomorphism is a subgroup of the codomain. Moreover, the* kernel *of any group homomorphism, defined as the set of elements mapped to the identity, is a subgroup of the domain.*
Finally, a group homomorphism is injective if and only if its kernel consists only of the identity.

Proof. Let $f : G \longrightarrow H$ be a group homomorphism from a group $(G, \star)$ to a group $(H, \cdot)$. Let $h_1 = f(g_1)$ and $h_2 = f(g_2)$ be elements of $\operatorname{Im} f$. So by Proposition 146, applied at the mark,

$$f(g_1 * (g_2)^{-1}) = f(g_1) \cdot f((g_2)^{-1}) \overset{!}{=} f(g_1) \cdot [f(g_2)]^{-1} = h_1[h_2]^{-1}.$$

So $h_1[h_2]^{-1}$ is also in $\operatorname{Im} f$. By Proposition 108, $\operatorname{Im} f$ is a subgroup of H. Next, consider

$$\ker f \overset{\text{def}}{=} \{g \in G \ : \ f(g) = e_H\}.$$

Let g_3, g_4 be in G such that $f(g_3) = f(g_4) = e_H$. Applying Proposition 146 at the mark,

$$f(g_3 * g_4^{-1}) = f(g_3) \cdot f(g_4^{-1}) \overset{!}{=} f(g_3) \cdot [f(g_4)]^{-1} = e_H \cdot e_H^{-1} = e_H.$$

So by Proposition 108, $\ker f$ is a subgroup of G. It remains to show that f is injective if and only if $\ker f = \{e_G\}$. The direction $\Rightarrow$ is easy: by Proposition 146 we know that $f(e_G) = f(e_H)$, so if f is injective, no other element can be mapped to e_H. As for the converse direction $\Leftarrow$, we reason as follows: Suppose $f(x) = f(y)$. Left-multiplying by the inverse of $f(x)$ and applying Proposition 146 at the mark, we get

$$e_H = [f(x)]^{-1} \cdot f(y) = f(x^{-1}) \cdot f(y) = f(x^{-1}y).$$

So xy^{-1} belongs to $\ker f$. But $\ker f = \{e_G\}$, so $xy^{-1} = e_G$. Which means $x = y$. □

Example 148. Consider the groups $(\mathbb{R}^*, \cdot)$ and $(\mathbb{R}, +)$. The function $f : \mathbb{R}^* \longrightarrow \mathbb{R}$ defined by $f(x) = \log(x^2)$ is a surjective group homomorphism: $f(xy) = \log(xy)^2 = 2\log x + 2\log y = f(x) + f(y)$. Since the neutral element of $(\mathbb{R}, +)$ is 0, $\ker f = \{x : \log x^2 = 0\} = \{x : x^2 = 1\} = \{-1, +1\}$. Note that for any group homomorphism $F : (\mathbb{R}^*, \cdot) \longrightarrow (\mathbb{R}, +)$, the kernel of F has to contain both -1 and 1, because from $(-1)^2 = 1$ and Proposition 146 it follows that $F(-1) + F(-1) = 0$, whence $F(-1) = 0$. In particular, $(\mathbb{R}^*, \cdot)$ is **not** isomorphic to $(\mathbb{R}, +)$.

Example 149. Let n be any positive integer. The map

$$\begin{array}{rccc} g_n : & (\mathbb{R}, +) & \longrightarrow & (\mathbb{R}_{>0}, \cdot) \\ & x & \longmapsto & e^{nx} \end{array}$$

is a group isomorphism: $g_n(x + y) = e^{n(x+y)} = e^{nx} \cdot e^{ny} = g_n(x) \cdot g_n f(y)$.

Example 150. Let $\mathbb{F}$ be any of $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}_p$, with p prime. Let $\mathbb{F}^* \stackrel{\text{def}}{=} \mathbb{F} \setminus \{0\}$. Consider

$$\begin{array}{rccc} f : & (GL_n(\mathbb{F}), \times) & \longrightarrow & (\mathbb{F}^*, \cdot) \\ & A & \longmapsto & \det A. \end{array}$$

By Cauchy–Binet's theorem, f is a group homomorphism. Its kernel is $SL_n(\mathbb{F})$. Its image is all of $\mathbb{F}^*$: In fact, for any x in $\mathbb{F}^*$, the matrix M with $m_{1,1} = x$, $m_{i,i} = 1$ for all $i > 1$, and $m_{i,j} = 0$ for all $i \neq j$, has determinant x.

Example 151 (Orthogonal matrices and rotations)**.** The *orthogonal group* is defined as

$$O_n(\mathbb{F}) \stackrel{\text{def}}{=} \{n \times n \text{ matrices A with entries in } \mathbb{F} \text{ such that } A\,A^T = I\},$$

with $\mathbb{F}$ as in the previous example. By Cauchy–Binet's theorem, $1 = \det A \det A^T = (\det A)^2$, so matrices in $O_n(\mathbb{F})$ have determinant ± 1 and $O_n(\mathbb{F})$ is also a subgroup of $GL_n(\mathbb{K})$. Set

$$\begin{array}{rccc} f : & (O_n(\mathbb{F}), \times) & \longrightarrow & (\{-1, 1\}, \cdot) \\ & A & \longmapsto & \det A. \end{array}$$

The kernel is called the *rotation group* $SO_n(\mathbb{F})$. By definition, $SO_n(\mathbb{F}) = O_n(\mathbb{F}) \cap SL_n(\mathbb{F})$.

Example 152. Let $(E, +)$ be the subgroup of $\mathbb{Z}$ formed by the even numbers (i.e. $E =< 2 >$). Let $f : (\mathbb{Z}, +) \to (E, +)$ be the map $f(x) = 2x$. This map is a bijective group homomorphism.

Example 153. Let $f : (\mathbb{Z}_{10}, +) \to (\mathbb{Z}_{10}, +)$ be the map defined by $f(x) = 2x$. This map is a group homomorphism. It is not injective: $f(0) = f(5) = 0$. It is not surjective, because its image corresponds to the subgroup of $\mathbb{Z}_{10}$ generated by 2.

Example 154. Let $f : (\mathbb{Z}_{11}, +) \to (\mathbb{Z}_{11}, +)$ be the map defined by $f(x) = 2x$. This map is a bijective group homomorphism.

Non-Example 155. Let $f : (\mathbb{Z}_{10}, +) \to (\mathbb{Z}_{11}, +)$ be the function defined by $f(x) = 2x$. This is not a group homomorphism! In fact, $f(5 + 7) = f(2) = 4$, whereas $f(5) + f(7) = 10 + 3 = 2$.

Example 156. Let $f : (\mathbb{Z}_{10}, +) \to (\mathbb{Z}_{12}, +)$ be the function defined by $f(x) = 2x$. This is a group homomorphism! It is not injective, but it is surjective.

Example 157. Let $f : (\mathbb{Z}_{10}, +) \to (\mathbb{Z}_{22}, +)$ be the function defined by $f(x) = 2x$. This is a group homomorphism! It is injective, but not surjective.

Proposition 158. *For any group homomorphism $f : G \longrightarrow H$, and for each $x \in G$, the period of $f(x)$ divides the period of x.*

Proof. Let t be the period of x. From $x^t = e_G$, we get $(f(x))^t = f(x^t) = f(e_G)$. But by Proposition 146 we know that $f(e_G) = e_L$. So $(f(x))^t = e_L$. By Lemma 123, this means that the period of $f(x)$ divides t. □

Corollary 159. *If $\gcd(m, n) = 1$, the only group homomorphism between $(\mathbb{Z}_m, +)$ and $(\mathbb{Z}_n, +)$ is the zero homomorphism.*

Proof. Let $f : \mathbb{Z}_m \longrightarrow \mathbb{Z}_n$ be an arbitrary group homomorphism and set $y \stackrel{\text{def}}{=} f(\overline{1})$. By Proposition 158, the period of y divides $\pi(1) = m$. By Proposition 129, the period of y divides also n. So, since $\gcd(m, n) = 1$, the period of y must be 1. So $y = 0$. □

Remark 160. It is a nice exercise to prove that the number of distinct group homomorphisms between $\mathbb{Z}_m$ and $\mathbb{Z}_n$ is $\gcd(m, n)$. Hints: (1) show that every group homomorphism $f : \mathbb{Z}_m \to \mathbb{Z}_n$ is of the form $f(z) = f(1) \cdot z$, with $0 \leq f(1) \leq n - 1$; (2) show that any function $f : \mathbb{Z}_m \to \mathbb{Z}_n$ defined by $f(z) = x \cdot z$, with $0 \leq x \leq n - 1$, is a group homomorphism if and only

$$x \cdot m \equiv 0 \mod n.$$

(3) But the above equation in x has exactly $\gcd(m, n)$ solutions, namely,

$$x = \frac{n}{\gcd(m, n)} k, \qquad \text{for } k \in \{0, 1, \ldots, \gcd(m, n) - 1\}.$$

Theorem 161. *Every infinite cyclic group is isomorphic to $(\mathbb{Z}, +)$.*
Every cyclic group with m elements is isomorphic to $(\mathbb{Z}_m, +)$.

Proof. Suppose $G =< a >$. If $\pi(a) = \infty$, then consider the function ϕ from $(\mathbb{Z}, +)$ to $(G, \star)$ that sends z to a^z. This is a surjective group homomorphism; injectivity follows from the fact that all powers of a are distinct, since $\pi(a) = \infty$. Thus $(\mathbb{Z}, +)$ and $(G, \star)$ are isomorphic. If instead $\pi(a) = m$, then consider the function ψ from $(\mathbb{Z}_m, +)$ to $(G, \star)$ that sends z to a^z. Again, it is easy to see that this is a surjective group homomorphism; moreover, by Lemma 123, $\psi(z) = e$ if and only if $z = 0$. So $\ker \psi = \{0\} = \{e_{\mathbb{Z}_m}\}$. This means ψ is injective. □

We conclude this section with a result that shows how important permutation groups are.

Theorem 162 (Cayley)**.** *Every group $(G, \cdot)$ is isomorphic to a subgroup of the group of all bijective functions from G to G. In particular, every group with n elements is (isomorphic to) some subgroup of $\mathcal{S}_n$.*

Proof. Set $\Gamma \stackrel{\text{def}}{=} \{\sigma : G \to G \text{ bijective}\}$. We have seen in Example 95 that Γ is a group with respect to composition. Now given an element $a \in G$, we may define a function γ_a by

$$\begin{array}{rccc} \gamma_a : & G & \longrightarrow & G \\ & x & \longmapsto & ax. \end{array}$$

Note that γ_a is injective (because $ax = ay$ implies $x = y$ by Prop. 105) and surjective (because every $g \in G$ can be written as $g = a(a^{-1}g) = \gamma_a(a^{-1}g)$.) So γ_a is an element of Γ. In fact, the inverse of the bijection γ_a is the map $\gamma_{a^{-1}}$, which sends x to $a^{-1}x$. Now set

$$T \stackrel{\text{def}}{=} \{\gamma_a \text{ such that } a \in G\}.$$

We claim that T is a **subgroup**, and not just a subset, of Γ. In fact, for any two elements γ_a, γ_b of T, the function $\gamma_a \circ \gamma_b$ is simply γ_{ab}; and so $\gamma_a \circ (\gamma_b)^{-1}$ is simply $\gamma_{ab^{-1}}$, which is in T. To complete the proof, it is easy to see that the function that sends a to γ_a is the desired bijection from G to T. □

2.5 Exercises

1. Prove that $(\mathbb{Q}, +)$ and $(\mathbb{Q}^*, \cdot)$ are not cyclic.
2. Prove the following fact: if in a group G there are elements $a_1, \ldots, a_n$ any two of which commute, i.e. $a_i a_j = a_j a_i$ for all i, j, then
$$< a_1, \ldots, a_n > = \{a_1^{z_1} a_2^{z_2} \cdots a_n^{z_n} \text{ such that } z_i \in \mathbb{Z}\}.$$
3. Prove that $(\mathbb{Q}, +)$ and $(\mathbb{Q}^*, \cdot)$ are not finitely generated.
4. In $GL_2(\mathbb{R})$, what is the period of the element $a = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$?
5. Write down an explicit group isomorphism between $(\mathbb{Z}_4, +)$ and $(U_5, \cdot)$.
6. Let x, y be two group elements such that $x^{2018} = y^{2019}$ and $xyx = yxy$. Prove that $x = y = e$.
7. For any groups G, H, K, prove that $G \times (H \times K)$ is isomorphic to $(G \times H) \times K$.
8. Prove Remark 160.
9. Prove that if a subgroup of $\mathcal{S}_5$ contains $(1, \ldots, 5)$ and $(1, 2)$, then it is the whole $\mathcal{S}_5$.

3 Normal subgroups, quotients, and Abelian groups

3.1 Normal subgroups

Definition 163 (Normal)**.** A subgroup H of G is called *normal (in G)* if for each $g \in G$, for each $h \in H$, $ghg^{-1} \in H$.

Non-Example 164. Consider in $G = GL_2(\mathbb{R})$ the subgroup $H = UT_2(\mathbb{R})$ of upper triangular matrices with nonzero determinant. This H is a subgroup that is not normal. In fact, choosing

$$h = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } g = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \text{one has } ghg^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \notin UT_2(\mathbb{R}).$$

Example 165. Given any group G, the subgroups $H = \{0\}$ and $H = G$ are always normal.

Example 166. Given any homomorphism $f : G \to H$, the subgroup $\ker f$ is always normal in G. In fact, if $f(k) = e_H$, then for every g in G we have

$$f(gkg^{-1}) = f(g)f(k)f(g^{-1}) = f(g)e_H f(g^{-1}) = f(g)f(g^{-1}) = f(gg^{-1}) = f(e_G),$$

which is equal to e_H by Proposition 146. So $gkg^{-1} \in \ker f$.

Remark 167. If the elements g and h commute, then $ghg^{-1} = hgg^{-1} = h$. So certainly in groups where the operation is commutative, like $(\mathbb{Z}, +)$ or $(\mathbb{Q}^*, \cdot)$, every subgroup is normal. However, there exist groups G where every subgroup is normal, yet the operation is not commutative. One such group, the 8-element *quaternion group*, is explained in the Exercises.

Lemma 168. *Let H be a subgroup of a group G. The following are equivalent:*
(1) *H is normal.*
(2) *For each g in G, for any $h \in H$, there is a k in H such that $gh = kg$.*
(3) *For each g in G, $gH \stackrel{\text{def}}{=} \{gh \text{ such that } h \in H\}$ and $Hg \stackrel{\text{def}}{=} \{hg \text{ such that } g \in G\}$ coincide.*
(4) *For each $a, b \in G$, $ab^{-1} \in H$ if and only if $a^{-1}b \in H$.*

Proof. (1) $\Rightarrow$ (2): We know that $ghg^{-1} \in H$. Setting $k \stackrel{\text{def}}{=} ghg^{-1}$, we have $kg = gh$.
(2) $\Rightarrow$ (3): $gH \subseteq Hg$, because any element of the form gh can also be rewritten in the form kg for some $k \in H$. Symmetrically, $Hg \subseteq gH$. So $gH = Hg$.
(3) $\Rightarrow$ (4): Suppose $ab^{-1} \in H$. Set $h \stackrel{\text{def}}{=} ab^{-1}$. Then $a = ab^{-1}b = hb \in Hb$. Since by assumption $Hb \subseteq bH$, it follows that $a = bk$ for some $k \in H$. So $a^{-1}b = (bk)^{-1}b = k^{-1}b^{-1}b = k^{-1}$ is in H. The converse implication is similar: if $k = b^{-1}a \in H$, then $a = bk \in bH \subseteq Hb$, so we can find $h \in H$ such that $a = hb$. Hence, $ab^{-1} = h \in H$.
(4) $\Rightarrow$ (1): For every h in H and for every G in G, we want to show that ghg^{-1} is in H. In other words, if we set $a \stackrel{\text{def}}{=} gh$ and $b \stackrel{\text{def}}{=} g$, we want to show that ab^{-1} is in H. But by the assumption, this is equivalent to proving that $a^{-1}b \in H$. But $a^{-1}b = h^{-1}g^{-1}g = h^{-1}$. □

Proposition 169. *If a subgroup H of G contains half of the elements of G, then H is normal.*

Proof. Let x be an element of G that is not in H. The set xH is disjoint from H and has the same number of elements of H (cf. Lemma 113), so xH is simply the complement of H. The same applies to Hx. But then $xH = Hx$, so by Lemma 168 H is normal. □

Corollary 170. *The set $A_n \stackrel{\text{def}}{=} \{\text{even permutations}\}$ is a normal subgroup of $\mathcal{S}_n$.*

Proof. This follows straightforwardly either from the definition, or from the fact that A_n has half the elements of $\mathcal{S}_n$ (cf. Proposition 83.) □

Proposition 171. *Let H, K be two normal subgroups of a group G. If $H \cap K = (e)$, then the smallest subgroup of G containing both H and K is isomorphic to $H \times K$.*

Proof. For any h in H and for any k in K, consider $hk(kh)^{-1}$. Since we can write it as $(hkh^{-1})k^{-1}$, and $hkh^{-1} \in K$ by the normality of K in G, we see that $hk(kh)^{-1}$ is a product of two elements in K; so it is itself in K. But at the same time $hk(kh)^{-1} = h(kh^{-1}k^{-1})$, which by the normality of H is a product of two elements of H; so $hk(kh)^{-1}$ is in H. Hence $hk(kh)^{-1}$ is in $H \cap K$, which by assumption consists only of the identity. But $hk(kh)^{-1} = e$ means $hk = kh$. So we have proven that any element of H commutes with any element of K. Now let us set

$$\begin{array}{rccc} \varphi: & H \times K & \longrightarrow & G \\ & (h,k) & \longmapsto & hk. \end{array}$$

Is it a group homomorphism? Indeed, since h' commutes with k,

$$\varphi(h,k) \cdot \varphi(h',k') = hk \cdot h'k' = hh' \cdot kk' = \varphi(hh', kk') = \varphi((h,k) \cdot (h',k')).$$

Let us check that φ is injective. Assume $hk = e$. Then $h^{-1} = h^{-1}e = h^{-1}hk = k$. Since $h^{-1} \in H$, we have that $k = h^{-1} \in H \cap K$, which implies $k = e$, so $h = e$. Hence φ is injective. So $H \times K$ is isomorphic to $\operatorname{Im}\varphi = \{hk \text{ such that } h \in H, k \in K\}$.

To conclude our proof, it remains to show that $\operatorname{Im}\varphi$ is the smallest subgroup containing H and K. Indeed $\operatorname{Im}\varphi$ contains any element h of H, which can be written as $h = he$. Symmetrically, it contains any element of K, by writing it as $k = ek$. Finally, any subgroup containing H and K must contain the products of their elements; so it must contain $\operatorname{Im}\varphi$. □

3.2 Quotients and the First Isomorphism Theorem

Definition 172. Let H be a normal subgroup of G. Let $\sim_H$ be the associated relation of equivalence

$$a \sim b \overset{\text{def}}{\iff} a^{-1}b \in H.$$

Because of normality, we can equivalently write "$\overset{\text{def}}{\iff} ab^{-1} \in H$", of course. The *quotient*

$$G/H \overset{\text{def}}{=} \{\overline{g} \text{ such that } g \in G\},$$

is the set of all classes of equivalences.

Proposition 173. *The classes of equivalence of $\sim$ are the "cosets" aH, as a ranges over G. In particular, if G is finite, then the quotient*

$$G/H \overset{\text{def}}{=} \{\overline{g} \textit{ such that } g \in G\},$$

has exactly $\frac{|G|}{|H|}$ elements.

Proof. For any a, b in G,

$$a \sim b \overset{\text{def}}{\iff} \exists h \in H \text{ such that } a^{-1}b = h \iff \exists h \in H \text{ such that } b = ah \iff b \in aH.$$

So the set of elements in a relation with a is precisely the set aH, which we called "left coset". Hence, the elements of G/H are the various left cosets. (Because of normality, each aH is equal to HA, so the elements of G/H are also the right cosets.) By Theorem 114, when G is finite, there are precisely $\frac{g}{h}$ left cosets. □

Theorem 174. *Let H be a normal subgroup of G. Then ${}^{G}/_{H}$ is a group with respect to*

$$\overline{a} * \overline{b} \stackrel{\text{def}}{=} \overline{ab}$$

Such operation makes the map

$$\begin{array}{rccc} \pi : G & \longrightarrow & {}^{G}/_{H} \\ g & \longmapsto & \overline{g}, \end{array}$$

called projection*, a surjective group homomorphism with kernel H.*

Proof. The fact that H is normal is crucial to verify that the operation is well defined. In fact, if $x' \sim x$ and $y' \sim y$, then $h_x \stackrel{\text{def}}{=} x'x^{-1}$ and $h_y \stackrel{\text{def}}{=} y'y^{-1}$ both belong to H. Now

$$x'y'(xy)^{-1} = x'y'y^{-1}x^{-1} = x'h_yx^{-1}.$$

Since H is normal, we can write $x'h_y = kx'$, for some $k \in H$. So we can continue the chain of equalities with

$$x'h_yx^{-1} = kx'x^{-1} = kh_x,$$

which is an element of H. Hence, $x'y'(xy)^{-1} \in H$, which means that $x'y' \sim xy$. This shows that the operation is well defined. The rest is easy: The neutral element is $\overline{e}$, where e is the neutral element of G, and the inverse of $\overline{x}$ is $p(x^{-1}) = \overline{x^{-1}}$.
As for the second claim, the very definition of the operation ensures that π is a homomorphism. Surjectivity is because of the way ${}^{G}/_{H}$ is defined. It remains to compute $\ker \pi$. When is $\overline{g}$ equal to the neutral element of ${}^{G}/_{H}$, which is $\overline{e_G}$? By the way $\sim$ was defined,

$$\overline{e} = \overline{g} \iff e^{-1}g \in H.$$

Hence, $\ker \pi = \{g \in G \ : g \in H\} = H$. □

Corollary 175. *Kernels of group homomorphisms and normal subgroups are the same objects.*

Proof. We have already seen that kernels of homomorphisms are normal. Converely, if H is normal in G, then H is the kernel of the projection $\pi : G \to {}^{G}/_{H}$. □

Example 176. In $(\mathbb{Z}, +)$, any two elements commute, so every subgroup is normal. Fix an integer $m \geq 2$ and consider the subgroup $H =< m >$. Since the group is additive, H consists of all multiples of m, and the associated equivalence relation on $\mathbb{Z}$ is

$$a \sim b \stackrel{\text{def}}{\iff} b - a \in H.$$

Thus $\sim$ is simply "congruence mod m". The quotient ${}^{\mathbb{Z}}/_{< m >}$ consists of exactly m elements, namely, $\overline{0}, \overline{1}, \ldots, \overline{m-1}$; and the projection π from $\mathbb{Z}$ to ${}^{\mathbb{Z}}/_{< m >}$ is simply the map that sends z to the remainder of the division of z by m. So we see that ${}^{\mathbb{Z}}/_{< m >}$ coincides with $\mathbb{Z}_m$.

Proposition 177. *Any quotient of $(\mathbb{Z}, +)$ is isomorphic either to $(\mathbb{Z}, +)$, or to $\{0\}$, or to $(\mathbb{Z}_m, +)$, with $m \geq 2$.*

Proof. By Proposition 177, all subgroups of $\mathbb{Z}$ are of the form $< m >$ for some $m \in \mathbb{N}$. They are all normal, because in $\mathbb{Z}$ any two elements commute. So all quotients of $\mathbb{Z}$ are isomorphic either to $\mathbb{Z}$ (case $m = 0$), or to $\{0\}$ (case $m = 1$), or to $\mathbb{Z}_m$ (case $m \geq 2$). □

Remark 178. We have defined the quotients only for *normal* subgroups. For any group G, for any $x \in G$, and for any normal subgroup H of G, from now on we will adopt the notation $\overline{x}$ to indicate the image of x under the projection π. We have two slogans to remember:

- $\overline{x} = e$ if and only if $x \in H$;
- $\overline{x} = \overline{y}$ if and only if $x^{-1}y \in H$. (Or equivalently, if and only if $xy^{-1} \in H$.)

First Isomorphism Theorem

Recall that for any normal subgroup N in a group G, the projection $\pi : G \to {}^{G}/_{N}$ is the surjective ring homomorphism that sends x to its class of equivalence $\overline{x}$.

Theorem 179 (First Homomorphism Theorem for Groups, Noether 1927)**.** *For any group homomorphism $f : A \to B$, there exist a (unique) group homomorphism*

$$g : {}^{A}/_{\ker f} \to B$$

such that (1) *g is injective,* (2) $\operatorname{Im} g = \operatorname{Im} f$*, and* (3) *$f = g \circ \pi$, where $\pi :\to {}^{A}/_{\ker f}$ is the projection..*
Thus, if there is a surjective *group homomorphism $f : A \longrightarrow B$, then B is isomorphic to ${}^{A}/_{\ker f}$.*

Proof. Let us start from the end. Let us force property number (3) by defining

$$g(\overline{a}) \stackrel{\text{def}}{=} f(a) \quad \text{for all } a.$$

Is this a good definition? If a, a' are distinct elements of A such that $\overline{a} = \overline{a'}$, is it true that $f(a) = f(a')$? By definition of quotient, $\overline{a'} = \overline{a}$ if and only if $a'a^{-1} \in \ker f$, if and only if $e_B = f(a'a^{-1})$. Multiplying both sides by $f(a)$, this is the same as saying, $f(a) = f(a'a^{-1})f(a)$; but the right-hand side equals $f(a'a^{-1}a)$, which is $f(a')$. So summing up,

$$\overline{a'} = \overline{a} \text{ in } {}^{A}/_{\ker f} \iff f(a) = f(a') \stackrel{\text{def}}{\iff} g(\overline{a}) = g(\overline{a'}).$$

The stream of implications from left to right tells us that g is a well-defined function; the converse implications, from right to left, tell us that g is injective.
It remains to see that $\operatorname{Im} g = \operatorname{Im} f$. But this is easy: for any $b \in B$, we have

$$b \in \operatorname{Im} g \iff \exists a \in A \text{ such that } g(\overline{a}) = b \stackrel{\text{def}}{\iff} \exists a \in A \text{ such that } f(a) = b \iff b \in \operatorname{Im} f.$$

This proves the first part. In the particular case where f is surjective, the group homomorphism g we obtained is not only injective but also surjective, since $\operatorname{Im} g = \operatorname{Im} f$; and so in this case the g we constructed is an isomorphism. □

Example 180. In Example 148 we described a surjective group homomorphism f from $(\mathbb{R}^*, \cdot)$ to $(\mathbb{R}, +)$, defined by $f(x) = \log(x^2)$. We also computed $\ker f = \{-1, 1\}$. Thus, by Theorem 179,

$$\frac{\mathbb{R}^*}{\{-1, 1\}} \text{ is isomorphic to } (\mathbb{R}, +).$$

Example 181. In Example 150 we introduced $SL_n(\mathbb{K})$ as kernel. By Theorem 179,

$$\frac{GL_n(\mathbb{F})}{SL_n(\mathbb{F})} \text{ is isomorphic to } \mathbb{F}^*.$$

Similarly (cf. Example 151), for any $n \geq 2$,

$$\frac{O_n(\mathbb{F})}{SO_n(\mathbb{F})} \text{ is isomorphic to } \mathbb{F}^*.$$

Example 182. Consider the surjective group homomorphism

$$\begin{array}{rccc} \pi_1 : & G \times H & \longrightarrow & G \\ & (g, h) & \longmapsto & g. \end{array}$$

Since $\ker \pi_1 = \{(g, h) \text{ such that } g = e_G\} = \{e_G\} \times H$, by Theorem 179 we have that

$$\frac{G \times H}{\{e_G\} \times H} \text{ is isomorphic to } G.$$

3.3 Abelian Groups and the Chinese Remainders theorem

Definition 183 (Abelian)**.** A group G is called *Abelian* if its operation satisfies the commutative property, that is, for each x, y in G one has $xy = yx$.

You are already familiar with several groups that satisfy this property. On the other hand, you also know many finite groups that are *not* Abelian, like the permutation group $\mathcal{S}_3$. In particular, "finitely-generated" does not imply "Abelian', except when the finite number of generators is "one":

Proposition 184. *Every cyclic group is Abelian.*

Proof. If $G =< a >$, for any z, w integers one has $a^z a^w = a^{z+w} = a^{w+z} = a^w a^z$. □

The converse is obviously false: $(\mathbb{Q}*, \cdot)$ is not cyclic, and not even finitely generated.
Now, it is easy to see (exercise!) that the Cartesian product of two Abelian groups is Abelian. In contrast, the products of two cyclic groups is sometimes cyclic, and sometimes not:

Theorem 185. *Let A, B be two cyclic groups with a and b elements, respectively. Then:*

- *If $\gcd(a, b) = 1$, then $A \times B$ is cyclic.*
- *If $\gcd(a, b) = d > 1$, then the period of every element of $A \times B$ divides $\frac{ab}{d}$, so in particular $A \times B$ is not cyclic.*

Proof. By definition of product group, $(x, y)^t = (x^t, y^t)$. So using Lemma 123,

$$(x, y)^t = (e_A, e_B) \iff \begin{cases} x^t = e_A \\ y^t = e_B \end{cases} \iff \begin{cases} t \text{ is a multiple of } \pi(x) \\ t \text{ is a multiple of } \pi(y). \end{cases} \tag{5}$$

- If $\gcd(a, b) = 1$, choose (x, y) in $A \times B$ such that $A =< x >$ and $B =< y >$. Clearly $\pi(x) = a$ and $\pi(y) = b$. Since they have no common factor, any number that is multiple of both a, b must also be a multiple of ab. Hence the coimplication (5) becomes

$$(x, y)^t = (e_A, e_B) \iff t \text{ is a multiple of } ab.$$

- If $\gcd(a, b) = d > 1$, set $a' \stackrel{\text{def}}{=} \frac{a}{d}$ and $b' \stackrel{\text{def}}{=} \frac{b}{d}$. For any (x, y) in $A \times B$, we know that $x^a = e_A$ and $y^b = e_B$ by definition of period. In particular, $x^{ab'} = e_A$, because ab' is a multiple of a; and $y^{ab'} = e_B$, because $ab' = a'db' = a'b$ is a multiple of b. So $(x, y)^{ab'} = (e_A, e_B)$. □

Corollary 186. *$\mathbb{Z}_a \times \mathbb{Z}_b$ is isomorphic to $\mathbb{Z}_{ab}$ $\iff$ $\gcd(a, b) = 1$.*

Proof. By Theorem 185, $\mathbb{Z}_a \times \mathbb{Z}_b$ has an element of period ab if and only if $\gcd(a, b) = 1$. □

Lemma 187. *Let $m_1, \ldots, m_n$ be positive integers such that $\gcd(m_i, m_j) = 1$ for all $i \neq j$. Set $m \stackrel{\text{def}}{=} m_1 m_2 \cdots m_n$. For any integer x,*

$$m \text{ divides } x \iff \text{each } m_i \text{ divides } x.$$

Proof. The direction '⇒' is trivial, so let us focus on '⇐'. Let p be a prime that divides m. By Euclid's Lemma 14, p divides *at least* one of the m_i's; but because the gcd of any two m_i's is 1, this p divides *at most* one of the m_i's. Conclusion: p divides *exactly* one of the m_i's. So if

$$m = p_1^{a_1} p_2^{a_2} p_k^{a_k}$$

then each m_i is either of the form $m_i = p_j^{a_j}$, or (up to reordering the p_j's) of the form

$$m_i = p_j^{a_j} p_{j+1}^{a_{j+1}} \cdots p_{j+h}^{a_{j+h}} \text{ for some } j \in \{1, \ldots, k\},\ h \in \{1, \ldots, k - j\}.$$

So if an integer x is a multiple of all m_i's, it means that the exponent of each p_j in the factorization of x is a_j or larger. So m divides x. □

Theorem 188 (Chinese Remainder Theorem)**.** *Let $m_1, \ldots, m_n \in \mathbb{N}$ with $\gcd(m_i, m_j) = 1$ for all $i \neq j$. Set $m \stackrel{\text{def}}{=} m_1 m_2 \cdots m_n$. For any x in $\mathbb{Z}$ with $0 \leq x \leq m-1$, let us denote by $[x]_i$ its equivalence class modulo m_i. Then the "multi-projection" function*

$$\begin{array}{rccccc} \Pi: & \mathbb{Z}_m & \longrightarrow & \mathbb{Z}_{m_1} \times & \ldots \times & \mathbb{Z}_{m_n} \\ & x & \longmapsto & ([x]_1, & \ldots, & [x]_n). \end{array}$$

is a group isomorphism. Moreover, $\gcd(x, m) = 1$ if and only if $\gcd([x]_i, m_i) = 1$ for all i.

Proof. First of all, Π is injective: in fact, for any integer y in $\{0, \ldots, m-1\}$, we have that

$$\Pi(x) = \Pi(y) \stackrel{\text{def}}{\Longleftrightarrow} [x]_i = [y]_i \text{ for all } i \stackrel{\text{def}}{\Longleftrightarrow} \text{ each } m_i \text{ divides } x-y \stackrel{!}{\Longleftrightarrow} m \text{ divides } x-y \iff x = y,$$

where the second-last equivalence is by the previous Lemma, and the last equivalence is because both x, y are between 0 and $m-1$. Notice that the finite sets $\mathbb{Z}_m$ and $\mathbb{Z}_{m_1} \times \ldots \times \mathbb{Z}_{m_n}$ have the same number of elements, so any injective function between them is automatically surjective. So to conclude that Π is a group isomorphism, we only need to show that $\Pi(x+y) = \Pi(x) + \Pi(y)$, for all x, y in $\{0, \ldots, m-1\}$. To do this, we need to perform $2n$ divisions:

- For all $i \in \{0, \ldots, n\}$, let $x = q_i m_i + r_i$, with $0 \leq r_i < m_i$.
- For all $i \in \{0, \ldots, n\}$, let $y = q_i' m_i + r_i'$, with $0 \leq r_i' < m_i$.

Then by definition

$$\Pi(x) + \Pi(y) = (r_1, r_2, \ldots, r_n) + (r_1', r_2', \ldots, r_n') = \left([r_1 + r_1']_1, [r_2 + r_2']_2 \ldots, [r_n + r_n']_n\right). \quad (6)$$

But at the same time, $x + y = (q_i + q_i')m_i + (r_i + r_i')$. So modulo m_i, $x+y$ is congruent to $r_i + r_i'$. In other words, in each $\mathbb{Z}_{m_i}$ we have $[x+y]_i = [r_i + r_i']_i$. But then

$$\Pi(x+y) = ([x+y]_1, [x+y]_2, \ldots, [x+y]_n) = \left([r_1 + r_1']_1, [r_2 + r_2']_2 \ldots, [r_n + r_n']_n\right). \quad (7)$$

Putting together Equations (6) and (7), we conclude that Π is a group isomorphism.

As for the second claim: Suppose that for some i we have $\gcd([x]_i, m_i) > 1$. Let p be any prime that divides both $[x]_i$ and m_i. Since $x = q_i m_i + [x]_i$, the prime p obviously divides x as well. Also, since p divides m_i, it divides m as well. So $\gcd(x, m) > 1$. Conversely, suppose that $\gcd(x, m) > 1$. Let p' be any prime that divides both x and m. By Euclid's Lemma, there is an i such that p' divides m_i. Since $x = q_i m_i + [x]_i$, it follows that p' divides also $[x]_i$. So $\gcd)([x]_i, m_i) > 1$. □

Corollary 189 (also cited as 'Chinese Remainder Theorem'; Sunzi, 3rd Century AD)**.** *Let $m_1, \ldots, m_n$ be positive integers such that $\gcd(m_i, m_j) = 1$ for all $i \neq j$. Set $m \stackrel{\text{def}}{=} m_1 m_2 \cdots m_n$. For any $a_1, \ldots, a_n$ in $\mathbb{N}$, the system*

$$\begin{cases} x \equiv a_1 \mod m_1 \\ x \equiv a_2 \mod m_2 \\ \quad \vdots \\ x \equiv a_n \mod m_n \end{cases} \quad (8)$$

admits a unique solution x_0 in $\{0, \ldots, m-1\}$. Moreover, any further integer solution is congruent to such x_0 modulo m.

Proof. By Theorem 188, for any $a_1, \ldots, a_n$ in $\mathbb{N}$ there is a unique element x in $\mathbb{Z}_m$ such that

$$\Pi(x) = ([a_1]_1, \ldots, [a_n]_n).$$

Thus the set of solutions to Sunzi;'s problem is given by all integers z such that $\overline{z} = x$ in $\mathbb{Z}_m$. This is simply the set $\{x + km \ : \ k \in \mathbb{Z}\}$. □

Remark 190. "There are certain things whose number is unknown. If we count them by threes, we have two left over; by fives, we have three left over; and by sevens, two are left over. How many things are there?" (Sunzi Suanjing, 500 AD). To re-phrase: Is there a natural number x congruent to 2 mod 3, congruent to 3 mod 5, and congruent to 2 mod 7? In view of the previous theorem, there should be, because $3, 5, 7$ are pairwise coprime. In fact, we expect a unique solution x with $0 < x < \cdot 5 \cdot 7 = 105$; and once we find x, we can get infinitely many other integer solutions simply by repeatedly adding 105. So how do we find the *smallest positive solution*? Here is a hint for a possible, simple (though not so fast) algorithm, called *sieving*. We start with the largest "mod", which is 7. The positive integers congruent to 2 mod 7 are

$$2, 9, 16, 23, \ldots, 7k + 2, \ldots$$

Within this list, we select the integers also congruent to 3 mod 5. The smallest is 23:

$$23, 58, 93, 128, \ldots, (7 \cdot 5)k + 23, \ldots$$

Finally, within this second list, we select the integers also congruent to 2 mod 3. The remaining list contains all positive solutions to Sunzi's problem, with 23 being the smallest:

$$23, 128, \ldots, (7 \cdot 5 \cdot 3)k + 23, \ldots ...$$

The totient function

Definition 191. For any integer $m \geq 2$, the *totient function* $\phi(m)$ counts the positive integers coprime with m and smaller than m. Equivalently, $\phi(m)$ is the size of the group U_m.

Remark 192. By definition, $1 \leq \phi(m) \leq m - 1$. The lower bound is because 1 is coprime with m for all m. As for the upper bound, $\phi(m) = m - 1$ if and only if m is prime. In fact, if m is composite, then $m = ab$ with $1 < a < m$, so $U_m \subseteq \mathbb{Z}_m - \{0, a\}$. So $\phi(m) \leq m - 2$.

An obvious consequence of Proposition 129 is the following extension of the Fermat Little Theorem 130, which represents the case $m = p$ prime (where as we said, $\phi(p) = p - 1$):

Theorem 193 ("Euler Theorem")**.** *For all a in U_m one has $a^{\phi(m)} = 1$.*

Proof. $|U_m| = \phi(m)$ and $e_{U_m} = 1$, so by Proposition 129 $a^{\phi(m)} = 1$ for all a. □

Remark 194. Not necessarily $\phi(m)$ is the smallest integer such that $a^{\phi(m)} = 1$ for all a in U_m. For example, $U_8 = \{1, 3, 5, 7\}$, so $\phi(8) = 4$, but every x of U_9 satisfies $x^2 = 1$.

The Chinese remainder theorem has another important consequence for the totient function:

Lemma 195. *Let $m_1, \ldots, m_n$ be integers larger than 1, such that $\gcd(m_i, m_j) = 1$ for all $i \neq j$. Then $\phi(m_1 m_2 \cdot m_n) = \phi(m_1) \cdot \phi(m_2) \cdots \phi(m_n)$.*

Proof. Set $m \stackrel{\text{def}}{=} m_1 m_2 \cdots m_n$. By the second part of Theorem 188, Π restricts to a bijection

$$\{ \text{ invertible in } \mathbb{Z}_m \} \;\cong\; \{ \text{ invertible in } \mathbb{Z}_{m_1} \} \times \{ \text{ invertible in } \mathbb{Z}_{m_2} \} \times \cdots \times \{ \text{ invertible in } \mathbb{Z}_{m_n} \}.$$

The $n + 1$ sets above have precisely $\phi(m)$, $\phi(m_1)$, $\phi(m_2)$, ..., $\phi(m_n)$ elements, respectively. □

Lemma 196. *If $m = p^a$ is a prime power, then $\phi(m) = p^a - p^{a-1} = p^a(1 - \frac{1}{p})$.*

Proof. Among the integers from 1 to p^a, those *not* coprime with p^a are simply the multiples of p, and there are p^{a-1} of them. So the remaining $p^a - p^{a-1}$ numbers are those coprime with p^a. In conclusion, $\phi(m) = p^a - p^{a-1}$, which is just another way to write $p^a(1 - \frac{1}{p})$. □

Theorem 197 (Euler)**.** *For any integer $m \geq 2$,*

$$\phi(m) \;=\; m \cdot \prod_{p \text{ prime divisor of } m} \left(1 - \frac{1}{p}\right).$$

Proof. Suppose m factors as

$$n = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}.$$

Setting $m_i \stackrel{\text{def}}{=} p_i^{a_i}$ for all $i = 1, 2, \ldots, k$, and applying Lemmas 195 and 196, we obtain

$$\phi(m) \;= \phi(m_1) \cdots \phi(m_k) = p_1^{a_1}\left(1 - \frac{1}{p_1}\right) \cdots p_k^{a_k}\left(1 - \frac{1}{p_k}\right) = n\left(1 - \frac{1}{p_1}\right) \cdots \left(1 - \frac{1}{p_k}\right). \quad \square$$

Example 198. The invertible elements in $\mathbb{Z}_{200}$ are exactly $200\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{5}\right) = 80$. Instead, the invertible elements in $\mathbb{Z}_{210}$ are $210\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{5}\right)\left(1 - \frac{1}{7}\right) = 48$.

Corollary 199. *For any integer n, $\sqrt{\frac{n}{2}} \leq \phi(n) \leq n - 1$.*

Proof. The upper bound is attained when n is prime. Let us show the lower bound[9]. By Lemmas 195 and 196 the quantity $\phi(n)$ is the product of factors of the type $p^{a-1}(p-1)$. We wish to bound from below each of these factors. Our first claim is:

$$\text{If } (p, a) \neq (2, 1), \text{ then } p^{a-1}(p-1) \geq (\sqrt{p})^a. \tag{9}$$

In fact, for $a \geq 2$, inequality (9) is obvious: since $a - 1 \geq \frac{a}{2}$, we have $p^{a-1}(p-1) \geq p^{a-1} \geq (\sqrt{p})^a$. If instead $a = 1$, inequality (9) simplifies to $p - 1 \geq \sqrt{p}$, so it becomes a calculus exercise to check that $(p-1)^2 > p$ if and only if $p \in (-\infty, \frac{3-\sqrt{5}}{2}) \cup (\frac{3+\sqrt{5}}{2}, \infty)$. So for $p = 2$ the inequality is actually false, but it is true for every other prime $p \geq 3 > \frac{3+\sqrt{5}}{2}$. Our next claim is:

$$\text{Unless } n \text{ is twice an odd integer}, \qquad \phi(n) \geq \sqrt{n}. \tag{10}$$

In fact, "n is twice an odd integer" if and only if "in the factorization of n, one of the factors is p^a with $(p, a) = (1, 2)$". So if n is *not* twice an odd integer, then every factor of the type $p^{a-1}(p-1)$ in $\phi(n)$, since$(p, a) \neq (2, 1)$, will be at least $(\sqrt{p})^a$, by inequality (9). So we compute

$$\phi(n) = p_1^{a_1 - 1}(p_1 - 1) \cdots p_k^{a_k - 1}(p_k - 1) \geq \sqrt{p_1}^{a_1} \cdots \sqrt{p_k}^{a_k} = \sqrt{p_1^{a_1} \cdots p_k^{a_k}} = \sqrt{n}.$$

We are now ready to prove the theorem. We distinguish two cases: If n is *not* twice an odd number, then by Inequality 10 $\phi(n) \geq \sqrt{n} \geq \sqrt{\frac{n}{2}}$. If instead $n = 2n'$ with n' odd, then in particular n' is not twice an odd number, so Inequality 10 tells us that $\phi(n') \geq \sqrt{n'}$ and

$$\phi(n) = \phi(2)\phi(n') = \phi(n') \geq \sqrt{n'} = \sqrt{\frac{n}{2}}. \qquad \square$$

Deeper thoughts 200. With much more effort, the lower bound above can be improved a lot: There are lower bounds proportional to $\frac{n}{\log \log n}$. There are many open problems on the totient function. In 1922, Carmichael conjectured that there is no number m with an *exclusive* totient: that is, a number m such that for all $n \neq m$ one has $\phi(n) \neq \phi(m)$. In 1932, Lehmer conjectured that for no composite number n, $\phi(n)$ divides $n - 1$. Since for prime numbers $\phi(p)$ does divide $p - 1$ (they are equal!), Lehmer's conjecture can be rephrased as "$\phi(n)$ divides $n - 1$ if and only if n is prime".

[9] Proof by F. Nicolas, *A simple, polynomial-time algorithm for the matrix torsion problem*, arXiv:0806.2068.

3.4 Finite Abelian groups are products of cyclic groups

In this section we will prove two famous results: (1) a converse of Lagrange theorem for Abelian groups; (2) Gauss' famous theorem that if m is prime, then U_m is cyclic. The gateway to both is a structural results on finite Abelian groups. Namely, we are going to show that they are all (isomorphic to) products of cyclic groups.

Remark 201. The Cartesian product of sets is "associative" and "commutative" up to isomorphism. By this we mean that there are obvious group isomorphisms

$$f_1 : \begin{array}{ccc} G \times H & \longrightarrow & H \times G \\ (g,h) & \longmapsto & (h,g) \end{array} \qquad \text{and} \qquad f_2 : \begin{array}{ccc} G \times (H \times K) & \longrightarrow & (G \times H) \times K \\ (g,(h,k)) & \longmapsto & ((g,h),k) \end{array}$$

Therefore we usually adopt the following conventions: (1) In case we have a Cartesian product of finite cyclic groups, we rearrange the groups by increasing number of elements; (2) when we have a Cartesian product of three or more groups, we omit the brackets. So we will write $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_5$ instead of $\mathbb{Z}_5 \times (\mathbb{Z}_2 \times \mathbb{Z}_2)$.

Example 202. The two 60-element groups $\mathbb{Z}_2 \times \mathbb{Z}_{30}$ and $\mathbb{Z}_6 \times \mathbb{Z}_{10}$ are not cyclic. Using Corollary 186 (and the notation of Remark 201) we can break them further.

$$\begin{aligned} \mathbb{Z}_2 \times \mathbb{Z}_{30} &\cong \mathbb{Z}_2 \times (\mathbb{Z}_3 \times \mathbb{Z}_{10}) \cong \mathbb{Z}_2 \times \mathbb{Z}_3 \times (\mathbb{Z}_2 \times \mathbb{Z}_5) \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5 \\ \mathbb{Z}_6 \times \mathbb{Z}_{10} &\cong (\mathbb{Z}_2 \times \mathbb{Z}_3) \times (\mathbb{Z}_2 \times \mathbb{Z}_5) \cong \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3 \times \mathbb{Z}_5. \end{aligned}$$

So they are isomorphic to one another! Remember though that **we cannot split $\mathbb{Z}_m$ further if m is a prime power**. So for example

$$\mathbb{Z}_{60} \cong \mathbb{Z}_3 \times \mathbb{Z}_{20} \cong \mathbb{Z}_3 \times \mathbb{Z}_4 \times \mathbb{Z}_5,$$

but we are not allowed to split $\mathbb{Z}_4$ further, because $\mathbb{Z}_4$ is *not* $\mathbb{Z}_2 \times \mathbb{Z}_2$, by Corollary 186. So $\mathbb{Z}_{60}$ is a different Abelian group than $\mathbb{Z}_2 \times \mathbb{Z}_{30}$.

At the moment we see only two distinct Abelian groups with 60 elements. Of course, if we proved that every Abelian groups is a product of cyclic groups, we could immediately conclude that there are only two distinct Abelian groups with 60 elements. So let us prove it!

Lemma 203. *Let $G =< g_1, \ldots, g_n >$ be an Abelian group. Let $a_1, \ldots, a_n$ be any list of integers, possibly with repetitions, with $\gcd(a_1, \ldots, a_n) = 1$. Then there is another list of n generators for G that includes the element $g_1^{a_1} g_2^{a_2} \cdots g_n^{a_n}$.*

Proof. By Proposition 93, there is an integer matrix A of determinant 1 whose first row is $a_1, \ldots, a_n$. Let $a_{i,j}$ be the element in the i-th row, j-th column of A. Define

$$x_i \stackrel{\text{def}}{=} g_1^{a_{i,1}} \cdot g_2^{a_{i,2}} \cdots g_n^{a_{i,n}}. \tag{11}$$

We claim $\{x_1, \ldots, x_n\}$ is the desired generating set. Let us prove it. First of all

$$x_1 = g_1^{a_{1,1}} \cdot g_2^{a_{1,2}} \cdots g_n^{a_{1,n}} = g_1^{a_1} g_2^{a_2} \cdots g_n^{a_n}.$$

Moreover, $g_1, \ldots, g_n$ are in G, so every x_i is obviously in G. It remains to show is that $G =< x_1, \ldots, x_n >$. To see this, consider the matrix $B \stackrel{\text{def}}{=} A^{-1}$. Since A has determinant 1, B has also

integer entries (this follows by the Cofactor Formula for calculating the inverse matrix). Let $b_{j,k}$ be the element in the j-th row, k-th column of B. Using Equation 11, we compute

$$\begin{aligned}
&(x_1)^{\,b_{j,1}} \cdot (x_2)^{b_{j,2}} \cdots (x_n)^{\,b_{j,n}} = \\
&= \left(g_1^{a_{1,1}} \cdots g_n^{a_{1,n}}\right)^{b_{j,1}} \cdot \left(g_1^{a_{2,1}} \cdots g_n^{a_{2,n}}\right)^{b_{j,2}} \cdots \left(g_1^{a_{n,1}} \cdots g_n^{a_{n,n}}\right)^{b_{j,n}} = \\
&= \left(g_1^{b_{j,1}a_{1,1}} \cdots g_n^{b_{j,1}a_{1,n}}\right) \cdot \left(g_1^{b_{j,2}a_{2,1}} \cdots g_n^{b_{j,2}a_{2,n}}\right) \cdots \left(g_1^{b_{j,n}a_{n,1}} \cdots g_n^{b_{j,n}a_{n,n}}\right) = \\
&= (g_1)^{\sum_{i=1}^n b_{j,k}a_{k,1}} \cdot (g_2)^{\sum_{i=1}^n b_{j,k}a_{k,2}} \cdot \ldots \cdot (g_n)^{\sum_{i=1}^n b_{j,k}a_{k,n}} = \\
&\stackrel{!}{=} (g_1)^0 \cdot (g_2)^0 \cdot \cdots \cdot (g_{j-1})^0 \cdot (g_j)^1 \cdot (g_{j+1})^0 \cdots (g_n)^0 = g_j,
\end{aligned}$$

where the marked equality is due to the fact that $\sum_{i=1}^n b_{j,k}a_{k,\ell}$ is the (j,ℓ)-entry of the matrix BA, which is the identity matrix; but the (j,ℓ)-entry of the identity matrix is always 0, unless $j = \ell$, in which case it is equal to 1. So in conclusion, for every $j \in \{1, \ldots, n\}$, we have

$$g_j = x_1^{b_{j,1}} \cdots x_n^{b_{j,n}}. \tag{12}$$

Equation 12 tells us that every g_j is in $< x_1, \ldots, x_n >$.
So $G = < g_1, \ldots, g_n > \subseteq < x_1, \ldots, x_n > \subseteq G$, which implies $G =< x_1, \ldots, x_n >$. □

Theorem 204 (Smith 1861, Kronecker 1870)**.** *Every finite Abelian group G is isomorphic to the product of cyclic groups.*

Proof by E. Schenkman. The idea is to proceed by induction on the *smallest number of generators* n of G. If $n = 1$, then G is cyclic, and we are done. Suppose now $n \geq 2$. Let g_1 be an element of smallest period among those elements that form a generating set of n elements for G. So to recap our assumptions:

- there are elements $g_2, \ldots, g_n$ so that $G =< g_1, \ldots, g_n >$;
- any subset $X \subset G$ with less than n elements cannot be a generating set for G;
- no element x with $\pi(x) < \pi(g_1)$ can be part of a size-n generating set for G.

Set $H \stackrel{\text{def}}{=} < g_1 >$ and $K \stackrel{\text{def}}{=} < g_2, \ldots, g_n >$. Being generated by less than n elements, K is by inductive assumption a product of cyclic groups. H is clearly cyclic. What we want to show is that $G \cong H \times K$. But via Proposition 171, all we need to show is that $H \cap K = (e)$. (The normality of H, K is automatic, because G is Abelian; the smallest subgroup containing both H and K is $< g_1, \ldots, g_n >= G$.) So let us prove that $H \cap K \subseteq (e)$, the other inclusion being obvious. By contradiction, suppose there exists a $z \neq e$ inside $H \cap K$. Since $z \in H =< g_1 >$, we can write $z = g_1^{a_1}$ for some $a_1 \in \{1, \ldots, k-1\}$. Also, $z \in K$, so $z = g_2^{a_2} g_3^{a_3} \cdot g_n^{a_n}$ for some $a_2, \ldots, a_n \in \mathbb{N}$. Set $d \stackrel{\text{def}}{=} \gcd(a_1, a_2, \ldots, a_n)$. Because d is the greatest common divisor of the a_i's, the integers $-\frac{a_1}{d}$, $\frac{a_2}{d}$, $\ldots$, $\frac{a_n}{d}$ have gcd equal to 1. By Lemma 203, we can find another generating set of size n for G that includes the element

$$x \stackrel{\text{def}}{=} g_1^{-\frac{a_1}{d}} g_2^{\frac{a_2}{d}} \cdots g_n^{\frac{a_n}{d}}.$$

By construction,

$$x^d = g_1^{-a_1} g_2^{a_2} \cdots g_n^{a_n} = (g_1^{a_1})^{-1} \cdot (g_2^{a_2} \cdots g_n^{a_n}) = z^{-1} \cdot z = e.$$

So $\pi(x)$ divides d. In turn, d divides a_1, which was smaller than k. This implies that

$$\pi(x) \leq d \leq a_1 < k = \pi(g_1),$$

a contradiction with how g_1 was chosen. □

Corollary 205. *For every finite Abelian group G, there exist natural numbers h, $m_1, \ldots, m_h$ and prime numbers $p_1 \leq \ldots \leq p_h$ (not necessarily distinct), such that G can be decomposed as*

$$G \cong \mathbb{Z}_{p_1^{m_1}} \times \mathbb{Z}_{p_2^{m_2}} \times \ldots \times \mathbb{Z}_{p_h^{m_h}}.$$

Proof. Using Corollary 186 (and the notation of Remark 201), we have seen that it is possible to break up every finite cyclic group until it is the product of cyclic groups *whose sizes are prime powers* (not necessarily distinct). Compare Example 202. □

Corollary 206 (Converse of Lagrange for Abelian groups)**.** *Let G be a finite Abelian group. If d is a divisor of $|G|$, then G has a subgroup with exactly d elements.*

Proof. By Corollary 205,

$$G \cong \mathbb{Z}_{p_1^{m_1}} \times \mathbb{Z}_{p_2^{m_2}} \times \ldots \times \mathbb{Z}_{p_h^{m_n}}, \text{ with } |G| = p_1^{m_1} p_2^{m_2} \cdots p_n^{m_n}.$$

Since d divides $|G|$, each prime factor of d is also a factor of $|G|$; so by the Unique Factorization theorem, d must decompose as

$$d = p_1^{d_1} p_2^{d_2} \cdots p_n^{d_n}, \text{ with } d_i \leq m_i \text{ for each } i.$$

Now, in the additive group $(\mathbb{Z}_{p_1^{m_1}}, +)$ there is an element of period $p_1^{d_1}$, namely, the element

$$a_1 \stackrel{\text{def}}{=} p_1^{m_1 - d_1}.$$

(In fact, by construction $p_1^{d_1} \cdot a_1 = p_1^{d_1} p_1^{m_1 - d_1} = p_1^{m_1} \equiv 0$.) In particular, the subgroup

$$A_1 \stackrel{\text{def}}{=} < p_1^{m_1 - d_1} >$$

has exactly $p_1^{d_1}$ elements. Similarly, inside $\mathbb{Z}_{p_i^{m_i}}$ the subgroup $A_i \stackrel{\text{def}}{=} < p_i^{m_i - d_i} >$ has exactly $p_i^{d_i}$ elements. It follows that the subgroup we are looking for is

$$H \stackrel{\text{def}}{=} A_1 \times A_2 \times \ldots \times A_n.$$

□

3.5 Exercises

1. Let G be a group with $2p$ elements, p prime. Prove that if every element of G has period 1 or 2, then G is Abelian. Use this to show that G contains a subgroup with p elements. (Hint: what can the period of an element $x \neq e$ be?)
2. Given $n \in \mathbb{N}$, find the smallest m such that S_m contains a cyclic subgroup of size n. (Hint: $(1,2,3,4,5,6)(7,8,9,10,11,12,13,14,15,16,17,18,19,20,21)$ has period 30, so in $\mathcal{S}_21$ there is a cyclic subgroup of size 30. However, one can do better: already $\mathcal{S}_11$ has a subgroup isomorphic to $\mathbb{Z}_{30}$.)
3. The *center* of G is defined as $Z(G) \stackrel{\text{def}}{=} \{g \in G : \text{ for all } x \text{ in } G,\ gx = xg\}$. Show that $Z(G)$ is always a normal subgroup of G.
4. Let G be a group. Show that the quotient of G by $Z(G)$ is cyclic if and only if G is Abelian.
5. How many Abelian groups are there with exactly 100 elements?
6. Let m be a *squarefree integer*, i.e. an integer that is not the multiple of any square of an integer. (E.g. 10 is square free, 20 is not.) Show that up to isomorphism, there is only one Abelian group of size m.

7. For the previous exercise, the Abelian assumption is crucial: Let G be the smallest group formed by two elements x, y, and subject only to the relations

$$x^7 = e = y^3 \quad \text{and} \quad yx = x^2y.$$

Thus G is not commutative. Show that G has exactly $21 = 3 \cdot 7$ elements, namely,

$$ey^a, xy^a, x^2y^a, x^3y^a, x^4y^a, x^5y^a, x^6y^a, \qquad \text{for } a = 1, 2, 3.$$

8. The *quaternion group* is the 8-element set $Q = \{e, -e, i, -i, j, -j, k, -k\}$, with the operation:

$\cdot$	e	$-e$	i	$-i$	j	$-j$	k	$-k$
e	e	$-e$	i	$-i$	j	$-j$	k	$-k$
$-e$	$-e$	e	$-i$	i	$-j$	j	$-k$	k
i	i	$-i$	$-e$	e	k	$-k$	$-j$	j
$-i$	$-i$	i	e	$-e$	$-k$	k	j	$-j$
j	j	$-j$	$-k$	k	$-e$	e	i	$-i$
$-j$	$-j$	j	k	$-k$	e	$-e$	$-i$	i
k	k	$-k$	j	$-j$	$-i$	i	$-e$	e
$-k$	$-k$	k	$-j$	j	i	$-i$	e	$-e$

Show that every subgroup of the quaternion group is normal.

4 C-rings, Fields, Domains, and Polynomials

4.1 Commutative Rings

A **commutative ring** or **C-ring** consists of a set A endowed with two operations $+$ and $\cdot$ that satisfy the following eight axioms:

(R0) The operations are *internal*. That is, for all x, y in A, the elements $x + y$ and $x \cdot y$ are both in A.

(R1) The operation $+$ is *associative*. That is, for all x, y, z in A, $x + (y + z) = (x + y) + z$.

(R2) The operation $+$ is *commutative*. That is, for all x, y, z in A, $x + y = y + x$.

(R3) The operation $+$ has a unique *neutral element*. That is, there exists an element z in A such that for all x in A, $x + z = x$. From now on we denote such element by "0".

(R4) Every element has a unique *additive inverse*. That is, for all x in A there exists exactly one element y in A such that $x + y = 0$. From now on we denote such element by "$-x$".

(R5) The operation $\cdot$ is *associative*. That is, for all x, y, z in A, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.

(R6) The operation $\cdot$ is *commutative*. That is, for all x, y, z in A, $x \cdot y = y \cdot x$.

(R7) The operation $\cdot$ *distributes* $+$: for all x, y, z in A, $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.

Notation. We write $a - b$ as a shortening of $a + (-b)$. Moreover, we usually write xy instead of $x \cdot y$. Note also that by associativity, it is not ambiguous to write $abcd$ instead of $a(b(cd))$ or of $(ab)(cd)$. In fact, no matter how you insert brackets, the result is always the same.

Remark 207. Some textbooks rephrase axiom (R3) as "The operation $+$ has a neutral element". Uniqueness is implied: Were there two neutral elements z and w, we would have $z + w = z$ (because w is neutral) yet also $z + w = w$ (being z neutral), so $z = w$. Similarly, some textbooks rephrase axiom (R4) as "Every element has an additive inverse". Also in this case, uniqueness is implied: Were there two y, y' in A such that $x + y = 0 = x + y'$, then we would have $y' = y' + 0 = y' + (x + y) = (y' + x) + y = 0 + y = y$.

Remark 208. If $(A, +, \cdot)$ is a C-ring, then $(A, +)$ is an Abelian group. Given any Abelian group $(G, +)$, by endowing G with the operation $\cdot$ defined by $a \cdot b = 0$ for all a, b, one obtains a C-ring $(G, +, \cdot)$. See also Remark 218.

Example 209. The empty set is not a commutative ring: In fact, by axiom (R4), any C-ring must contain at least one element, namely, the neutral element 0.
The set $\{0\}$, instead, *is* a C-ring. So the smallest C-ring has one element.

Example 210. Let m be a positive integer. The set $\mathbb{Z}_m = \{0, 1, \ldots, m\}$ is a C-ring, with the operations of addition and multiplication "modulo m". So for any positive integer m, there is a C-ring with exactly m elements.

Example 211. The sets $(\mathbb{Z}, +, \cdot)$, $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$, are *C*-rings with the usual addition and multiplication. So there are also infinite C-rings.

Now we can see some general properties of C-rings. Since every C-ring is also an Abelian group with respect to addition, whenever $a + b = a + c$ we can conclude that $b = c$. (In fact, we can sum $-a$ to both sides...). The analogous property with respect to multiplication does not always work. Besides, even in $\mathbb{Z}$ you know that $3 \cdot 0 = 7 \cdot 0$, but $3 \neq 7$. In fact:

Proposition 212. *Let A be a C-ring. For all a in A, $a \cdot 0 = 0$.*

Proof. Being 0 neutral element, $a \cdot 0 + 0 = a \cdot 0 = a \cdot (0 + 0) = a \cdot 0 + a \cdot 0$, where in the last step we used distributivity. So by Cancellation with respect to sum, $0 = a \cdot 0$. □

Proposition 213. *Let A be a C-ring. For all a, b in A, $(-a)b = a(-b) = -(ab)$.*

Proof. Since the additive inverse is unique, to check that $(-a)b = -(ab)$ it suffices to prove that $(-a)b$ is an additive inverse of ab; that is, we need to show that $(-a)b + ab = 0$. This can be done using distributivity: $(-a)b + ab = (-a + a)b = 0b = 0$, where in the last step we applied Proposition 212. Similarly, $a(-b) + ab = a(-b + b) = a \cdot 0 = 0$, which shows that also $a(-b)$ is the additive inverse of ab. □

Proposition 214. *Let A be a C-ring. For all a, b in A, $(-a)(-b) = ab$.*

Proof. Obviously $-ab + ab = 0$. On the other hand, by Proposition 213, $-ab = (-a)b$; so

$$-ab + (-a)(-b) = (-a)b + (-a)(-b) = (-a)(b + -b) = (-a)0 = 0.$$

So both ab and $(-a)(-b)$ are the additive inverse of $-ab$. Hence, they must be equal. □

4.2 Invertible elements and Fields

In the definition of C-ring, there are two blatant asymmetries between the two operations. First of all, we required only $+$ to have a neutral element, which we denoted by 0; there was no mention of a neutral element for multiplication. Second of all, we required every element to have an additive inverse, though there was no request of a multiplicative one. But nothing prevents us from focusing on "special" rings with these extra properties.

Definition 215. A **C-ring with 1** is a C-ring that satisfies the additional axiom

(R8) The operation $\cdot$ has a (necessarily unique[10]) *neutral element*. That is, there exists a unique element z in A such that for all x in A, $xz = x$. We denote such z by "1".

Example 216. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ are C-ring with 1. For any positive integer m, $\mathbb{Z}_m$ is a C-ring with 1. Instead, the set $2\mathbb{Z}$ of EVEN integers is a C-ring "without 1".

Proposition 217. $0 \neq 1$ *(unless $A = \{0\}$).*

Proof. Let $a \neq 0$ in A. By Proposition 212, $a \cdot 0 = 0 \neq a = a \cdot 1$. □

Remark 218. As a follow up to Remark 218: Given any Abelian group $(G, +)$, is it always possible to endow G with an operation $\cdot$ such that $(G, +, \cdot)$ is a C-ring *with 1*? The answer is negative, because of the following observation. Suppose A is a C-ring with 1 such that in $(A, +)$ every element has finite (additive) period. Let k be the period of 1. Then for every x in A by the distributive property we have

$$x + x + \ldots + x \text{ (k times)} = x \cdot (1 + 1 + \ldots + 1) = x \cdot 0 = 0,$$

which means that the additive period of any element of $(A, +)$ divides the additive period of 1. In contrast, consider now the Abelian group $\mathbb{Q}/\mathbb{Z}$. It is easy to see that in this group every element has finite (additive) period: In fact, the class of any fraction $\frac{m}{n}$, if $\gcd(m, n) = 1$, has period n. However, it is also easy to see that there is no integer n such that the period of every element of $\mathbb{Q}/\mathbb{Z}$ divides n. So this $\mathbb{Q}/\mathbb{Z}$ cannot be turned into a C-ring with 1.

[10] Were there two distinct neutral elements z and w, we would have $zw = z$ (because w is neutral) yet also $zw = w$ (being z neutral), so $z = w$; a contradiction.

Definition 219. Let A be a C-ring with 1. An element x in A is called *invertible* if there exists an element y in A, called a *multiplicative inverse*, such that $xy = 1$.

For example, 1 is always invertible, because $1 \cdot 1 = 1$. Instead 0 is never invertible, because $0 \times y = 0 \neq 1$ for all y in A, by Proposition 212. Note that the definition of "invertible" makes sense only if there is an element 1 in the ring.

Proposition 220. *If it exists, the multiplicative inverse is also unique.*

Proof. Let u, u' be multiplicative inverses for a. Then $u' = u'1 = u'(au) = (u'a)u = 1u = u$. □

Notation. From now on we denote the multiplicative inverse of x (whenever it exists!) by "x^{-1}". Recall that we decided to write $a-b$ as a shortening for $a+(-b)$; similarly, some authors introduce the notation $\frac{a}{b}$ as a shortening for ab^{-1}. We stress that the notation $\frac{a}{b}$ only makes sense because of the commutativity axiom (R6): Were the product not commutative, then we would need to distinguish ab^{-1} from $b^{-1}a$, so the notation "$\frac{a}{b}$" would be ambiguous. For these reasons, most textbooks prefer to use the notation ab^{-1} rather than $\frac{a}{b}$.

Definition 221. A **field** is a C-ring with 1 that satisfies another additional axiom:

(R9) Every element $a \neq 0$ of F is invertible.

Example 222. $\mathbb{Q}, \mathbb{R}, \mathbb{C}$ are fields; $\mathbb{Z}$ is not a field. In $\mathbb{Z}$, only 1 and -1 are invertible.

Is $\mathbb{Z}_m$ a field? It turns out that the answer depends on m.

Proposition 223. *$\mathbb{Z}_m$ is a field $\iff$ m is a prime number.*

Proof. "$\Leftarrow$" Let $a \neq 0$ in $\mathbb{Z}_m$. Since m is prime, $\gcd(a, m) = 1$. So a is invertible by Prop. 26.
"$\Rightarrow$" Had m a divisor d with $1 < d < m$, then we would have $\gcd(d, m) = d \neq 1$, so by Prop. 26 this d in $\mathbb{Z}_m$ is not invertible. □

Corollary 224. *The smallest field is $\mathbb{Z}_2$.*

Proof. Since 2 is a prime number, $\mathbb{Z}_2$ is a field. On the other hand, a field must contain by definition at least two different elements, 0 and 1. □

4.3 Zerodivisors and Domains

Definition 225. Let A be a C-ring with at least two elements.

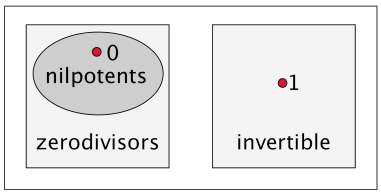
- An element $a \in A$ is called *zero-divisor* if there exists some element $b \neq 0$ such that $ab = 0$.
- An element a of A is called *nilpotent* if there is a positive integer k such that $a^k = 0$.

Note that by Proposition 212, if $a^k = 0$, then $a^n = 0$ for all $n \geq k$, so we equivalently could have written "a is called nilpotent if there is a positive integer k such that $a^n = 0$ for all $n \geq k$".

Example 226. 0 is obviously nilpotent. It is also a zero-divisor because if $x \neq 0$ (here we use that A has at least two elements), by Proposition 212 $0 \cdot x = 0$. If A is a C-ring with 1, then 1 cannot be a zerodivisor, because if $x \neq 0$, then $1 \cdot x \neq 0$; and it cannot be nilpotent either, because $1^k = 1$ for all positive k, and $1 \neq 0$. (Here again we used that A has at least two elements, cf. Proposition 217.)

Proposition 227. *Nilpotent $\Longrightarrow$ Zerodivisor $\Longrightarrow$ Not invertible.*
The two converse implications are false.

Proof. Let a be a nilpotent element and let r be the smallest positive integer for which $a^r = 0$. If $r = 1$, then $a = 0$, which is a zero-divisor. If $r \geq 2$, then $a^{r-1} \neq 0$ by definition of r, and $a \cdot a^{r-1} = a^r = 0$. So either way, a is a zero-divisor. As for the second implication: Let a be a zerodivisor. Let $b \neq 0$ such that $ab = 0$. Were a invertible, we could multiply by a^{-1} and obtain $b = 0$, a contradiction. Counterexamples for the converse implications: In $\mathbb{Z}_{10}$, the element 5 is a zerodivisor (because $2 \cdot 5 = 0$), but not nilpotent (because $5^n = 5$ for all positive integers n); in $\mathbb{Z}$, any $z \notin \{-1, 0, 1\}$ is not invertible, but not a zero-divisor. □



Note that a *field* is a C-ring where the only non-invertible element is zero. Inspired by this and by Proposition 227 above, we give two new definitions:

Definition 228. A C-ring with at least two elements is called
- a *domain*, if the only zero-divisor is zero.
- *reduced*, if the only nilpotent is zero.

Corollary 229. *All fields are domains, and all domains are reduced.*

Proof. Straightforward from the definitions and Proposition 227. □

Example 230. $\mathbb{Z}_2$ and $\mathbb{Q}$ are fields. $\mathbb{Z}$ and $2\mathbb{Z}$ are domains, but not fields. $\mathbb{Z}_6$ and $\mathbb{Z}_{10}$ are reduced, not domains. (cf. also Proposition 232 below). $\mathbb{Z}_4$ and the C-ring $\{0\}$ are not even reduced.

It turns out that domains exactly as those C-rings where one can perform "cancellation with respect to product", on the condition that what you cancel by is non-zero.

Proposition 231. *Let A be a C-ring. The following are equivalent:*

(1) *A is a domain, i.e. the only zerodivisor is 0.*

(2) *For all a, b in A, if $ab = 0$, then either $a = 0$ or $b = 0$.*

(3) *For every $a \neq 0$, if $ab = ac$ then $b = c$.*

Proof.
(1) ⇒ (2). By contradiction, suppose there exist a, b in A such that $ab = 0$, but both $a \neq 0$ and $b \neq 0$. Then a and b are zerodivisors.
(2) ⇒ (3). If $ab = ac$, with $a \neq 0$, then $a(b - c) = 0$. So by the assumption, either $a = 0$ or $b - c = 0$. But $a \neq 0$, so $b - c = 0$.
(3) ⇒ (1). By contradiction, suppose there is a zerodivisor $a \neq 0$. So for some $b \neq 0$ we have $ab = 0$. So by Proposition 212, $ab = a \cdot 0$, which implies $b = 0$; a contradiction. □

Recall that by Proposition 223, $\mathbb{Z}_m$ is a field if and only if m is prime.

Proposition 232 (Euclid's Lemma rivisited)**.** *Let $m \geq 2$.*
- *$\mathbb{Z}_m$ is a domain if and only if m is prime.*

- *$\mathbb{Z}_m$ is reduced if and only if m is a product of (one or more) distinct primes.*

Proof.

- If m is a prime number and $ab = 0$ in $\mathbb{Z}_m$, then m divides ab. By Euclid's Lemma 13, m is either a factor of a or of b. In other words, either $a \equiv 0 \bmod m$ or $b \equiv 0 \bmod m$. Conversely: if m is not prime, then it is a product of two smaller numbers, $m = ab$. So in $\mathbb{Z}_m$ we have $ab = 0$, which means that a and b are zerodivisors.
- Suppose $m = p_1 \cdots p_k$, with $p_i \neq p_j$ for all $i \neq j$. If n^k is a multiple of m, then each p_i divides n^k, so by Euclid's lemma each p_i divides n. Hence, n is a multiple of m. So $n = 0$ in $\mathbb{Z}_m$. Conversely, suppose that $m = p^2 z$, for some integer z and some prime p. Setting $x \stackrel{\text{def}}{=} pz$, we have that $x^2 = p^2 z^2 = mz$. Thus $x^2 = 0$ in $\mathbb{Z}_m$. □

4.4 Polynomials

What is a monomial? What does "x" mean? When are two polynomials equal? In this section, we will try to answer these simple questions. Let us first recall a notion from calculus.

Definition 233. Let A be an arbitrary set. A *sequence in* A is a function $a : \mathbb{N} \to A$. The usual convention is to write a_i instead of $a(i)$. If 0 is an element of A, we say that a sequence is *eventually zero* if there exists an integer M such that $a_i = 0$ for all $i > M$. A common convention is to write down eventually zero sequences as finite vectors, by listing the images $a_0, a_1, \ldots, a_M$ and by forgetting the infinite sequence of zeroes that comes next.

Definition 234 (C-ring of polynomials with coefficient in a C-ring)**.** Let A be a C-ring with 1. The sequences $f = (a_0, a_1, a_2, \ldots, a_n, a_{n+1}, \ldots)$ in A that are eventually zero are called *polynomials with coefficients in* A. The set of all polynomials with coefficients in A is denoted by $A[x]$. With the notation above, the set $A[x]$ can be written as

$$A[x] \stackrel{\text{def}}{=} \{f = (a_0, a_1, a_2, \ldots, a_n) \text{ such that } n \in \mathbb{N} \text{ and } a_i \in A\}.$$

The original ring A can be thought of as a subset of $A[x]$, by identifying any element c of A with the "constant polynomial" $(c, 0, 0, \ldots)$ of $A[x]$.

Theorem 235. *$A[x]$ is a C-ring with 1, when endowed with the following operations:*

- $(a_0, \ldots, a_n) + (b_0, \ldots, b_m)$ *is the sequence* $c_0, c_1, \ldots$ *where* $c_i \stackrel{\text{def}}{=} a_i + b_i$*;*
- $(a_0, \ldots, a_n) \cdot (b_0, \ldots, b_m)$ *is the sequence* $c_0, c_1, \ldots$ *where* $c_i \stackrel{\text{def}}{=} \sum_{k=0}^{i} a_k \cdot b_{i-k}$.

The neutral element with respect to the sum is the sequence $(0, 0, 0, \ldots...)$*; the neutral element with respect to the product is the sequence* $(1, 0, 0, \ldots)$.

The proof is left as exercise. In view of the theorem, we adopt the notation $0 \stackrel{\text{def}}{=} (0, 0, \ldots)$ and $1 \stackrel{\text{def}}{=} (1, 0, 0, \ldots)$. We are now ready to explain who "**x**" is:

Definition 236. We call x the sequence $(0, 1, 0, \ldots...)$. So for example

$$x^2 = x \cdot x = (0, 1, 0, 0 \ldots) \cdot (0, 1, 0, 0 \ldots) = (0, 0, 1, 0).$$

Similarly,

$$x^3 = x \cdot x^2 = (0, 1, 0, 0) \cdot (0, 0, 1, 0) = (0, 0, 0, 1).$$

By induction, x^n is the sequence $c_0, c_1, \ldots$, where $c_n = 1$ and $c_i = 0$ for all $i \neq n$.

Remark 237. Note that

$$(a_0, a_1) = (a_0, 0) + (0, a_1) = a_0 \cdot (1, 0) + a_1 \cdot (0, 1) = a_0 \cdot 1 + a_1 \cdot x.$$

Similarly,

$$(a_0, a_1, a_2) = (a_0, 0, 0) + (0, a_1, 0) + (0, 0, a_2) = a_0 \cdot 1 + a_1 \cdot x + a_2 \cdot x^2.$$

More generally,

$$(a_0, \ldots a_n) = \sum_{k=0}^{n} a_k \cdot x^k.$$

Notation. We can now write down polynomials the way you are used to. In fact, in view of the identity above, we will write down the polynomial $(a_0, \ldots a_n)$ as

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n.$$

Definition 238. Let f be a polynomial in $A[x]$. Say $f = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$ with the notation above. Let $b \in A$. We call *evaluation of f at b* the element

$$f(b) \stackrel{\text{def}}{=} a_0 + a_1b + a_2b^2 + \ldots + a_nb^n.$$

Notationally, you can think of $f(b)$ as the result of "plugging in b for x".

This way, every polynomial f in $A[x]$ naturally induces a function from A to A; namely, the function $\tilde{f}$ that sends b to $f(b)$.

Remark 239. With our definition, two polynomials are equal if they have the same coefficients in the same positions. For example, the two polynomials of $\mathbb{Z}_3[x]$

$$f = x + 2 \qquad \text{and} \qquad g = x^3 + 2$$

are different: the polynomial f corresponds to the sequence (2,1), whereas g corresponds to the sequence (2,0,0,1). However, the two induced evaluations

$$\begin{array}{rcl} \tilde{f} : \mathbb{Z}_3 & \to & \mathbb{Z}_3 \\ b & \mapsto & b+2 \end{array} \qquad \text{and} \qquad \begin{array}{rcl} \tilde{g} : \mathbb{Z}_3 & \to & \mathbb{Z}_3 \\ b & \mapsto & b^3+2. \end{array}$$

are *equal* as functions, because they yield same outputs if they are given same inputs! In fact, by Fermat's Little Theorem (Theorem 130), one has $b^3 \equiv b \mod 3$ for all $b \in \mathbb{N}$.

Degree of a polynomial

Definition 240 (Degree)**.** The degree of a nonzero polynomial f is the maximum index k such that the coefficient of x^k is not zero. If the degree of f is n, we will often refer to a_nx^n as the *leading term* of f, and to a_n as the *leading coefficient* of f.

Example 241. Every constant polynomial has degree zero, except for the zero polynomial, which does not have a degree.

Let us see how the degree behaves with respect of sum and product.

Lemma 242. *Let A be a C-ring with 1. Let f, g be nonzero polynomials in $A[x]$. Then*

$$\textit{either } f + g = 0 \quad \textit{or } \deg(f+g) \leq \max\{\deg f, \deg g\}.$$

If in addition $\deg f \neq \deg g$, then

$$f + g \neq 0 \quad \textit{and } \deg(f+g) = \max\{\deg f, \deg g\}.$$

Proof. Write $f = (a_0, \ldots, a_n)$ with $a_n \neq 0$, and write $g = (b_0, \ldots, b_m)$ with $b_m \neq 0$. Then $n = \deg f$ and $m = \deg g$. Now:

- If $n < m$, then $f + g$ is the polynomial $(a_0 + b_0, \ldots, a_n + b_n, b_{n+1}, \ldots, b_m)$, of degree m.
- If $n > m$, then $f + g$ is the polynomial $(a_0 + b_0, \ldots, a_m + b_m, a_{m+1}, \ldots, a_n)$, of degree n.
- If $n = m$, then $f + g$ is the polynomial $(a_0 + b_0, \ldots, a_m + b_m)$. In this case we cannot be sure that the degree is n, because it could be that $a_m = -b_m$, so that $a_m + b_m = 0$ and the degree is then smaller than m. For this reason, we only claim $\deg(f+g) \leq \max\{\deg f, \deg g\}$. □

Lemma 243. *Let A be a C-ring with 1. Let f, g be non-zero polynomials in $A[x]$. Then*

$$\textit{either } f \cdot g = 0 \quad \textit{or } \deg(f \cdot g) \leq \deg f + \deg g.$$

If we know in addition that A is a domain, then

$$f \cdot g \neq 0 \quad \textit{and } \deg(f \cdot g) = \deg f + \deg g.$$

Proof. As in the previous proof, write $f = (a_0, \ldots, a_n)$ with $a_n \neq 0$, and write $g = (b_0, \ldots, b_m)$ with $b_m \neq 0$. Then

$$f \cdot g = (a_0 b_0,\ a_0 b_1 + a_1 b_0,\ \ \ldots\ ,\ a_n b_m).$$

If A is a domain, from $a_n \neq 0$ and $b_m \neq 0$ it follows that $a_n \cdot b_m \neq 0$, whence $\deg(f \cdot g) = n + m$. If instead A is not a domain, it could be that $a_n \cdot b_m = 0$, in which case the degree is lower, or it could even be that $fg = 0$, in which case fg does not have a degree. □

Example 244. In $\mathbb{Z}_4[x]$, consider the degree-five polynomial $f = 2x^5 + 1$. Then

$$f \cdot f = 4x^{10} + 4x^5 + 1 = 0x^{10} + 0x^5 + 1 = 1.$$

So $\deg(f \cdot f) = 0$, which is lower than $\deg f + \deg f$, and in fact it is even lower than $\deg f$! Of course, this happened because $\mathbb{Z}_4$ is not a domain, and inside such C-ring we have $2 \cdot 2 = 0$.

The proof of the next Lemma is left as exercise:

Lemma 245. *Let A be any C-ring with 1. Let f, g be polynomials in $A[x]$. Suppose that $f \neq 0$ and that the leading coefficient of g is not a zero-divisor. Then $fg \neq 0$ and* $\deg(f \cdot g) = \deg f + \deg g$. *In particular,* $\deg g \leq \deg(f \cdot g)$.

Definition 246 (Monic)**.** A polynomial is called *monic* if its leading coefficient is 1.

Proposition 247. *Let A be any C-ring with 1. All monic polynomials of positive degree (including x, x^2, x^3 etc.) are not invertible. In particular, $A[x]$ is never a field.*

Proof. Let g be a monic polynomial of degree $d > 0$. Let f be an arbitrary polynomial in $A[x]$. Since 1 is invertible, it is not a zero-divisor ((cf. Prop. 227), so by Lemma 245

$$\deg(fg) = \deg f + \deg g = \deg f + d \geq d > 0.$$

In particular $fg \neq 1$, because 1 has degree zero. □

Theorem 248. *Let A be a C-ring with 1. Then*

$$A \text{ is a domain} \quad \Longleftrightarrow \quad A[x] \text{ is a domain}\ .$$

Moreover, if A is a domain, then $\{$*invertible elements of* $A[x]\} = \{$*invertible elements of* $A\}$.

Proof. '⇒'. This is the second part of Lemma 243 above.
'⇐'. Note that $A \subseteq A[x]$, by viewing the elements of A as degree-zero polynomials of $A[x]$. Now let a, b be in A. If $ab = 0$ in A, then $ab = 0$ in $A[x]$, so either $a = 0$ or $b = 0$.
Last claim, "⊇": If $ab = 1$ in A, then since $A \subseteq A[x]$, the equality $ab = 1$ also holds in $A[x]$.
Last claim, "⊆": Suppose $fg = 1$ in $A[x]$. Since A is a domain, by Lemma 243 we have that $\deg f + \deg g = \deg(fg) = \deg 1 = 0$. So $\deg f = \deg g = 0$, that is, both f, g are in A. □

Non-Example 249. The polynomial $f = 2x^5 + 1$ of Example 244 satisfies $f \cdot f = 1$. So

$$\{\text{invertible elements of } \mathbb{Z}_4[x]\} \supsetneq \{\text{invertible elements of } \mathbb{Z}_4\}.$$

There is no contradiction with Theorem 248 though, because $\mathbb{Z}_4$ is not a domain.

Remark 250. One may wonder if the property 'the invertibles of $A[x]$ and of A are the same" occurs *only* if A is a domain. The answer is negative. We will see that in $\mathbb{Z}_6[x]$ the only invertible elements are 1 and 5, which are already in $\mathbb{Z}_6$. Yet $\mathbb{Z}_6$ is not a domain.

4.5 Division of polynomials and cyclicity of U_p

We would like to divide a polynomial by another, as we did with integers, with the idea that the remainder should have lower degree of what we divide by. There is however an immediate problem: in $\mathbb{Z}[x]$, say, how could we possibly divide x^2 by $3x$? If we try to write

$$x^2 = q \cdot (3x) + r, \tag{13}$$

with $\deg r < 1 = \deg(3x)$. But $\mathbb{Z}[x]$ is a domain by Theorem 248, so

$$2 = \deg(x^2) = \deg(x^2 - r) = \deg(q \cdot 3x) = \deg q + \deg(3x) = \deg q + 1.$$

So $\deg q = 1$. Let's rewrite $q = ax + b$, with a, b integers. Equation (13) is an identity of polynomials, which means that the coefficients in the respective degrees should match. Thus from Equation (13) we get a system of three equations in $\mathbb{Z}$:

$$\begin{aligned} 1 &= 3a \\ 0 &= 3b \\ 0 &= r. \end{aligned}$$

But the first equation is already impossible! So the system has no solution: In $\mathbb{Z}[x]$ we cannot divide x^2 by $3x$. And we would face the same problem in $\mathbb{Z}_6[x]$. End of story?! **Wait.** The problem was that in $\mathbb{Z}$ we tried to divide by $3x$, a polynomial whose leading coefficient is not invertible in $\mathbb{Z}$. Since in $\mathbb{Z}$ we cannot divide all numbers by 3, it makes sense that in $\mathbb{Z}[x]$ we cannot divide all polynomials by $3x$...

Theorem 251. *Let A be a C-ring with 1. Let f, g in $A[x]$ be polynomials* such that the leading coefficient of g is invertible. *Then, there exist a unique pair (q, r) in $A[x] \times A[x]$ such that:*

- $f = q \cdot g + r$;
- *either $r = 0$, or $\deg r < \deg g$.*

Proof. Existence. If $\deg f < \deg g$, or if $\deg f$ is undefined because $f = 0$, the claim is obvious: Set $q \stackrel{\text{def}}{=} 0$, $r \stackrel{\text{def}}{=} f$ and we are done. So we can assume that $\deg f \geq \deg g$. Also, if $\deg g = 0$, then g coincides with its leading coefficient, so it is invertible and the claim is again obvious: set $q \stackrel{\text{def}}{=} f \cdot g^{-1}$ and $r \stackrel{\text{def}}{=} 0$. Hence, from now on we can assume that $\deg f \geq \deg g \geq 1$. We proceed by strong induction on the degree of f. Set $n \stackrel{\text{def}}{=} \deg f$, $m \stackrel{\text{def}}{=} \deg g$ and write

$$f = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \quad \text{and} \quad g = b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m.$$

Since we know that b_m is invertible, consider

$$f' \stackrel{\text{def}}{=} f \; - \; a_n \cdot (b_m)^{-1} \cdot x^{n-m} \cdot g.$$

This f' is the difference of two polynomials of same degree and same leading coefficient (a_n). Since the leading terms cancel out, either $f' = 0$, or $\deg f' < n$. Either way, by induction, the theorem holds for the pair (f', g): Namely, there exist q', r' in in $A[x]$ such that $f' = q' \cdot g + r'$, with either $r' = 0$ or $\deg r' < \deg g$. But then

$$\begin{aligned} f \; &= f' + a_n \cdot (b_m)^{-1} \cdot x^{n-m} \cdot g = (q' \cdot g + r') + a_n \cdot (b_m)^{-1} \cdot x^{n-m} \cdot g = \\ &= (q' + a_n \cdot (b_m)^{-1} \cdot x^{n-m}) \cdot g + r'. \end{aligned}$$

If we set $r \stackrel{\text{def}}{=} r'$, $q \stackrel{\text{def}}{=} (q' + a_n \cdot (b_m)^{-1} \cdot x^{n-m})$, then $f = qg + r$, with either $r = 0$ or $degr' < \deg g$, as desired.

UNIQUENESS. Suppose that

$$\begin{aligned} f &= q_1 \cdot g + r_1 \qquad \text{with either } r_1 = 0 \text{ or } \deg r_1 < \deg g, \text{ and also} \\ f &= q_2 \cdot g + r_2, \qquad \text{with either } r_2 = 0 \text{ or } \deg r_2 < \deg g. \end{aligned} \tag{14}$$

We claim that $q_1 = q_2$. From the claim, it follows immediately that $r_1 = f - q_1 g = f - q_2 g = r_2$, which is the desired conclusion. Let us prove the claim by contradiction. Suppose $q_2 - q_1 \neq 0$. From Equations (14) we get $q_1 \cdot g + r_1 = q_2 \cdot g + r_2$, so

$$r_1 - r_2 = (q_2 - q_1) \cdot g.$$

The leading coefficient of g is invertible, so not a zero-divisor (Prop. 227). Hence by Lemma 245

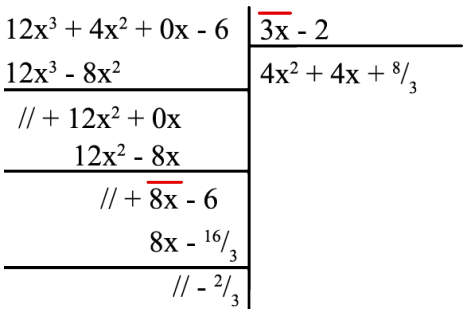
$$\deg(r_1 - r_2) = \deg(q_2 - q_1) + \deg g \ \geq \ \deg g.$$

This contradicts Lemma 242, because each r_i is either 0, or of degree smaller than $\deg g$. □

So to any pair (f, g) of polynomials, with the leading coefficient of g invertible, there is this unique other pair (q, r). But how do we concretely find it? There is an algorithm simply derived from the previous theorem. Suppose you have to divide f by g, so the input is the pair (f, g).

1. Initialize $q \stackrel{\text{def}}{=} 0$.
2. If the leading term of g is not invertible, return an error message and stop.
3. If $\deg g = 0$, output $(f \cdot g^{-1}, 0)$ and stop.
4. If $\deg f < \deg g$, or if $f = 0$, output $(0, f)$ and stop.
5. While $\deg f \geq \deg g \geq 1$:
   - divide the leading term of f by the leading term of g. Let m_1 be the resulting monomial. Replace q by $q + m_1$. (Computer scientists say: "increment q by m_1").
   - Write $-m_1 g$ under f and sum them: this kills the leading term of f. Replace f by the lower-degree polynomial $f - m_1 g$ (or "increment f by $-m_1 g$").
   - Go back to step 4.
6. Output $(q, f - qg)$.

Below is an example of how I graphically compute the division in $\mathbb{Q}[x]$ of $f = 12x^3 + 4x^2 - 6$ by $g = 3x - 2$.



The remainder I obtain in the end is $r = -\frac{2}{3}$ (last line to the left); the quotient is $q = 4x^2 + 4x + \frac{8}{3}$ (last line to the right). Every new iteration of the algorithm produces a new horizontal bar on the left, and a new monomial (of smaller and smaller degree) composing q on the right. Bottom line: There is only one reason I use this notation – namely, because it is consistent with the way I was taught to graphically represent long divisions of integers, when I was in elementary school. Obviously, if you were taught long division of integers in a different way, you should perhaps adopt a way to keep track of the long division of polynomials that is consistent with how you divided integers.

The Euclidean division has several spectacular consequences. Let us start with one.

Definition 252. Let A be a C-ring with 1. Let f be a nonzero polynomial in $A[x]$. An element a of A is called a *root* of f if $f(a) = 0$. (That is, if "plugging in $x = a$ we get an expression that is equal to zero".)

For example, 3 is a root of $x^2 - 5x + 6$, because $3^2 - 5 \cdot 3 + 6 = 0$.

Theorem 253 (Ruffini)**.** *Let A be a C-ring with 1. Let f be a polynomial in $A[x]$. Then for any $a \in A$, we can write*

$$f = (x - a) \cdot q + f(a).$$

In particular,

$$a \text{ is a root of some polynomial } g \iff (x - a) \text{ divides } g.$$

Proof. The leading coefficient of $x - a$ is 1, so we can apply the Euclidean division to f and $g = x - a$: there exist polynomials q, r such that

$$f = (x - a) \cdot q + r$$

and either $r = 0$, or $\deg r < \deg(x - a) = 1$. In both cases, r must be a constant, which remains unchanged if we plug in $x = a$. So let's plug in $x = a$: we get

$$f(a) = 0 \cdot q(a) + r = r.$$

So the remainder of the Euclidean division of f by $x - a$ is exactly $f(a)$. In particular, $f(a) = 0$ if and only if f is a multiple of $x - a$. □

Theorem 254. *Let A be a domain. If $a_1, \dots a_n$ are distinct roots of some nonzero polynomial $f \in A[x]$, then $\deg f \geq n$ and*

$$f = g(x - a_1)(x - a_2) \cdot (x - a_n)$$

for some nonzero polynomial $g \in A[x]$ of degree $\deg f - n$.

Proof. By induction on n. The case $n = 1$ is given by Ruffini's theorem 253: if a_1 is a root of f, then $f = q \cdot (x - a_1)$, with $q \neq 0$ (otherwise $f = 0$). So by Lemma 243 we have $\deg f = \deg q + 1 \geq 1$. Setting $g \stackrel{\text{def}}{=} q$ we are done. Now suppose we have already proven the theorem for every nonzero polynomial with $n - 1$ distinct roots. Let f be a polynomial with n distinct roots, $a_1, \dots, a_n$. By Ruffini's theorem applied to a_n, we have

$$f = q \cdot (x - a_n),$$

and since we are in a domain, $\deg f = \deg q + 1$. Now, if we plug in $x = a_1$, which is a root of f, we get

$$0 = q(a_1) \cdot (a_1 - a_n).$$

But by assumptions $a_1 - a_n \neq 0$ and A is a domain: hence, $q(a_1) = 0$. The same applies also to $a_2, a_3, \dots, a_{n-1}$: we get

$$0 = f(a_i) = q(a_i) \cdot (a_i - a_n),$$

which implies $q(a_i) = 0$. In conclusion, q has $n - 1$ distinct roots. By the inductive assumption, $\deg q \geq n - 1$ and

$$q = g(x - a_1) \cdots (x - a_{n-1})g$$

for some $g \in A[x]$. But then $\deg f = \deg q + 1 \geq (n - 1) + 1 = n$ and

$$f = q(x - a_n) = g(x - a_1) \cdots (x - a_{n-1})(x - a_n).$$ □

Non-Example 255. Consider the polynomial $x^2 - 4$ in $\mathbb{Z}_{12}$. It has degree two, but four different roots: $2, 4, 8, 10$. Note that $(x - 2)(x - 10) = x^2 - 4$, but also $(x - 4)(x - 8) = x^2 - 4$.

Application: Cyclicity of U_p

We conclude with an unexpected application to groups. In Theorem 185 we saw that if G is the product of two cyclic groups with a and b elements, and $\gcd(a,b) \neq 1$, then the period of every element of G divides $s \stackrel{\text{def}}{=} \frac{ab}{d}$. So the polynomial equation $x^s = e$ has more solutions (namely, ab) than its degree. This connects to our new Theorem 254.

Theorem 256 (Gauss' theorem)**.** *For any p prime, the group $(U_p, \cdot) = (\mathbb{Z}_p - \{0\}, \cdot)$ is cyclic.*

Proof. U_p is Abelian, so by Corollary 205 it is isomorphic to a product of finite cyclic groups

$$G \stackrel{\text{def}}{=} \left(\mathbb{Z}_{p_1^{m_1}} \times \mathbb{Z}_{p_2^{m_2}} \times \ldots \times \mathbb{Z}_{p_h^{m_h}},\ +\right).$$

We claim that with the notation above, the p_i's are indeed all distinct, so that

$$p - 1 = p_1^{m_1} p_2^{m_2} \cdots p_h^{m_h}$$

is actually the prime decomposition of $p-1$ into powers of **different** primes. The claim would immediately imply the conclusion that G (and thus U_p) is cyclic, via Corollary 186. So let us prove the claim. By contradiction, assume that two p_i's are the same. Without loss of generality, we can assume $p_1 = p_2$ and $m_1 \leq m_2$. By Theorem 185, any element x of

$$G = \left(\mathbb{Z}_{p_1^{m_1}} \times \mathbb{Z}_{p_1^{m_2}} \times \ldots \times \mathbb{Z}_{p_h^{m_h}}, +\right)$$

has period dividing $t \stackrel{\text{def}}{=} \frac{p-1}{p_1^{m_1}}$. Since the period is maintained under isomorphisms, then also in $(U_p, \cdot)$ every element has period dividing t. In other words, the equation $x^t = 1$ has $p-1$ solutions in U_p. So the polynomial $x^t - 1$ in $\mathbb{Z}_p[x]$ has more roots than its degree, since $t < p-1$ by definition of t. But p is prime, $\mathbb{Z}_p$ is a domain, and so is $\mathbb{Z}_p[x]$ by Theorem 248; so by Theorem 254, no polynomial in $\mathbb{Z}_p[x]$ has more roots than its degree. A contradiction. □

Corollary 257. *Let p be a prime number. In $\mathbb{Z}_p$, the product of two non-squares is a square.*

Proof. By theorem 256, there exists a (nonzero) element $a \in \mathbb{Z}_p$ such that every nonzero element b of Z_p can be written as $b = a^s$, for some $s \in \mathbb{N}$. Clearly, s is even if and only if b is a square. But if $b = a^s$ and $c = a^t$ with s, t both odd, their product is $bc = a^{s+t}$, with $s + t$ even. □

Non-Example 258. $U_8 = \{1, 3, 5, 7\}$ is not cyclic: it is isomorphic to $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$. The degree-two polynomial $x^2 - 1$ has four roots in $\mathbb{Z}_8[x]$, as all invertible elements of $\mathbb{Z}_8$ have square 1. Also, in $\mathbb{Z}_8$ the non-squares 5 and 7 have product 3, which is not a square.

Remark 259. The converse of Gauss' theorem is false: It is not true that if U_m is cyclic, then m is prime. For example if $m = 2p$ with p prime different than 2, by the Chinese remainder theorem one has that U_m is isomorphic to $U_2 \times U_p$, which is the same as U_p.

However, since $|U_m| = m - 1$ implies m prime, one can come up with a converse as follows:

Theorem 260 (Euler–Gauss, Lucas–Lehmer)**.** *For any integer $m \geq 2$,*

$$m \text{ is prime} \quad \Longleftrightarrow \quad \text{some } a \in U_m \text{ has period } m-1.$$

Proof. If m is prime, U_m has $m-1$ elements and is cyclic by Gauss' theorem 256, so there is an element of period $m-1$. Conversely, if m is not prime, then the totient function of m is at most $m-2$, so $|U_m| \leq m-2$ and no element of U_m has period $m-1$. □

Deeper thoughts 261. Gauss proved that U_m is cyclic if and only if m belongs to one of the following sets:

- $\{2, 4\}$;
- odd primes and all their powers;
- the double of a power of an odd prime.

Many integers are not covered by these sets, for example the powers of 2 larger than 4, or composite numbers like 15. Recall that $U_8 = \{1, 3, 5, 7\}$ is isomorphic to $(\mathbb{Z}_2 \times \mathbb{Z}_2, +)$. Exercise for you: Show that U_{12} is also isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_2$, while U_{15} and U_{16} are both isomorphic to $\mathbb{Z}_2 \times \mathbb{Z}_4$.

Deeper thoughts 262. The multiplicative group of invertible elements in $\mathbb{Z}_p$ is cyclic, but our proof leaves no insight on *who the generator is*. Finding the "primitive root", that is, the generator is in general a very difficult problem. Similarly, for U_m (whether m is prime or not) it is hard to find a set of generators of smallest cardinality. A table of smallest generating sets of U_m for $m \in \{2, 3, \ldots, 128\}$ can be found at the link
https://en.wikipedia.org/wiki/Multiplicative_group_of_integers_modulo_n
There is no upper bound on the smallest size of a generating set, because if m is a product of k different odd primes, then it can be seen via the Chinese Remainder Theorem that U_m is generated by at least m generators.

4.6 Exercises

1. Prove that if $A \subsetneq B$ are C-rings with 1, then $A[x] \subsetneq B[x]$.

2. Prove Lemma 245.

3. Prove Theorem 235.

4. Is the converse of Theorem 254 true? Can there be polynomials f of degree one with no roots? What about degree two or higher?

5. Find a gcd of $x^3 - x + 1$ and x^4 in $\mathbb{Z}_2[x]$. Call it $d(x)$. Find polynomials a, b such that
$$d(x) = (x^3 - x + 1)a(x) + x^4 b(x).$$

6. Find a gcd of $x^2 + 1$ and $x^3 + 1$ in $\mathbb{Z}_3[x]$. Then do the same in $\mathbb{Z}_2[x]$.

7. How many polynomials of degree four in $\mathbb{Z}_3[x]$ are there?

8. Let A be a C-ring with 1. Let $B = A[x]$ and $C = A[y]$. Write down an isomorphism between $B[y]$ and $C[x]$.

9. Prove the following stronger version of Fermat's little theorem 130: Let a, m be any integers such that $\gcd(a, m) = 1$. Then m is prime if and only if in $\mathbb{Z}_m[x]$,
$$(x + a)^m = x^m + a.$$
Hint: Use an exercise from Chapter 0: If n is any integer ≥ 2, then
$$n \text{ is prime } \iff \text{ for all integers } k \in \{1, \ldots, n-1\}, \ \binom{n}{k} \text{ is a multiple of } n.$$