Prof. Dr. rer. nat. Antoine Wiehe

Discrete Algebraic Structures

Preface

These lectures notes were written during the winter semester 2023/2024 while I was giving the lectures for the course "Discrete Algebraic Structures" for the first time. The notes contained in their first version imprecisions, typos, and unclear statements. The current version benefited greatly from the help of many students including Johannes Blum, Thilo Droste, Thomas Friese, Alexander Kranz, Alexander Schlüter, Linus Schmidt, David Söding, and Oliver Spitz, who reported on the mistakes. Many thanks to all of them.

Hamburg, February 2024
Antoine Wiehe

Contents

1 Foundations of Mathematics

Things to remember / to know

- Have an informal definition of a set.
- Be able to prove an inclusion $A \subseteq B$ and an equality $A = B$.
- Know the notions of union, intersection, difference, power set.
- Know the properties in Figure 6 and how to prove them.
- Know the notions of injective, surjective, bijective functions and to prove those properties for a given function.
- Know how to compose functions, know how injectivity/surjectivity and composition interact (Lemma 1.18).

1.1 Sets

A *set* is, informally, a mathematical box that can contain other mathematical objects. The question of what formally counts as a set is hard to answer so we remain vague about this. We write a set by enclosing in curly braces { and } the elements that belong to the set, separated by commas. This is called an *extensional definition* of the set. For example, $\{0, 1, -1\}$ is a set with 3 elements, containing the numbers $0, -1, 1$. The order of the elements does not matter, as well as the number of times an element appears. So the sets $\{0, 1, -1\}$ and $\{1, 0, 1, -1\}$ are equal.

set

extensional definition

We have standard notations for some sets of numbers that we encounter regularly:

- $\mathbb{N}$: all the counting numbers $\mathbb{N} = \{1, 2, 3, \dots\}$,
- $\mathbb{N}_0$: counting numbers from 0, $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$
- $\mathbb{Z}$: all the integer numbers $\mathbb{Z} = \{0, 1, -1, 2, -2, \dots\}$,
- $\mathbb{Q}$: the rational numbers $\mathbb{Q} = \{0, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{3}{4}, 1, \dots\}$,
- $\mathbb{R}$: real numbers $\mathbb{R} = \{0, 1, \sqrt{2}, \pi, e, \dots\}$
- $\mathbb{C}$: complex numbers

Sets don't have to contain only things of the same "type": $\{0, \{1, 2\}\}$ is a set whose elements are 0 (a number) and $\{1, 2\}$ (itself a set). This set has 2 elements.

Notation 1.1. We use the notation $a \in A$ to say that a is an element of A, and $a \notin A$ to say that a is not an element of A.

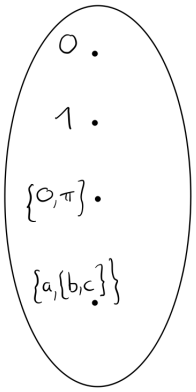


Figure 1: A representation of the set $\{0, 1, \{0, \pi\}, \{a, \{b, c\}\}\}$.

Sets are usually represented by drawing "bubbles" (another accepted technical term for it is "potatoes") with dots inside representing the elements. See Figure 1 for an example.

Definition 1.2. We say that two sets A and B are *equal if, and only if,* they contain the same elements. That is, $A = B$ exactly when for every a, we have $a \in A$ if, and only if, $a \in B$.

This might look completely harmless, but this is is a very subtle difference compared to some programming languages.

For example, in Python, we have the following:

```
A = {0,1,2}
B = {0,1,2}
print(A is B) # prints False
```

Even if the two sets contain the same elements, A *is not* B because they are stored at different places in the computer's memory.

A set is *empty* if it contains nothing. Thus, by Definition 1.2, there is a *unique* empty set, which is denoted by $\emptyset$.

empty set

Definition 1.3. We say that B is a *subset* of A if, for every a, we have that if $a \in B$ then $a \in A$. This is denoted by $B \subseteq A$. We say that B is a *proper subset* if $B \subseteq A$ and there exists an element $a \in A$ that is not in B. We can write $B \subsetneq A$ or $B \subset A$ if B is a proper subset of A.

subset

proper subset

Note that $\emptyset \subseteq A$ and $A \subseteq A$ are true for any set A.

1.2 Constructing new sets

1.2.1 Set-builder notation

When a set has infinitely many elements, one cannot simply list all of them. So we use an *intentional definition* of the elements. For this, we use the notation $\{x \in A \mid$ some property about $x\}$. This is called the *set-builder notation*.

intentional definition

set-builder notation

Example 1.4. The set of even numbers can be written as $\{n \in \mathbb{N} \mid n \text{ is divisible by } 2\}$. Another example is $\{n \in \mathbb{N} \mid n \geq 6 \text{ and } n \text{ is odd}\}$, which is the set $\{7, 9, 11, 13, 15, \dots\}$.

On the left side of $|$, one must *always* have $x \in \dots$ and not just x. This is very important, because of *Russell's paradox*: imagine we allow to write $\{x \mid x \notin x\}$. If this were a "legitimate" set, call it A, then we would run into a contradiction:

Russell's paradox

- If A belongs to A (i.e., if $A \in A$), then by definition of A it should be the case that $A \notin A$, contradicting our assumption!
- If $A \notin A$, then A should be in A, i.e., $A \in A$, again a contradiction.

Such problems were understood at the end of the 19th century and the bugfix applied by mathematicians was to enforce the rule mentioned above.

1.2.2 Common constructions

Given sets A, B, there are many ways of obtaining a new set.

Definition 1.5. Let a, b be two "things". Then (a, b) is an *ordered pair*. If $a_1, \dots, a_n$ are "things", then $(a_1, \dots, a_n)$ is an *n-tuple*.

ordered pair

tuple

A pair (a, b) is *ordered*, it matters which is the first component and which is the second. So $(a, b) \neq (b, a)$, unless $a = b$. Similarly, tuples are ordered lists of elements.

Definition 1.6. Let A and B be two sets. Then the following are also sets:

- $A \times B$ is the set containing all the pairs (a, b), where $a \in A$ and $b \in B$. This is called the *product* of A and B. If $A = B$, we write A^2 instead of $A \times A$.

  product

- $A \cup B$ is the set containing the elements of A and of B: $x \in A \cup B$ if, and only if, $x \in A$ *or* $x \in B$. This is called the *union* of A and B.

  union

- $A \setminus B$ is the set containing the elements that are in A and not in B. This is the *difference* of A and B, see Figure 2. In set-builder notation, this can be written as $\{x \in A \mid x \notin B\}$.

  difference

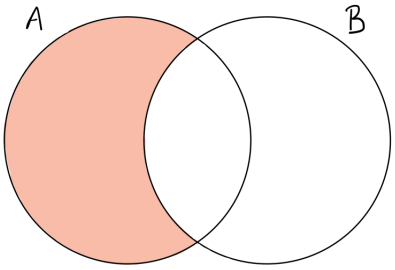


Figure 2: $A \setminus B$

- $A \cap B$ is the set containing the elements that are both in A and B: $x \in A \cap B$ if, and only if, $x \in A$ *and* $x \in B$. This is called the *intersection* of A and B, see Figure 3. In set-builder notation, we can write $A \cap B$ as either $\{x \in A \mid x \in B\}$ or as $\{x \in B \mid x \in A\}$. We say that A and B are *disjoint* if $A \cap B = \emptyset$.

intersection

disjoint

- $A \Delta B$ is the set containing the elements that are either in A or in B but not both. That is, $x \in A \Delta B$ if, and only if, $x \in A \cup B$ and either $x \notin A$ or $x \notin B$. This is the *symmetric difference* of A and B, see Figure 4. In set-builder notation, this is written $A \Delta B = \{x \in A \cup B \mid x \notin A \text{ or } x \notin B\}$.

symmetric difference

- $\mathcal{P}(A)$ is the set such that $C \in \mathcal{P}(A)$ if, and only if, $C \subseteq A$. This is called the *power set* of A.

power set

Remark 1.7. Note that we don't give a definition of $A \cup B$ or $\mathcal{P}(A)$ in terms of set-builder notation, in fact this is impossible. One has to simply accept that those sets exist, we say their existence is an *axiom*. One could give a set-builder definition of $A \times B$, but this is not important for this course.

Venn diagram

To visually understand all these concepts, we typically draw *Venn diagrams*: such diagrams are obtained by drawing circles (aka potatoes) for each of the sets that are involved, and we draw those potatoes in a way that they intersect as much as possible. See for example Figure 5, where we draw one potato for each of the sets A, B, C in a way that all possible intersections $(A \cap B) \setminus C, (A \cap C) \setminus B, (B \cap C) \setminus A$, and $A \cap B \cap C$ appear on the diagram. We typically only draw Venn diagrams with at most 3 sets.

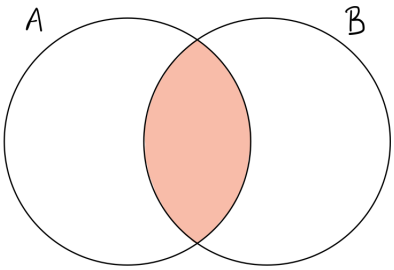


Figure 3: $A \cap B$

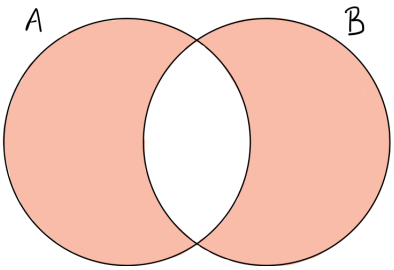


Figure 4: $A \Delta B$

Example 1.8 (Power set). If $A = \{1, 2, 3\}$, then $\mathcal{P}(A)$ contains the following elements: $\emptyset$, $\{1\}$, $\{2\}$, $\{3\}$, $\{1, 2\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2, 3\}$. In general, $\emptyset$ and A are always elements of $\mathcal{P}(A)$.

One can generalize the operations of product, intersection, and union to more than just 2 sets:

- The product $A \times B \times C$ is the set of triples (a, b, c) where $a \in A, b \in B, c \in C$. More generally, the product $A_1 \times \cdots \times A_n$ is the set of n-tuples $(a_1, \ldots, a_n)$ where $a_1 \in A_1, \ldots, a_n \in A_n$. We also write $\prod_{i=1}^{n} A_i$ instead of $A_1 \times \cdots \times A_n$ and A^n instead of $A \times A \times \cdots \times A$.
- If $A_1, \ldots, A_n$ are sets, then $\bigcap A_i$ is the set of elements a such that a is in A_i for *every* $i \in \{1, \ldots, n\}$.
- If $A_1, \ldots, A_n$ are sets, then $\bigcup A_i$ is the set of elements a such that a is in A_i for *at least one* $i \in \{1, \ldots, n\}$.

Here we see some example of the properties that these operations have. This is also our first example of a *proof*.

Proposition 1.9 (Distributivity property)**.** *Let A, B, and C be sets. Then $A \cap (B \cup C)$ is equal to $(A \cap B) \cup (A \cap C)$.*

See Figure 5 for a Venn diagram representing that statement.

Proof. By Definition 1.2, in order to prove that $A \cap (B \cup C)$ is equal to $(A \cap B) \cup (A \cap C)$ we must prove that x is in one of the sets if, and only if, it is in the other. In other words, we must prove that both $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$ and $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$ are true.

We first prove the first inclusion. For this, we take an arbitrary element x in $A \cap (B \cup C)$. By the definition of $\cap$, this means that x is in A and x is in $B \cup C$. By applying the definition of $B \cup C$, we obtain that x is in B or x is in C.

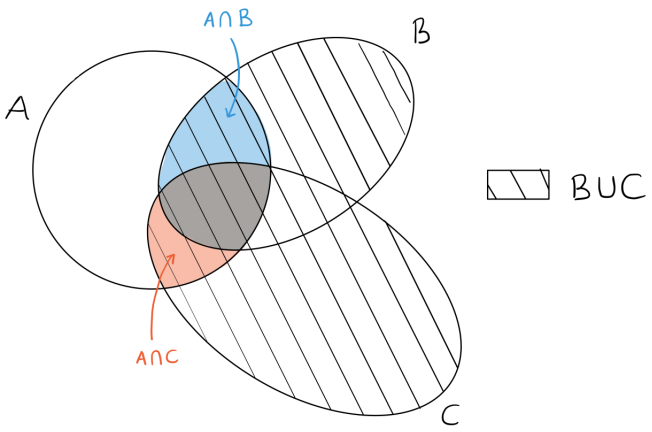


Figure 5: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

- Assume we are in the first case, i.e., x is in B. Then x is in $A \cap B$ by definition of $\cap$, and therefore x is in $(A \cap B) \cup (A \cap C)$ by definition of $\cup$.
- Assume we are in the second case, i.e., x is in C. Then x is in $A \cap C$ by definition of $\cap$, and therefore x is in $(A \cap B) \cup (A \cap C)$ by definition of $\cup$.

Thus, in both cases, we obtain that x is in $(A \cap B) \cup (A \cap C)$, so that every element of $A \cap (B \cup C)$ is an element of $(A \cap B) \cup (A \cap C)$.

We now prove the inclusion $(A \cap B) \cup (A \cap C) \subseteq A \cap (B \cup C)$. Take an arbitrary element x in $(A \cap B) \cup (A \cap C)$. By definition of $\cup$, it must be that x is in $A \cap B$ or in $A \cap C$, so again we distinguish two cases:

- Assume first that x is in $A \cap B$, so that by definition of $\cap$ we get that x is in A and x is in B. By definition of $\cup$, we have that x is in $B \cup C$, so that we obtain that x is in $A \cap (B \cup C)$.
- The case where x is in $A \cap C$ is similar: we have $x \in A$ and x in C, so $x \in B \cup C$, and therefore $x \in A \cap (B \cup C)$.

Since in both cases we obtained that x is in $A \cap (B \cup C)$, we finished the proof of the inclusion, and this concludes the proof. □

Remark 1.10. This is an example of a *direct proof*: we started from the assumptions (here, simply that A, B, C are sets), and we proved that the conclusion is true. When you are asked to prove an equality between two sets (in this example, $A \cap (B \cup C)$ and $(A \cap B) \cup (A \cap C)$), you can always proceed in this way by *double inclusion*.

direct proof

For all sets A, B, C, the following equalities are true:

$$\begin{array}{ll} A \cap (A \cup B) = A & A \cup (A \cap B) = A \\ A \cap (B \cup C) = (A \cap B) \cup (A \cap C) & A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \\ A \setminus (B \cup C) = (A \setminus B) \cap (A \setminus C) & A \setminus (B \cap C) = (A \setminus B) \cup (A \setminus C) \end{array}$$

Figure 6: Important properties of set operations

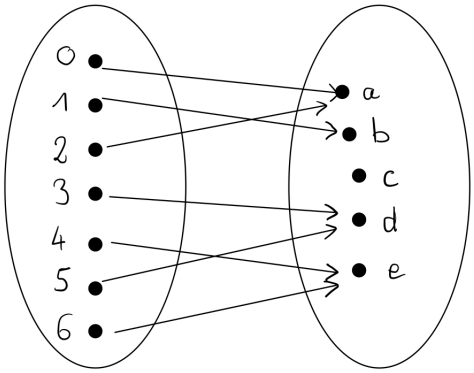


Figure 7: Representing a function as arrows between potatoes.

Similarly to Proposition 1.9, there are a number of important properties relating the operations of intersection, union, and difference. These properties are summarized in Figure 6. We do not provide a proof of these equalities here and the proofs are left as an elementary exercise.

1.3 Functions

Definition 1.11. A *function* from A to B, denoted by $f\colon A \to B$, is a subset $f \subseteq A \times B$ with the property that for every $a \in A$, there exists a unique $b \in B$ such that $(a, b) \in f$. Since b is uniquely determined by a, we write $f(a) = b$, and we say that b is the *image* of a under f, and that a is *a preimage* of b under f.

function

preimage of an element under a function

If A and B are finite and $f\colon A \to B$, we can represent each of A and B as a potato, and we put an arrow from a to b if $f(a) = b$. See Figure 7 for an example of how one might draw a function. In that example, we have for example $f(0) = a, f(1) = b, f(2) = a, \ldots$

Remark 1.12. Note the choice of articles: we say that b is *the* image of a, as there is only one b such that $(a, b) \in f$ by the definition that f is a function. However, a is *a* preimage of b, as there could be another $a' \in A$ such that $(a', b) \in f$, i.e., such that $f(a') = b$. We write $f^{-1}(\{b\})$ for the set of all $a \in A$ such that $f(a) = b$. In set-builder notation, we have $f^{-1}(\{b\}) = \{a \in A \mid f(a) = b\}$.

Definition 1.13. Let A be any set. The function $f\colon A \to A$ such that $f(a) = a$ for all $a \in A$ is called the *identity function on* A and is often written id_A.

identity function

Definition 1.14. A function $f\colon A \to B$ is *injective* if for every $a \neq a'$ in A, then $f(a) \neq f(a')$. A function is *surjective* if for every $b \in B$, there exists $a \in A$ such that $f(a) = b$. A function is *bijective* if it is both injective and surjective.

injectivity

surjectivity

bijectivity

Another way to phrase these properties:

- f is injective if for every $b \in B$, there exists *at most one* $a \in A$ such that $f(a) = b$;
- f is surjective if for every $b \in B$, there exists *at least one* $a \in A$ such that $f(a) = b$;
- f is bijective if for every $b \in B$, there exists *exactly one* $a \in A$ such that $f(a) = b$.

Remark 1.15. One also say that f is *one-to-one* if it is injective, and that it is *onto* if it is surjective.

one-to-one

onto

When considering functions $f\colon \mathbb{R} \to \mathbb{R}$, one way to remember what injective/surjective means:

- injective means that any horizontal ray intersects the graph of the function *at most once*,
- surjective means that any horizontal ray intersects the graph of the function *at least once*.

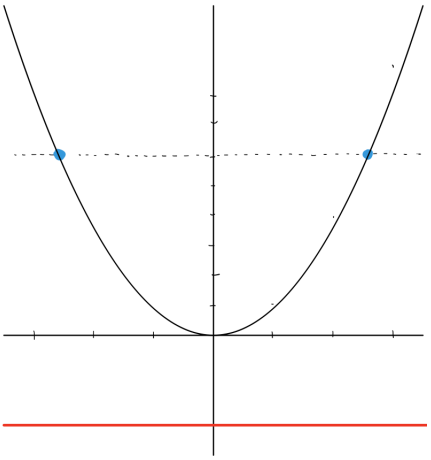


Figure 8: The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is neither injective (because of the two blue dots) nor surjective (the curve does not intersect with the red line).

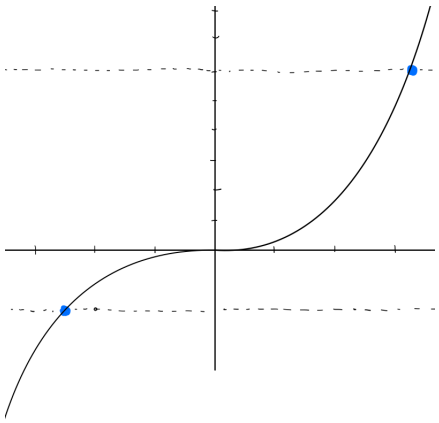


Figure 9: The function $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g(x) = x^3$ is both injective and surjective.

Example 1.16. Here are examples:

- The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) := x^2$ is not injective, since $1 \neq -1$ but $f(1) = f(-1)$.
- It is also not surjective, because there is no $a \in \mathbb{R}$ such that $f(a) = -1$.
- The function $\arctan\colon \mathbb{R} \to \mathbb{R}$ is injective but not surjective, since $|\arctan(x)| \in (-\pi/2, \pi/2)$. However, it is surjective if we consider it as a function $\mathbb{R} \to (-\pi/2, \pi/2)$, see Figure 10.
- Similarly, $f\colon \mathbb{N} \to \mathbb{N}$ defined by $f(x) = 2x$ is injective but not surjective (since, for example, there is no x such that $f(x) = 3$).

Definition 1.17 (Function composition)**.** Let $f\colon A \to B$ and $g\colon B \to C$ be functions. The *composition of f and g*, denoted by $g \circ f\colon A \to C$, is the function defined by $(g \circ f)(x) = g(f(x))$.

composition of two functions

Lemma 1.18. *Let $f\colon A \to B$ and $g\colon B \to C$ be injective functions. Then $g \circ f$ is injective. The same is true if one replaces injective by surjective.*

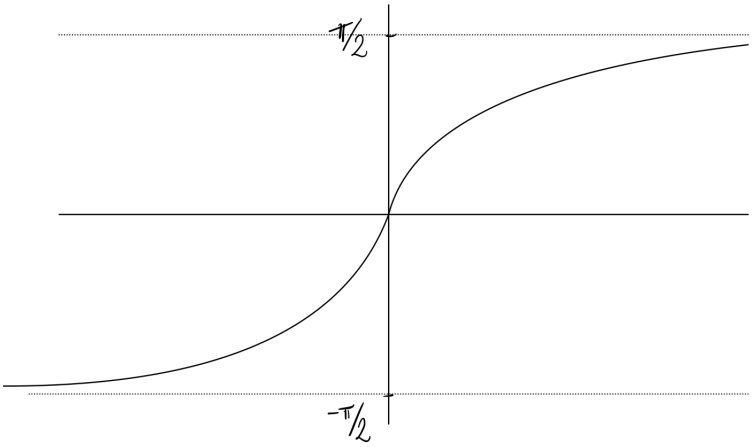


Figure 10: Graph of the function arctan, a bijection from $\mathbb{R}$ to $(-\pi/2, \pi/2)$.

Proof. We first do the proof for the injective case. Let $a, a' \in A$ be different. Since f is injective by assumption, we have that $b := f(a)$ and $b' := f(a')$ are different. Since g is injective, we have $g(b) \neq g(b')$, so that $g(f(a)) \neq g(f(a'))$. By rewriting using the definition, this means $(g \circ f)(a) \neq (g \circ f)(a')$, which is what we wanted to prove.

Now let us turn to the surjective case. Let $c \in C$. To prove that $g \circ f$ is surjective, we want to show that there exists $a \in A$ such that $(g \circ f)(a) = c$. Since g is surjective, there exists $b \in B$ such that $g(b) = c$. Since f is surjective, there exists $a \in A$ such that $f(a) = b$. We obtain $(g \circ f)(a) = g(f(a)) = g(b) = c$, which is what we wanted. □

Proposition 1.19. *Let $f\colon A \to B$ be a bijective function. Then there exists a function $g\colon B \to A$ such that $g \circ f = id_A$ and $f \circ g = id_B$. Moreover, if $h\colon B \to A$ satisfies the same properties, then $g = h$.*

Proof. We need to define the function $g\colon B \to A$. For this, fix an arbitrary $b \in B$. Since f is surjective, there exists an $a \in A$ such that $f(a) = b$; we now define $g(b) := a$. Note in particular that $f(g(b)) = f(a) = b$. The function g is now defined, and we already see that it satisfies $f(g(b)) = b$ for all $b \in B$ so that in particular $f \circ g = \mathrm{id}_B$.

We now prove that $g \circ f$ is the identity function on A. Rephrasing what this means, we must prove that for every $a \in A$, we have $g(f(a)) = a$. Let $a \in A$ be arbitrary. We let $b := f(a)$. By the previous paragraph, we already know that $f(g(b)) = b = f(a)$. Since f is injective, the definition of injectivity gives that $g(b) = a$. And since $b = f(a)$, this can be rewritten as $g(f(a)) = a$, which is exactly what we wanted to prove.

Let now h be another function satisfying the same properties as g. For any $b \in B$, let a be the preimage of b under f (so $f(a) = b$). Then $h(b) = h(f(a)) = (h \circ f)(a) = a = (g \circ f)(a) = g(b)$. Therefore, $h(b) = g(b)$ for all $b \in B$. □

The function g in Proposition 1.19 is called the *inverse* of f, and is denoted by f^{-1}. Note that this is not always defined! According to Proposition 1.19, we know this exists when f is a bijection. What we see next in Lemma 1.21 is that f^{-1} exists precisely when f is a bijection.

inverse of a function

Remark 1.20. Pay attention to the fact that for a function $f\colon A \to B$ and $b \in B$, the notation $f^{-1}(\{b\})$ is *always* defined, but it is in general a subset of A (once more, it is the set of all preimages of b under f). When f is bijective, then $f^{-1}(b)$ is used to mean *the only* $a \in A$ *such that* $f(a) = b$.

Lemma 1.21. *Let $f\colon A \to B$. Assume that there exists $g\colon B \to A$ such that $f \circ g = id_B$ and $g \circ f = id_A$. Then f is bijective.*

Proof. id_B is surjective,[1] so $f \circ g$ is surjective. By Exercise 9, we have that f is surjective. Similarly, id_A is injective,[2] so $g \circ f$ is injective. By Exercise 8, we have that f is injective, and therefore f is a bijection. □

Let $f\colon A \to B$ be a function. We call A the *domain* of f, and B is the *codomain* of f. For any $C \subseteq A$, we write $f(C)$ as the set $\{b \in B \mid \text{there exists } c \in C \text{ such that } f(c) = b\}$, and $f(A)$ is the *image* of f.[3] Thus, f is surjective if and only if $f(A) = B$. For $D \subseteq B$, we define $f^{-1}(D)$ as the set $\{a \in A \mid f(a) \in D\}$.

codomain

image of a function

1.4 Exercises

1. Prove that $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

2. Prove that $A \cap (A \cup B) = A$.

3. Prove that if $B \subseteq A$, then $A \cap B = B$ and $A \cup B = A$.

4. Prove that $A \setminus B \subseteq C$ if, and only if, $A \subseteq B \cup C$.

5. Prove that $A \Delta B = (A \cup B) \setminus (A \cap B)$.

6. Prove that $A \setminus (A \setminus B) = A \cap B$.

7. Let f be the function depicted in Figure 7. Compute $f^{-1}(a), f^{-1}(b), f^{-1}(c), f^{-1}(d), f^{-1}(e)$.

8. Show that if $g \circ f$ is injective, then f is injective.

9. Show that if $g \circ f$ is surjective, then g is surjective.

[1] It is in fact bijective, but we only need the surjectivity here.
[2] It is in fact bijective, but we only need the injectivity here.
[3] In your linear algebra course, if f is a linear map between vector spaces this is sometimes denoted by $\mathrm{im}(f)$.

10. Show that if $f\colon A \to B$ is injective and A and B are finite, $A \neq \emptyset$, then there exists $g\colon B \to A$ that is surjective and such that $g \circ f = \mathrm{id}_A$.

11. Show that if $g\colon A \to B$ is surjective and A and B are finite, then there exists $f\colon B \to A$ injective such that $g \circ f = \mathrm{id}_B$.

12. For the next few exercises, f is a function $f\colon A \to B$. Show that if $C, D \subseteq B$, then $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$. Show that the same holds for $\cup$ instead of $\cap$.

13. Show that if $C, D \subseteq A$, then $f(C \cap D) \subseteq f(C) \cap f(D)$. Find an example of a function $f\colon A \to B$ and sets $C, D \subseteq A$ such that $f(C \cap D) \neq f(C) \cap f(D)$.

14. (*) Show that for any $D \subseteq B$, $f(f^{-1}(D)) \subseteq D$, and give an example showing that $f(f^{-1}(D)) \neq D$ in general. Think about how this relates to f being or failing to be surjective.

15. (*) Show that for any $C \subseteq A$ it is true that $C \subseteq f^{-1}(f(C))$. Show that $C \neq f^{-1}(f(C))$ in general. Think about how this relates to f being or failing to be injective.

16. Show that for any $C, D \subseteq B$ such that $C \subseteq D$, then $f^{-1}(C) \subseteq f^{-1}(D)$.

2 Mathematical Logic

Things to remember / to know

- Know the notion of quantifiers and boolean connectives, and how to write propositional formulas and predicates.
- Be able to turn a sentence in English/German into a mathematical formula.
- Know the notion of truth table.
- Know the notion of logical equivalence and how to prove that two formulas are logically equivalent.
- Know the notion of tautology and satisfiability, and how they relate.
- Know the notion of proofs by contradiction, contraposition, and induction.
- Know the De Morgan's rules.

2.1 Statements

In mathematics, a statement usually has the form "If *some hypotheses* are met, then *some conclusion*." For example, the sentence "Hamburg is a country" and "Hamburg is a city" are both statements (without any hypotheses), and "If it rains outside, then I take an umbrella" is a statement whose hypothesis is "it rains outside" and whose conclusion is "I use an umbrella." Expressing statements in a natural language can lead to ambiguities: think about the sentence "Free entrance for students of computer science and data science." Does this apply to people studying both? Or at least one? How does it compare to "Free entrance for students of computer science or data science"? This ambiguous meaning is why for example we sometimes see "and/or" used in texts.

Mathematicians invented a precise language to clear up any possible confusion. Like any language, it has a *syntax* and a *semantics*. The syntax is the grammar, it says that we are allowed to write. The semantics is what gives a meaning to the things that we write.

2.2 Propositions and predicates: the grammar of the language

To formalize the notion of statements, we use letters such as A, B, C, P, Q, R as variables to denote any kind of statement, which could either take the value True or the value False. These are called *propositional variables* or sometimes also *atomic propositions*.[4] Then, we translate typical conjunctions words like "or", "and" and "implies" into mathematical symbols. The word "or" is written as $\vee$, "and" is written $\wedge$, and "not" is written $\neg$. For example, A could mean "It rains outside", and B could mean "I use an umbrella": those

propositional variables

[4] The word "atomic" means that those propositions cannot be decomposed further.

are atomic propositions. One would translate the sentence "If it rains outside then I use an umbrella" by $A \Rightarrow B$ (pronounced A *implies* B), while "It does not rain outside or I use an umbrella" is written $\neg A \vee B$ (pronounced "not A or B").

We now give a formal definition.

Definition 2.1. A *propositional formula* is an expression built from propositional variables and such that if P and Q are propositional formulas, then $P \wedge Q$, $P \vee Q$, $P \Rightarrow Q$, and $\neg P$ are also propositional formulas.

propositional formula

The statement "It rains outside" is a *closed* statement, it does not depend on anything. But typically in mathematics, one wants to make statements that can be true or false depending on some parameters. For example, whether the statement "n is an odd number" is true depends on n: we say that it is a *predicate*, and we would denote such a predicate by $P(n)$ instead of just P, to emphasize that it depends on a yet unspecified value of n.

predicate

Another important logical construction in every day language but also in mathematics is the notion of *quantifiers*. For example, in the sentences "every cat is an animal" or "some cats have three legs," the quantification is what lets us know whether the statement is *universal* (talking about all objects of some sort) or *existential* (simply stating that at least one object exists).

quantifier

Let us see a mathematical example. Consider the mathematical statement "For every natural number n, if n is even then n^2 is even." In symbols, one would write

$$\forall n \in \mathbb{N}(n \text{ is even} \Rightarrow n^2 \text{ is even}).$$

The symbol $\forall$ is read "Every" or "For all" and is called *universal quantifier*. And we translate "some" or "there exists" by the *existential quantifier* $\exists$. A complete translation table is in Figure 11.

universal quantifier

existential quantifier

Natural language	Mathematical language
Every/All element $x \in A$ satisfies P	$\forall x \in A\ P(x)$
At least one/There exists some $x \in A$ satisfying P	$\exists x \in A\ P(x)$
Both P and Q are true	$P \wedge Q$
At least one of P and Q is true	$P \vee Q$
P is false	$\neg P$
If P is true, then Q is true	$P \Rightarrow Q$

Figure 11: Translation from natural language to mathematical language.

Following Figure 11, the statement "Everything that is a planet is inhabited" could be translated as $\forall p \in$ Planets (p is inhabited), while the statement "There exists a planet different from Earth that is inhabited" can be translated as $\exists p \in$ Planets ($p \neq$ Earth $\wedge\ p$ is inhabited), where we assume that Planets is the set containing all planets.

For a more mathematical statement, consider for example

$$\forall n \in \mathbb{N}(n \text{ is prime} \Rightarrow n \text{ is odd})$$

(which we can see is false, even if we haven't formally defined the semantics of the mathematical language yet),

$$\forall n \in \mathbb{N}(n \text{ is prime} \Rightarrow (n \text{ is odd} \vee n = 2)),$$

or

$$\forall f\colon A \to B\ \forall g\colon B \to C\ ((f \text{ is injective} \wedge g \text{ is injective}) \Rightarrow g \circ f \text{ is injective})$$

(which we can imagine are true). Notice the generous use of brackets, which are there to make sure that there is no ambiguity in reading the formula.

2.3 Truth tables: the meaning of the language

So far we only described the syntax of the logical language and roughly explained what it corresponds to in an every day language, but there is no meaning to the formulas yet. We know that $(A \wedge B) \vee C$ is a formula and so is $(A \Rightarrow (B \vee C))$, but what do they mean and when are they true?

The symbols $\wedge, \vee, \neg$ are called *boolean connectives*,[5] and their logical function can be summarized by their *truth tables* (see Figure 12): for every value of P and Q, we record in a table what is the value of $P \wedge Q$ (and similarly for $P \vee Q, P \Rightarrow Q, \neg P$). In maths, the meaning of "or" is *inclusive*: for example, one would consider that the propositional statement "I ate a cake or a fruit" is true, even after eating a cake *and* a fruit. But in every day language, for example on the menu of a restaurant, "cheese or dessert" definitely means that you have to choose one of the two. We say that this type of "or" is *exclusive*.

boolean connectives

truth table

inclusive or

exclusive or

The truth table for $P \Rightarrow Q$ might also be surprising at first: mathematicians consider that the statement "If red is a smell then 5 is a vegetable" is true, since the hypothesis "red is a smell" is false. This is the principle known as *ex falso quodlibet* or also *principle of explosion*: from a false hypothesis, one can deduce everything. Mathematicians also say that "If Hamburg is a city then red is a color" is true, even if the hypothesis and the conclusion are completely unrelated!

principle of explosion

We can generalize this and consider truth tables of any expression involving boolean connectives. We have one column for each propositional variable, and then one column for each intermediate expression, and finally a column for the final result. For example, the expression $P \wedge (Q \vee R)$ has three propositional variables, and $Q \vee R$ is an intermediate expression. The truth table of $P \wedge (Q \vee R)$ is then the following:

[5] After George Boole, a British mathematician from the 19th century who laid the foundations for an algebraic treatment of logic.

P	Q	$P \wedge Q$
False	False	False
True	False	False
False	True	False
True	True	True

P	Q	$P \vee Q$
False	False	False
True	False	True
False	True	True
True	True	True

P	Q	$P \Rightarrow Q$
False	False	True
True	False	False
False	True	True
True	True	True

P	$\neg P$
False	True
True	False

Figure 12: Truth tables of the boolean connectives

P	Q	R	$Q \vee R$	$P \wedge (Q \vee R)$
False	False	False	False	False
True	False	False	False	False
False	True	False	True	False
True	True	False	True	True
False	False	True	True	False
True	False	True	True	True
False	True	True	True	False
True	True	True	True	True

In computers, the values False and True are usually encoded as 0 and 1.

2.4 Logical Equivalence

Definition 2.2. Let P and Q be two propositional formulas that have the same propositional variables $A_1, \ldots, A_n$. We say that P and Q are *logically equivalent* if the truth tables for P and Q have an identical last column. We write $P \equiv Q$ to say that P and Q are logically equivalent.

logical equivalence

This means that for all statements one could imagine to replace the propositional variables in P and Q, both P and Q have the same meaning. For example, $A \vee B$ and $B \vee A$ are logically equivalent, which one can check by looking at the truth table of $\vee$. This means that whatever A and B mean, the sentences we would get in both cases have exactly the same meaning. A non-trivial example is the following:

Proposition 2.3. *$\neg A \vee B$ is logically equivalent to $A \Rightarrow B$.*

Proof. We simply need to compare the truth tables of the two expressions. Here is the one for $\neg A \vee B$ (there are two variables, and one intermediate expression $\neg A$):

A	B	$\neg A$	$\neg A \vee B$
False	False	True	True
True	False	False	False
False	True	True	True
True	True	False	True

We see that the last column agrees with the last column in the truth table for $A \Rightarrow B$. □

Remark 2.4. We won't formalize this here, but as far as mathematicians know, the *only* general way to check whether two statements are logically equivalent is to build their truth tables and check that they give the same result. Even with only 10 propositional variables $A_1, \ldots, A_{10}$, we would have to build a truth table with more than 1000 rows! This is related to the Millenium problem "P vs. NP," [a] one of the biggest open questions in computer science. There are several courses at TUHH dedicated to such questions: *Computability and Complexity Theory* (given by Prof. Kliesch at the Bachelor's level), and the two courses *Advanced Complexity* and *Constraint Satisfaction Problems* given by myself at the Master's level.

[a] https://www.claymath.org/millennium/p-vs-np/

Definition 2.5. Let P be a proposition. We say that it is:

- a *tautology* if the last column of its truth table only contains True.
- *satisfiable* if the last column of its truth table contains True at least once.

tautology

satisfiable

Remark 2.6 (Law of excluded middle). The proposition $A \vee \neg A$ is a tautology. We show this by building the truth table (there is one propositional variable, and one intermediate expression $\neg A$):

A	$\neg A$	$A \vee \neg A$
False	True	True
True	False	True

and we see that the last column always contains True. The fact that his statement is a tautology is called the *law of excluded middle*: there is no option between "A is true" and "A is false."

law of excluded middle

The following statements relates the notions of logical equivalence and of tautology.

Proposition 2.7. *Two propositions P, Q are logically equivalent if, and only if, $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ is a tautology.*

Proof. There are two directions to prove.

Suppose first that P, Q are logically equivalent. This means that in their respective truth tables, the last column receives exactly the same value. Therefore, whenever the column for P is True, then so is the column for Q. This means that $P \Rightarrow Q$ is True.

Similarly, whenever Q is True, then the column for P is True, so that $Q \Rightarrow P$ is True. It follows that $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ is always True.

The other direction is similar. Let $A_1, \ldots, A_n$ be the propositional variables in P and Q. Suppose that for some True/False values of $A_1, \ldots, A_n$, the corresponding value of P is True. Since $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$ is always True, we have in particular that $P \Rightarrow Q$ is True, and from the truth table of $\Rightarrow$ we conclude that Q must be True (for this evaluation of $A_1, \ldots, A_n$). Symmetrically, if Q is True for some evaluation then P must be True as well. Therefore, the last columns of the truth tables of P and Q are identical, so that P and Q are logically equivalent. □

Notation 2.8. We write $P \Leftrightarrow Q$ as a shortcut for $(P \Rightarrow Q) \wedge (Q \Rightarrow P)$.

The concepts of tautologies and satisfiable formulas are also related:

Proposition 2.9. *Let P be a proposition. Then P is a tautology if, and only if, $\neg P$ is* not *satisfiable.*

2.5 Simplification of Negation

In natural language, we would typically never say "Not every animal is a dog". We have the intuition that "not every animal is ..." really means "some animal is not ...". So "not every animal is a dog" is equivalent to "some animal is not a dog." Similarly, "there is no cat that is a fish" is equivalent to "every cat is not a fish." We have similar simplification rules in the mathematical language, see Figure 13.

The first two rows in Figure 13 are called De Morgan's laws[6] and are repeated in the following lemma. We don't give a proof for the last two rows and rather rely on intuition. We will see a generalization of this fact and its proof in Lemma 6.47.

Lemma 2.10 (De Morgan's laws)**.** *The formulas $\neg(P \wedge Q)$ and $\neg P \vee \neg Q$ are logically equivalent. The formulas $\neg(P \vee Q)$ and $\neg P \wedge \neg Q$ are logically equivalent.*

Proof. Build the truth tables and compare the last columns. □

2.6 Proofs by contraposition and contradiction

Definition 2.11 (Contrapositive)**.** The *contrapositive* of a formula $P \Rightarrow Q$ is $\neg Q \Rightarrow \neg P$.

contrapositive

[6] After Augustus De Morgan, a British logician from the 19th century.

Original statement	Equivalent statement	Explanation
$\neg(P \vee Q)$	$\neg P \wedge \neg Q$	If (P or Q) is false, then *both* P and Q must be false.
$\neg(P \wedge Q)$	$\neg P \vee \neg Q$	If it is not the case that both P and Q are true, then one of P or Q must be false.
$\neg(\neg P)$	P	If it is false that P is false, then P must be true.
$\neg(\forall x \in A\ P(x))$	$\exists x \in A\ \neg P(x)$	If a property P is not true for *all* $x \in A$, then *there exists some* $x \in A$ for which it is false
$\neg(\exists x \in A\ P(x))$	$\forall x \in A\ \neg P(x)$	If a property P is not true for *any* $x \in A$, then the property must be false *for all* $x \in A$

Figure 13: Simplification of negations in mathematical statements

Theorem 2.12 (Correctness of the proof by contraposition)**.** *$A \Rightarrow B$ and $\neg B \Rightarrow \neg A$ are logically equivalent. In other words, whatever A and B stand for, proving that A implies B is equivalent to proving that $\neg B$ implies $\neg A$.*

Proof. We could simply build the truth table for $\neg B \Rightarrow \neg A$ and see that it is the same as $A \Rightarrow B$.

Another way to prove the result is to use Proposition 2.3. We know that $\neg B \Rightarrow \neg A$ is logically equivalent to $\neg(\neg B) \vee \neg A$, by Proposition 2.3. One sees easily that $\neg(\neg B)$ and B are logically equivalent, so $\neg(\neg B) \vee \neg A$ and $B \vee \neg A$ are equivalent. Now, $B \vee \neg A$ and $\neg A \vee B$ are equivalent. By Proposition 2.3, this is equivalent to $A \Rightarrow B$. □

What this means is that when one is asked to prove $A \Rightarrow B$, it is logically correct to prove $\neg B \Rightarrow \neg A$ instead. This is what we did in **??**. Of course, a direct proof would always be possible, but sometimes it is easier in one of the two directions.

We can similarly show that a proof by contradiction is also "mathematically legal."

Theorem 2.13 (Correctness of the proof by contradiction)**.** *For every statement A, A and $\neg A \Rightarrow$ False are logically equivalent. In other words, whatever A stands for, proving A is equivalent to proving that $\neg A$ leads to a contradiction.*

Proof. We build the truth table for the second expression.

A	$\neg A$	$\neg A \Rightarrow$ False
False	True	False
True	False	True

□

The difference between a proof by contraposition and a proof by contradiction is very subtle and in fact there is essentially no difference between them. Whenever something can be proved using a proof by contraposition, then a proof by contradiction is also possible, and conversely. In some sense, it always best to use a proof by contradiction, as this is the proof where we have the most assumptions available: we can assume both the hypotheses of our statement *and* we can also assume the negation of the conclusion.

Let us see another example of an indirect proof right away:

Theorem 2.14. *For all $p, q \in \mathbb{N}$, we have $\sqrt{2} \neq \frac{p}{q}$.*

Proof. Suppose that the theorem is *not* true, and derive a contradiction. First, we write the statement in mathematical language:

$$\forall p \in \mathbb{N}\, \forall q \in \mathbb{N}\, \sqrt{2} \neq \frac{p}{q}$$

The negation of this statement is

$$\exists p \in \mathbb{N}\, \exists q \in \mathbb{N}\, \sqrt{2} = \frac{p}{q},$$

which means that there exist natural numbers $p, q \in \mathbb{N}$ such that $\sqrt{2} = \frac{p}{q}$.

We can assume that p and q do not share any factor (otherwise we could simplify the fraction $\frac{p}{q}$. Then $2 = \frac{p^2}{q^2}$, or by multiplying both sides by q^2 we get $2q^2 = p^2$, and in particular p^2 must be an even number. Then p itself must be an even number, so it can be written as $2p'$. So $q^2 = \frac{p^2}{2} = \frac{(2p')^2}{2} = 2(p')^2$, from which we get that q^2 is even, so q is even. We obtain a contradiction to our assumption that p and q do not share any factor, since both are divisible by 2. □

It *is* possible to give a direct proof that $\sqrt{2}$ is irrational (i.e., that it is not equal to a fraction), but it definitely requires more effort.

2.7 Induction

Induction is a proof technique that is used to prove statements of the form

$$\forall n \in \mathbb{N}_0\, P(n)$$

where $P(n)$ is an arbitrary predicate. For example, we could take the predicate $\binom{n}{3} = \binom{n}{n-3}$, or $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.

To prove such a statement, it is not enough to simply check a few cases with $n = 0, 1, 2, 3$ since this can lead to some mistakes. For example, if $P(n)$ is the statement "$n^2 + n + 41$ is a prime number", then the statement is true for $n = 0$ up to $n = 40$, but fails for $n = 41$.

A reasoning by induction has the following structure:

- A *base case*: we show that $P(n)$ is true for small values $n \leq n_0$. Often, we only do it for $n = 1$, but this depends on P itself (we will see some examples). — induction base case
- An *induction step*: we assume that for an arbitrary $n \geq n_0$ the statement $P(n)$ is true, and based on this assumption we prove that $P(n+1)$ is true. In other words, we prove that the statement $\forall n \in \mathbb{N}_0 \, ((n \geq n_0 \wedge P(n)) \Rightarrow P(n+1))$. We say that $P(n)$ is the *induction hypothesis*. — induction step; induction hypothesis

If we can accomplish these two steps, it is intuitively clear that $P(n)$ must be true for all $n \in \mathbb{N}_0$, but we give a formal proof.

Proposition 2.15. *Let $P(n)$ be a predicate. Suppose that $P(0), P(1), \ldots, P(n_0)$ are true for some $n_0 \in \mathbb{N}_0$, and that for every $n \geq n_0$, we have that $P(n)$ implies $P(n+1)$. Then $P(n)$ is true for all $n \in \mathbb{N}_0$.*

Proof. The proof goes by contradiction: we take for granted that the assumptions are true, and we assume that the conclusion is false. The conclusion being false means that there exists $n \in \mathbb{N}_0$ such that $P(n)$ is false. We can choose such an n that is the smallest possible, that is, n such that for every number $m < n$, we have that $P(m)$ is true.[7] It cannot be the case that $n \leq n_0$, because our assumptions tell us that $P(0), \ldots, P(n_0)$ are all true. So it must be the case that $n > n_0$. Let m be $n-1$. Then we have $m \geq n_0$, and since $m < n$ we must have that $P(m)$ is true. By the assumption that $P(m)$ implies $P(m+1)$ and since $m+1 = n$, we obtain that $P(n)$ is true. This is a contradiction. □

We give an example of a standard proof by induction.

Lemma 2.16. *For all $n \in \mathbb{N}$, $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.*

Proof. Let $P(n)$ be the predicate $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$. We prove the base case for $n = 1$. In that case, the left-hand side of the equation is just 1, and the right-hand side is $\frac{1\times 2}{2} = 1$, so the result is true.

We now do the induction step, for which one needs to show that $P(n)$ implies $P(n+1)$ for all $n \geq 1$. So let us assume that $P(n)$ is true. Then we have $1+2+\cdots+n+(n+1) = (1+2+\cdots+n)+(n+1)$, and the left bracket is equal to $\frac{n(n+1)}{2}$ by our induction hypothesis $P(n)$. Then $\frac{n(n+1)}{2}+(n+1) = \frac{n(n+1)}{2}+\frac{2(n+1)}{2} = \frac{n(n+1)+2(n+1)}{2} = \frac{(n+2)(n+1)}{2} = \frac{(n+1)(n+2)}{2}$. We notice that this is equal to the right-hand side of the equation in Lemma 2.16, where we replaced n by $n+1$. So $P(n+1)$ is true and we are finished with the proof. □

We now see an example where more than one base case needs to be checked. In this example, we use the so-called *strong induction principle*: during the induction step, instead of only assuming that $P(n)$ is true, we assume that $P(m)$ is true for all $m \leq n$. — strong induction principle

[7] Phrased differently, this means that all $P(0), \ldots, P(n-1)$ are true. See the comment about the "strong induction principle" a few paragraphs below.

The proof of Proposition 2.15 shows that the strong induction principle is indeed a valid way to prove things (see the footnote).

Lemma 2.17. *Let $(a_n)_{n\in\mathbb{N}_0}$ be the sequence defined by $a_0 = 1, a_1 = 3$ and $a_n = 4a_{n-1} - 3a_{n-2}$ for $n \geq 2$. Then for all $n \in \mathbb{N}_0$, we have $a_n = 3^n$.*

Proof. Let $P(n)$ be the predicate "$a_n = 3^n$". We prove the result by induction, with base cases $n = 0$ and $n = 1$. Since $3^0 = 1 = a_0$ and $3^1 = 3 = a_1$, the base cases $P(0)$ and $P(1)$ are true.

Let us prove the induction step using the strong induction principle. We let $n \geq 1$, we assume that $P(n-1)$ and $P(n)$ are true, and we prove $P(n+1)$. We have $a_{n+1} = 4a_n - 3a_{n-1}$ and by using the induction hypotheses we can rewrite this as $a_{n+1} = 4 \cdot 3^n - 3 \cdot 3^{n-1} = 4 \cdot 3^n - 3^n = (4-1) \cdot 3^n = 3 \cdot 3^n = 3^{n+1}$, so $P(n+1)$ is true. □

Being able to prove a statement by induction is a crucial skill that you need to master this semester, practice time and again until you can write a correct induction proof in a structured way:

1. define what is the predicate $P(n)$ that you are proving,
2. describe what are the base cases and prove that the predicate holds for these base cases,
3. describe the induction step and prove it.

2.8 Exercises

1. Prove that P and $\neg\neg P$ are logically equivalent.

2. Prove that $P \vee \neg P$ is a tautology.

3. Prove that $((P \Rightarrow Q) \Rightarrow P) \Rightarrow P$ is a tautology.

4. Prove or disprove that $A \Rightarrow (B \Rightarrow C)$ and $(A \wedge B) \Rightarrow C$ are logically equivalent.

5. Prove that $((P \vee Q) \Rightarrow R) \Rightarrow (P \Rightarrow R)$ is a tautology.

6. Prove that $(P \Rightarrow R) \Rightarrow ((P \vee Q) \Rightarrow R)$ is *not* a tautology.

7. Prove that $\neg P$ and $P \Rightarrow$ False are logically equivalent.

8. Prove that $P \wedge (Q \vee R)$ and $(P \wedge Q) \vee (P \wedge R)$ are logically equivalent. Does it remind you of something?

9. Argue that $\forall x \in \mathbb{N}_0\ (P(x) \wedge Q(x))$ is the same as $(\forall x \in \mathbb{N}_0\ P(x)) \wedge (\forall x \in \mathbb{N}_0\ Q(x))$.

10. Give examples of predicates $P(x)$ and $Q(x)$ such that $\forall x \in \mathbb{N}_0 (P(x) \vee Q(x))$ is not equivalent to $(\forall x \in \mathbb{N}_0\ P(x)) \vee (\forall x \in \mathbb{N}_0\ Q(x))$.

11. Prove by induction that for all $n \in \mathbb{N}$, the equality $1+3+5+7+\cdots+(2n-1) = n^2$ holds. (For $n = 1$, the left side is 1, for $n = 2$ it is $1 + 3$, for $n = 3$ it is $1 + 3 + 5$, and so forth.)

12. Let $(a_n)_{n\in\mathbb{N}_0}$ be the sequence defined by $a_0 = 1, a_1 = 1$, and $a_n = 4a_{n-2} - 3a_{n-1}$ for every $n \geq 2$. Show that for all $n \in \mathbb{N}_0$, we have $a_n = 1$.

3 Relations

Things to remember / to know

- Know the notion of a relation.
- Know the definitions of reflexivity, symmetry, transitivity, antireflexivity, antisymmetry, and how to prove that those properties hold for a given relation.
- Know how to compose binary relations, invert binary relations, compute the various closures (reflexive, symmetric, transitive, reflexive transitive) of a relation.
- Know the definition of an equivalence relation, be able to compute equivalence classes for an equivalence relation, be able to compute the factor of a set by an equivalence relation.
- Know the definition of the various forms of orders (quasi, partial, linear, strict) and how to prove that a given relation is an order.
- Know how to draw the Hasse diagram of an order.
- Know how to determine upper bounds, lower bounds, maximal elements, minimal elements in an order.

3.1 General Definitions

Definition 3.1. A *relation* is a subset R of $A_1 \times \cdots \times A_n$, for some $n \geq 1$ and sets $A_1, \ldots, A_n$. We say that n is the *arity* of the relation.

relation

arity

Unfolding the definition, this means that a relation R is a set of tuples $(a_1, \ldots, a_n)$, where $a_1 \in A_1, \ldots, a_n \in A_n$. Most of the time, we are interested in *binary* relations, i.e., relations $R \subseteq A \times B$ that have arity 2. In that case, R is a set of pairs (a, b) with $a \in A$ and $b \in B$. These relations can be depicted in two different ways:

binary

- we can represent A and B as bubbles/potatoes and put a line between elements of A and B whenever the pair is in the relation, see Figure 14.
- If $A = B$, we can represent R as a *directed network*, also called *directed graph*: dots represent the elements of A, and we put an arrow from a to b whenever $(a, b) \in R$, see Figure 15. Those directed networks are one of the most important concepts in computer science, one can use them to give models to a number of real-life situations[8] and then study their properties mathematically.

directed graph

[8] For example social networks, road infrastructure, telecommunication infrastructure, and many oth-

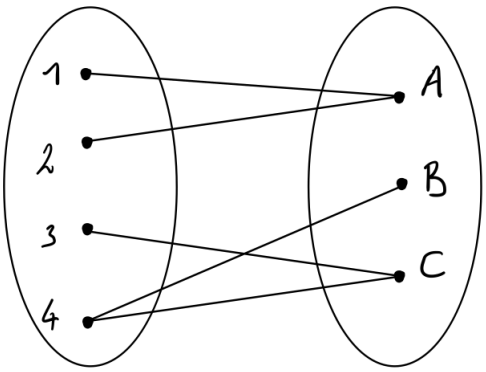


Figure 14: Depiction of the binary relation $\{(1, A), (2, A), (3, C), (4, B), (4, C)\}$

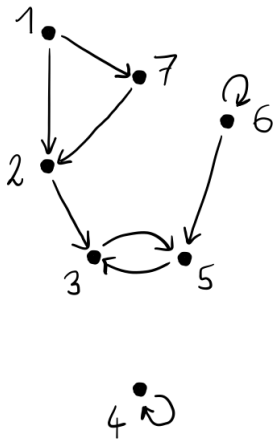


Figure 15: The binary relation $\{(1, 2), (1, 7), (2, 3), (3, 5), (4, 4), (5, 3), (6, 5), (6, 6), (7, 2)\}$ as a directed graph.

Note that a function is a particular example of a relation! Recall Definition 1.11, where we defined a function $f\colon A \to B$ to be a subset $f \subseteq A \times B$ with two essential properties: for every $a \in A$, there exists some $b \in B$ such that $(a, b) \in f$, and moreover b is uniquely determined. So relations generalize functions in two ways: there can exist a for which no $b \in B$ exists such that $(a, b) \in R$, or there could be multiple such bs.

Another important example is the *equality relation* on a set A. It is the relation $\Delta_A = \{(a, b) \in A^2 \mid a = b\}$,[9] (where Δ stands for *D*iagonal). For example, $\Delta_{\{0,1\}} = \{(0, 0), (1, 1)\}$. equality relation

The following are common properties that some relations can have.

Definition 3.2. Let $R \subseteq A \times A$ be a binary relation. We say that R is:

- *reflexive* if for every $a \in A$, we have that (a, a) is in R. reflexivity
- *transitive* if for every $a, b, c \in A$ such that $(a, b), (b, c) \in R$, then we also have $(a, c) \in R$. transitivity
- *symmetric* if for every $a, b \in A$ such that $(a, b) \in R$, then also (b, a) in R. symmetry
- *antireflexive* if for every $a \in A$, one has that (a, a) is *not* in R. antireflexivity
- *antisymmetric* if for every $a, b \in A$ such that both $(a, b), (b, a)$ are in R, then $a = b$. antisymmetry

ers.

[9] The notation $\{(a, a) \in A^2 \mid a \in A\}$ could also be used instead of $\{(a, b) \in A \times A \mid a = b\}$.

3.2 Algebraic operations on relations

Just like functions, relations can be composed.

composition of relations

Definition 3.3. Let $R \subseteq A \times B$ and $S \subseteq B \times C$. Then $R + S$ is the relation defined by

$$\{(a, c) \in A \times C \mid \exists b \in B : (a, b) \in R \text{ and } (b, c) \in S\}.$$

Remark 3.4. Some authors use the symbol $R \circ S$ for the composition of relations, but this can lead to confusion since the composition of relations is defined in the opposite way as function composition (think about why this is the case).

Note that $R + S$ is not always defined, and in general it is not the same thing as $S + R$.

Lemma 3.5. *Let $R \subseteq A \times B, S \subseteq B \times C, T \subseteq C \times D$. Then $(R + S) + T$ and $R + (S + T)$ are the same relation on $A \times D$.*

Proof. We show the inclusion $(R + S) + T \subseteq R + (S + T)$. Let $(a, d) \in (R + S) + T$. By definition, this means there exists $c \in C$ such that $(a, c) \in R + S$ and $(c, d) \in T$. Again by definition, there exists $b \in B$ such that $(a, b) \in R$ and $(b, c) \in S$. Since $(b, c) \in S$ and $(c, d) \in T$, we have $(b, d) \in S + T$. Finally, $(a, b) \in R$ and $(b, d) \in S + T$ means that $(a, d) \in R + (S + T)$.

The other direction is similar. □

If $R \subseteq A \times A$ and $n \geq 1$, we write $n \cdot R$ for the composition $R + R + \cdots + R$, where R appears n times. Note that by Lemma 3.5 (and by induction), the notation $R + R + \cdots + R$ is unambiguous and no parentheses are needed.

Moreover, while only some functions can be inverted (i.e., the bijective functions), every binary relation has an inverse.

inverse of a relation

Definition 3.6. Let $R \subseteq A \times B$ be a relation. The *inverse* of R is the relation

$$\{(b, a) \in B \times A \mid (a, b) \in R\}$$

and is denoted by $-R$.

Note that while the "inverse" of a bijective function $f \colon A \to A$ satisfies that for every $g \colon A \to A$, we have $(g \circ f) \circ f^{-1} = g$, this notion of inverse for relations does not obey the same cancellation property. Indeed, it can be the case that $(S + R) - R \neq S$ for some relations $R, S \subseteq A \times A$.

3.3 Closure operations

Most relations are not transitive, reflexive, or symmetric. Given a relation R that fails to have one of these properties, we consider the question of what is the *smallest* relation T such that $R \subseteq T$ and such that T has the property that we want. We call this a *closure operation*.

Definition 3.7. Let $R \subseteq A \times A$. Then we define:

- The *transitive closure* of R is the relation $R^+ \subseteq A \times A$ such that $(a, b) \in R^+$ if, and only if, there exist $n \in \mathbb{N}$ and elements $c_0, \ldots, c_n$ such that $c_0 = a, c_n = b$, and for every $i \in \{0, \ldots, n-1\}$, we have that $(c_i, c_{i+1}) \in R$. — transitive closure
- The *reflexive closure* of R is the relation $R \cup \Delta_A$. — reflexive closure
- The *symmetric closure* of R is the relation $R \cup (-R)$. — symmetric closure

In other words, if we view R as a graph:

- the transitive closure of R contains all the pairs (a, b) that are connected by a path of length 1 or more.
- the reflexive closure of R is like R, but we added a loop (a, a) at every vertex.
- the symmetric closure of R is obtained by adding to R an arrow from b to a whenever there exists an arrow from a to b.

Note that all these types of closure are relations that contain R.

Proposition 3.8. *Let $R, T \subseteq A \times A$ be relations such that $R \subseteq T$ and T is transitive. Then $R^+ \subseteq T$. In other words, R^+ is the smallest transitive relation containing R.*

Proof. To prove $R^+ \subseteq T$, we take an arbitrary element $(a, b) \in R^+$ and show that it is in T. Every such $(a, b) \in R^+$ comes by definition from a sequence $c_0, \ldots, c_n$ for some $n \in \mathbb{N}$. We do a proof by induction, showing that for any length n of the sequence, we have $(a, b) \in T$. Formally, we define the following predicate $P(n)$:

$$\forall a, b, c_0, \ldots, c_n \in A \, [(a = c_0 \wedge b = c_n \wedge \forall i \in \{0, \ldots, n-1\}(c_i, c_{i+1}) \in R) \Rightarrow (a, b) \in T]$$ [10]

and we show by induction $\forall n \in \mathbb{N} \; P(n)$, with base case $n = 1$.

If $n = 1$, let $a, b, c_0, c_1 \in A$ be such that $c_0 = a, c_1 = b$, and $(c_0, c_1) \in R$. So $(a, b) \in R$, and by assumption on T we have $(a, b) \in T$.

[10] We commit a slight abuse of notation and omit some quantifiers at the beginning. A proper way to write the beginning of the formula would be $\forall a \in A \forall b \in A \forall c_0 \in A \; \ldots \; \forall c_n \in A$.

Let us now prove the induction step: we assume that $P(n)$ is true, and we show that $P(n+1)$ is true. Let $a, b, c_0, \dots, c_{n+1} \in A$ be such that $c_0 = a, c_{n+1} = b$, and for every $i \in \{0, \dots, n\}$, we have $(c_i, c_{i+1}) \in R$. Let b' be c_n. Then since we have a "path" $c_0, \dots, c_n$ from a to b', we can apply the induction hypothesis $P(n)$, which tells us that $(a, b') \in T$. And we have $(b', b) \in R$, so that $(b', b) \in T$ as T contains R. We finally use the assumption that T is transitive: since $(a, b'), (b', b)$ are in T, we must have $(a, b) \in T$. This concludes the proof of the induction step.

So we have that $\forall n \in \mathbb{N}\ P(n)$ is true, and it follows that $R^+ \subseteq T$. □

Similarly, one could show that the symmetric closure of R is the smallest symmetric relation containing R, and that the reflexive closure of R is the smallest reflexive relation containing R (see Exercises 5 and 6).

3.4 Equivalence relations

Definition 3.9. A relation $R \subseteq A \times A$ is an *equivalence relation* if it is reflexive, transitive, and symmetric.

equivalence relation

There is an equivalence relation that you know well: the equality relation Δ_A. We always have $a = a$ (so equality is reflexive), it doesn't matter whether we write $a = b$ or $b = a$ (equality is symmetric), and if $a = b$ and $b = c$, then certainly $a = c$ (equality is transitive). Equivalence relations are relations that "approximate" equality.

This is the *smallest* equivalence relation on A: given any equivalence relation $E \subseteq A^2$, one has $\Delta_A \subseteq E$. There is also a *largest* equivalence relation on A, the relation $\nabla_A = A^2$ (pronounced *nabla*): by definition, every $E \subseteq A^2$ is such that $E \subseteq \nabla_A$. The relations Δ_A and ∇_A are called *trivial*, and the others *non-trivial*.

trivial equivalence relation

Example 3.10. Let $E \subseteq \mathbb{N} \times \mathbb{N}$ be the relation $\{(a, b) \in \mathbb{N}^2 \mid b - a \text{ is even}\}$. This relation contains for example $(0, 2), (1, 3), (7, 17)$ but does not contain $(2, 1)$. This is an equivalence relation: for $a \in \mathbb{N}$, we have $a - a = 0$, which is even, so E contains every pair (a, a) and is therefore reflexive. If $b - a$ and $c - b$ are even, then $c - a = (c - b) + (b - a)$ is the sum of two even numbers, so it is even. Thus, E is transitive. Moreover, if $b - a$ is even, then $a - b$ is also even.

Notation 3.11. Equivalence relations are often written as $\sim, \simeq$, or $\equiv$ instead of R (as a way to easily remember that they are equivalence relations). And instead of writing $(a, b) \in R$, one would usually write $a \sim b, a \simeq b$, or $a \equiv b$.

3.4.1 Equivalence classes and partitions

Definition 3.12. Let $\sim$ be an equivalence relation on A, and let $a \in A$. The *equivalence class* of a is the set of all $b \in A$ such that $a \sim b$.

equivalence class

Note that for any two b, b' in the equivalence class of some a, we have $b \sim b'$: indeed, $a \sim b$ and $a \sim b'$, so by symmetry and transitivity of equivalence relations we must have $b \sim b'$.

Notation 3.13. If $\sim$ is an equivalence relation on A and $a \in A$, the equivalence class of a is often denoted by $[a]_\sim$.

Example 3.14. Returning to Example 3.10, the equivalence class of 0 is the set of all numbers b such that $b-0$ is even, i.e., it is exactly the set of all even numbers. In other words, $[0]_E = \{n \in \mathbb{N} \mid n \text{ is even}\}$. The equivalence class of 1 is the set of all numbers b such that $b-1$ is even, so that b is an even number plus 1. This equivalence class is exactly the set of all odd numbers, $[1]_E = \{n \in \mathbb{N} \mid n \text{ is odd}\}$.

The number of different equivalence classes of an equivalence relation is called its *index*.

index of an equivalence relation

Lemma 3.15. *Let C_1, C_2 be two equivalence classes of some equivalence relation $\sim$. Then either $C_1 \cap C_2 = \emptyset$, or $C_1 = C_2$.*

Comment about the proof

The statement is of the form $P \Rightarrow (Q \vee R)$: P is the proposition saying that $\sim$ is an equivalence relation and that C_1 and C_2 are two equivalence classes, Q is the proposition saying that $C_1 \cap C_2 = \emptyset$, and R is the proposition saying that $C_1 = C_2$. Convince yourself by a proof table that $P \Rightarrow (Q \vee R)$ is logically equivalent to $(P \wedge \neg Q) \Rightarrow R$. Thus, it is equivalent to prove this second statement instead.

Proof. Let a_1 be such that $[a_1]_\sim = C_1$, and a_2 be such that $[a_2]_\sim = C_2$. Suppose that $C_1 \cap C_2 \neq \emptyset$, so there exists $b \in C_1 \cap C_2$. By definition of C_1 and C_2, we have $a_1 \sim b$ and $a_2 \sim b$.

We show that $C_1 \subseteq C_2$. Let $c \in C_1$. Since C_1 is the equivalence class of a_1, we must have $a_1 \sim c$. Similarly, we must have $a_1 \sim b$. By symmetry, this means that $b \sim a_1$, and by transitivity of $\sim$ we obtain $b \sim c$ (we also remarked this fact after Definition 3.12. Thus, again by transitivity, we have $a_2 \sim c$, which means that $c \in C_2$.

The other inclusion $C_2 \subseteq C_1$ is proved in exactly the same way, and therefore $C_1 = C_2$. □

We say that the equivalence classes of $\sim$ form a *partition* of the set A:

partition

Definition 3.16. Let A be a set. A partition of A is a set $\mathcal{P}$ of non-empty *subsets* of A such that:

- For every $a \in A$, there exists $C \in \mathcal{P}$ such that $a \in C$. We say that $\mathcal{P}$ *covers* A.
- For every $C_1, C_2 \in \mathcal{P}$, either $C_1 \cap C_2 = \emptyset$ or $C_1 = C_2$.

Note that a partition is a subset of $\mathcal{P}(A)$, and therefore it is an element $\mathcal{P}(\mathcal{P}(A))$.

Notation 3.17. Let A be a set. We write $\mathrm{Part}(A)$ for the set of all partitions of A, and $\mathrm{Eq}(A)$ for the set of equivalence relations on A.

Theorem 3.18. *For every set A, there is a bijection between* $\mathrm{Part}(A)$ *and* $\mathrm{Eq}(A)$.

Proof. We define a bijective function $\varphi\colon \mathrm{Eq}(A) \to \mathrm{Part}(A)$.

Let $E \subseteq A^2$ be an equivalence relation. We define $\varphi(E)$ to be $\{[a]_E \in \mathcal{P}(A) \mid a \in A\}$. This is a well-defined partition by Lemma 3.15: the classes are disjoint, and every element a is an element of its own equivalence class, so that $\varphi(E)$ covers A.

We show that φ is injective. Suppose that E and E' are different equivalence relations. Therefore, there must exist $(a, b) \in E \Delta E'$ (the symmetric difference of E and E'). Without loss of generality, we have $(a, b) \in E \setminus E'$. Then $[a]_E = [b]_E$, but $[a]_{E'} \neq [b]_{E'}$. Therefore $\varphi(E)$ contains $[a]_E$ and cannot contain $[a]_{E'}$: if $[a]_{E'} \in \varphi(E)$, then we have two sets $[a]_E, [a]_{E'} \in \varphi(E)$ that are different but have a non-empty intersection (as they both contain a), contradicting the fact that $\varphi(E)$ is a partition. So we have $\varphi(E) \neq \varphi(E')$.

Let us now show the surjectivity of φ. Let $\mathcal{P}$ be a partition of A. Define $E \subseteq A^2$ as $E = \{(a, b) \in A^2 \mid \exists C \in \mathcal{P} : a \in C \wedge b \in C\}$. We need to prove that E is an equivalence relation and that $\varphi(E)$ is equal to $\mathcal{P}$. We show that E is reflexive. Let $a \in A$. Since $\mathcal{P}$ is a A, there exists a $C \in \mathcal{P}$ such that $a \in C$. Then $(a, a) \in E$ as is witnessed by this C, so that E is reflexive. The symmetry of E is clear from the definition, which is symmetric in a and b. Finally, if $(a, b), (b, c) \in E$, there exist $C, D \in \mathcal{P}$ such that $a, b \in C$ and $b, c \in D$. Since $\mathcal{P}$ is a partition, and since $b \in C \cap D$, it must be that $C = D$. So $a \in C$ and $c \in C$, so that $(a, c) \in E$.

To prove that $\varphi(E)$ is equal to $\mathcal{P}$, one must prove two inclusions. We leave as an exercise to check that if $a \in C$ for some $C \in \mathcal{P}$, then $[a]_E = C$.[11] Let $[a]_E$ be an equivalence class of E. Since $\mathcal{P}$ is a partition, we have some $C \in \mathcal{P}$ such that $a \in C$. By the claim above this means that $[a]_E = C$, therefore $\varphi(E) \subseteq \mathcal{P}$.

For the other direction let $C \in \mathcal{P}$. Since $C \neq \emptyset$, there exists $a \in C$. We have $[a]_E = C$ by the claim above, and therefore $C \in \varphi(E)$, so that $\mathcal{P} \subseteq \varphi(E)$. □

[11] Proceed by double inclusion.

3.4.2 Factoring by an equivalence relation

Definition 3.19. Let $\sim$ be an equivalence relation on A. The set $A/\sim$ is the set whose elements are the equivalence classes of elements of a.

factoring by an equivalence relation

We say that $A/\sim$ is obtained by *factoring A by the equivalence relation $\sim$*. It is an extremely useful concept in mathematics/theoretical computer science: sometimes understanding a set A can be difficult. By finding an appropriate equivalence relation $\sim$, the set $A/\sim$ is smaller than A so can be understood (for example by induction), and the equivalence classes are also smaller and therefore can be understood (by induction). It remains to understand the structure of A from the structure of $A/\sim$ and of the equivalence classes of $\sim$. Some examples of this are Theorems 3.22 and 3.23, where we show a certain connection between functions $A \to B$ and functions $A/\sim \to B$. Another important example is Theorem 6.13.

Example 3.20. Take $\sim$ to be the relation from Example 3.10. Any two even numbers a, b are equivalent, and any two odd numbers are equivalent. But if a is even and b is odd, then $b - a$ is odd and therefore $a \not\sim b$. It follows that $\mathbb{N}/\sim$ contains exactly two elements: the equivalence class containing all the even numbers, and the equivalence class containing all the odd numbers.

Given an equivalence relation $\sim$ on A, there exists a function $f\colon A \to A/\sim$ such that $f(a)$ is defined to be the equivalence class of a. This map is called the *canonical factor map*. This function is always surjective, and sometimes denoted by π.

canonical factor map

Proposition 3.21. *Let $f\colon A \to B$ be a function. The relation $R = \{(a, b) \in A^2 \mid f(a) = f(b)\}$ is an equivalence relation.*

Proof. We have to show that R is reflexive, symmetric, and transitive. For the first, let $a \in A$. Then $f(a) = f(a)$, so $(a, a) \in R$ by definition. For the second, if $(a, b) \in R$ then we have $f(a) = f(b)$ by definition, so that $f(b) = f(a)$, and therefore $(b, a) \in R$. Finally, suppose that $(a, b) \in R$ and $(b, c) \in R$. Then $f(a) = f(b)$ and $f(b) = f(c)$, so that $f(a) = f(c)$ and $(a, c) \in R$. □

The relation in Proposition 3.21 is called the *kernel* of f and denoted by $\ker(f)$.

kernel

Theorem 3.22. *Let $f\colon A \to B$, and let $\pi\colon A \to A/\ker(f)$ be the canonical factor map. Then there exists a unique injective function $g\colon A/\ker(f) \to B$ such that $g \circ \pi = f$. If f is surjective, then g is a bijective function.*

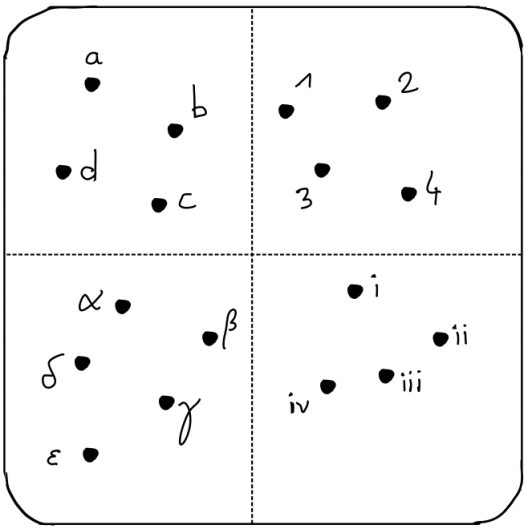


Figure 16: A partition on the set $A = \{a, b, c, d, \alpha, \beta, \gamma, \delta, \epsilon, 1, 2, 3, 4, i, ii, iii, iv\}$, decomposing the set into 4 parts (depicted by a dashed line). One can see this partition as an equivalence relation E on that same set, where $(x, y) \in E$ if x, y are in the same quadrant. The set A/E is $\{\{a, b, c, d\}, \{\alpha, \beta, \gamma, \delta, \epsilon\}, \{1, 2, 3, 4\}, \{i, ii, iii, iv\}\}$.

Comment about the proof

In the following proof, we define the function $g\colon A/\ker(f) \to B$ explicitly. To do this, we define g on an equivalence class C by picking an arbitrary $a \in C$ (and then defining $g(C)$ in terms of a, namely by $g(C) = f(a)$). There is one very important thing to consider when doing this: when defining $g(C)$, one must be careful that the result does not depend on the choice of the element a! Indeed, what if there were two $a, b \in C$ such that $f(a)$ and $f(b)$ are different? So when defining a function whose inputs are equivalence classes, the first thing we always have to check is whether the definition makes sense.

Proof. We define a map $g\colon A/\ker(f) \to B$. This map takes as an argument an arbitrary equivalence class C, and must output an element in B. For this, take an *arbitrary* $a \in C$, and define $g(C)$ to be $f(a)$.

We first prove that g is well-defined. Suppose that a, b are elements of C. Then we have $(a, b) \in \ker(f)$, since any two elements in the same equivalence class must be equivalent under the equivalence relation. By definition of $\ker(f)$, this means that $f(a) = f(b)$, and therefore g is well-defined.

Let us now prove that g is injective. Let C, D be two distinct equivalence classes. By Lemma 3.15, we have $C \cap D = \emptyset$. Pick an arbitrary $a \in C$ and $b \in D$. Since $C \cap D = \emptyset$,

we must have that $(a,b) \notin \ker(f)$, so $f(a) \neq f(b)$. It follows that $g(C) \neq g(D)$, and therefore g is injective.

Moreover, by definition we have $(g \circ \pi)(a) = g([a]) = f(a)$, so $g \circ \pi = f$.

We now prove that g is unique. Suppose that $h\colon A/\ker(f) \to B$ is another function such that $h \circ \pi = f$. Let C be an equivalence class of $\ker(f)$ and let $a \in C$. We must have $(h \circ \pi)(a) = f(a)$, and therefore $h(C) = h(\pi(a)) = f(a)$. Thus, $h = g$.

Finally, suppose that f is surjective. By Exercise 9 in Chapter 1, and since $g \circ \pi = f$, we have that g is surjective. We already proved that it is injective as well, so that g is a bijective function. □

Theorem 3.23. *Let A, B be sets, and let $E \subseteq A^2$ be an equivalence relation. There is a bijection between the set of functions $A/E \to B$ and the set of all functions $g\colon A \to B$ such that $E \subseteq \ker(g)$.*

Comment about the proof

In this proof, we define a function φ that takes as input *another function*, and then we prove that this function is a bijection. Be careful: this does not mean that the arguments of φ are bijections. It means that if $f \neq g$, then $\varphi(f) \neq \varphi(g)$ (φ is injective), and that every function has a preimage under φ.

Proof. Let F be the set of functions $A/E \to B$, and let G be the set of functions $g\colon A \to B$ such that $E \subseteq \ker(g)$. We define a bijective function $\varphi\colon G \to F$. Note that the arguments of φ are themselves functions, and so are the values of φ! So in the following it makes sense to write $\varphi(g)([a]_E)$, for $g \in G$ and $a \in A$.

Let $g\colon A \to B$ be a function such that $E \subseteq \ker(g)$. We define $\varphi(g)$ as the following function $A/E \to B$: given as input $[a]_E$ (the equivalence class of an element a), let $\varphi(g)([a]_E)$ be $g(a)$.

Again, we must prove that $\varphi(g)$ is well-defined. Suppose that $[a]_E = [b]_E$, i.e., that $(a,b) \in E$. Since $E \subseteq \ker(g)$, we have $(a,b) \in \ker(g)$, or in other words $g(a) = g(b)$. Therefore, $\varphi(g)$ is well-defined.

We now prove that φ is injective. Suppose that $f, g \in G$ are such that $f \neq g$. Then there exists $a \in A$ such that $f(a) \neq g(a)$. Then $\varphi(f)([a]_E) = f(a) \neq g(a) = \varphi(g)([a]_E)$, so $\varphi(f) \neq \varphi(g)$, i.e., φ is injective.

We now prove that φ is surjective. Let $f \in F$, that is, $f\colon A/E \to B$. We must find $g \in G$ such that $\varphi(g) = f$. Define $g\colon A \to B$ by $g(a) = f([a]_E)$. We have to check two things: $g \in G$, and $\varphi(g) = f$. For the first point, we must check that $E \subseteq \ker(g)$. Let $(a,b) \in E$. Then $[a]_E = [b]_E$, so $g(a) = f([a]_E) = f([b]_E) = g(b)$, i.e., $(a,b) \in \ker(g)$. For the second point, note that for every $a \in A$, we have

$$\begin{aligned} \varphi(g)([a]_E) &= g(a) && \text{by definition of } \varphi \\ &= f([a]_E) && \text{by definition of } g. \end{aligned}$$

So we have $\varphi(g) = f$. □

3.5 Orders

We have seen in Chapter 3.4 how to generalize the concept of equality. We see here how the concept of orders (such as the order we know over numbers) can be generalized.

Definition 3.24. A relation $R \subseteq A \times A$ is a *quasiorder* if it is reflexive and transitive. A relation $R \subseteq A \times A$ is an *order* if it is reflexive, transitive, and antisymmetric. We say that R is a *linear order* if it is an order such that for every $a, b \in A$, we have $a = b$ or $(a, b) \in R$ or $(b, a) \in R$. A *strict order* is an antireflexive, transitive, and antisymmetric relation.

quasiorder
order
linear order
strict order

The relation $R = \{(a, b) \in \mathbb{N}^2 \mid a \leq b\}$ is an order, even a linear order. When an order is not linear, one may sometimes emphasize this by saying that it is a *partial* order. Here is an example of an order that is not linear:

partial order

Example 3.25. Let R be the relation $\{(a, b) \in \mathbb{N}^2 \mid a \text{ divides } b\}$. So R contains for example $(2, 4), (2, 6), (1, 17), (3, 27), \ldots$ but it does not contain $(4, 6)$. We prove that it is an order. For any $a \in \mathbb{N}$, we have that $a = a \cdot 1$ so a divides a, and $(a, a) \in R$. To check transitivity, let $a, b, c \in \mathbb{N}$ be such that a divides b, and b divides c. Therefore, by definition one can write $b = a \cdot k$ and $c = b \cdot \ell$ for some $k, \ell \in \mathbb{N}$. Then $c = (a \cdot k) \cdot \ell = a \cdot (k \cdot \ell)$, so that a divides c and $(a, c) \in R$. Finally, assume that a divides b and b divides a. Then $a = b \cdot k$ and $b = a \cdot \ell$ for some $k, \ell \in \mathbb{N}$. It follows that $a = a \cdot (k \cdot \ell)$ and $b = b \cdot (k \cdot \ell)$. If $a \neq 0$, then we can divide the two sides of the first equation by a and we get $k \cdot \ell = 1$, and similarly if $b \neq 0$ we can divide the two sides of the second equation by b and get $k \cdot \ell = 1$. In either case, we obtain that $k = \ell = 1$ and therefore $a = b$. Now, it remains to deal with the case where both $a = 0$ and $b = 0$. But then $a = b$, so we are done.

Two elements a, b such that $(a, b) \in R$ or $(b, a) \in R$ or $a = b$ are called *comparable*, and *incomparable* otherwise. In the example above, 2 and 4 are comparable, but 2 and 3 are not.

comparable
incomparable

We also say that (A, R) is a *partially-ordered set* if $R \subseteq A \times A$ is an order. The abbreviation *poset* is also common. Usually, we use the symbol $\leq$ to denote an order on a set and instead of writing $(a, b) \in R$, we write $a \leq b$.

partially-ordered set
poset

Example 3.26. The sets $(\mathcal{P}(A), \subseteq)$ and $(\mathcal{P}_{\text{fin}}(A), \subseteq)$ are posets for every set A, where $\mathcal{P}_{\text{fin}}(A)$ is the set of all subsets of A that are finite.

$\mathcal{P}_{\text{fin}}(A)$

3.5.1 Hasse diagrams

Like every binary relation $R \subseteq A \times A$, one can represent an order as a directed graph. However, because of the fact that an order is transitive and reflexive, there are many arrows that are redundant: if we know the relation that we represent is an order, then we can simply not draw the arrows that are implied by the existence of other arrows. This leads to the notion of Hasse diagram.[12]

Definition 3.27. Let $(A, \leq)$ be a poset and let $a \neq b$. We say that b *covers* a, or that b is a cover of a, if we have $a \leq b$ and for every $c \in A$ such that $a \neq c$ and $a \leq c \leq b$, we have $b = c$.

cover of an element

For example, 2 covers 1 in the linear order $(\mathbb{N}, \leq)$, but not in $(\mathbb{Q}, \leq)$.
The *reflexive transitive closure* of a relation $R \subseteq A^2$ is $R^+ \cup \Delta_A$.

reflexive transitive closure

Lemma 3.28. *Let A be a finite set, and let O be an order on A. There $R = \{(a, b) \in O \mid b \text{ covers } a\}$ is an antireflexive relation whose reflexive transitive closure is O, and such that for every $T \subseteq O$ whose reflexive transitive closure is equal to O, then $R \subseteq T$.*

Proof. Since A is finite, O must be finite: we have $O \subseteq A^2$ so $|O| \leq |A^2| = |A|^2$.

It is clear that R is antireflexive, since a does not cover a by definition. We show that the reflexive transitive closure of R is O. Let $(a, b) \in O$. If $a = b$, then (a, b) is in the reflexive transitive closure of R. Suppose that $a \neq b$. Let $n \in \mathbb{N}$ be maximal such that there exist $c_0, \ldots, c_n \in A$ with $c_0 = a, c_n = b$, and such that $c_i \neq c_{i+1}$ and $(c_i, c_{i+1}) \in O$ for all $i \in \{0, \ldots, n-1\}$. The fact that such a maximal n exists comes from the fact that A is finite! Indeed, we must have $|\{c_0, \ldots, c_n\}| \leq |A|$. We claim that each c_{i+1} covers c_i, for $i \in \{0, \ldots, n-1\}$. Indeed, suppose now, so that by definition there exist $i \in \{0, \ldots, n-1\}$ and some $d \in A \setminus \{c_i, c_{i+1}\}$ such that $(c_i, d), (d, c_{i+1}) \in O$. But then we could have taken a longer sequence $c_0, \ldots, c_i, d, c_{i+1}, \ldots, c_n$, a contradiction to our choice of n. Thus, we have $(c_i, c_{i+1}) \in R$ for all $i \in \{0, \ldots, n-1\}$, and R^+ contains (a, b).

We finally prove that if $T \subseteq O$ is such that the reflexive transitive closure of T is O, then $R \subseteq T$. Let $(a, b) \in R$. We have $(a, b) \in O$ so by assumption on T there exist $c_0, \ldots, c_n \in A$ such that $c_0 = a, \ldots, c_n = b$ and for every $i \in \{0, \ldots, n-1\}$, we have $(c_i, c_{i+1}) \in T \subseteq O$. Since b covers a, it must be that $n = 1$, so that $(a, b) \in T$. □

The relation in Lemma 3.28 is therefore the smallest relation $R \subseteq A^2$ such that the reflexive transitive closure of R is O. We call this relation the *transitive reduction* of O.

transitive reduction

[12] Named after Helmut Hasse, a German mathematician from the 20th century.

Definition 3.29. Let $O \subseteq A^2$ be an order on a finite set and let R be the transitive reduction of O. The *Hasse diagram* of O is obtained by drawing R as a directed graph in a way that if $(a, b) \in R$, then a is lower than b.

Hasse diagram

3.5.2 Maximal elements and upper bounds

Definition 3.30. Let $(A, \leq)$ be a partial order, and let $a \in A$. We say that a is *maximal* if for every $b \in A$ such that $a \leq b$, we have $b = a$. Similarly, a is *minimal* if for every $b \in A$ such that $b \leq a$, we have $a = b$.

maximal

minimal

Note that A might not have a single minimal element, or a single maximal element!

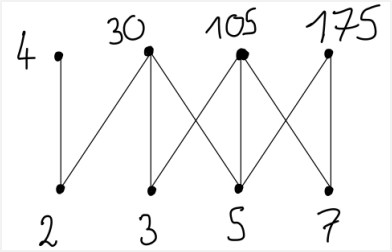
Definition 3.31. Let $(A, \leq)$ be a partially-ordered set, let $B \subseteq A$. We say that $u \in A$ is an *upper bound* of B if for every $b \in B$, we have $b \leq u$. We say that u is a *least upper bound* of B it is an upper bound, and for every other upper bound u' of B, we have $u \leq u'$.

upper bound

least upper bound

Note that a set B might not have an upper bound, and if it has one, it might not have a least upper bound.

Example 3.32. Consider the set $\{2, 3, 4, 5, 7, 30, 105, 175\}$ and order it by the divisibility relation as in Example 3.25. The Hasse diagram of this poset is the following:



Then $\{2, 7\}$ does not have an upper bound, $\{3, 5\}, \{5, 7\}$ have an upper bound but no least upper bound, and $\{2, 3\}, \{3, 7\}$ have a least upper bound. The elements $4, 30, 105, 175$ are all maximal in this poset.

If B has two elements a, b and has a least upper bound, then we write $a \vee b$ (pronounced "join") for the least upper bound. This is the same symbol as for "or" in Chapter 2 and it is not a coincidence! In the set $\{\text{True}, \text{False}\}$ together with the order stating $\text{False} \leq \text{True}$, the notion of least upper bound coincides exactly with the notion of "or" we have seen before.

The terms *lower bound* and *greatest lower bound* are defined similarly. We write $a \wedge b$ (pronounced "meet") for the greatest lower bound of a and b, if it exists.

greatest lower bound

3.5.3 Linear completions

As we have seen, a poset is in general not linear. It is always possible to complete a partial order into a linear order, however this completion is not unique.

Proposition 3.33. *Let $(A, \leq)$ be a poset. There exists a linear order $\preceq$ with the property that for all $a, b \in A$, if $a \leq b$ then $a \preceq b$.*

Proof. We do the proof in the case where A is finite; the case where A is infinite is beyond the scope of this course. Consider the set $\mathcal{R}$ of all relations R containing $\leq$ and that are partial orders. $\mathcal{R}$ is non-empty ($\leq$ is in $\mathcal{R}$) and for every $R \in \mathcal{R}$, we have $R \subseteq A^2$ and therefore $|R|$ is finite. Take an arbitrary $R \in \mathcal{R}$ having a maximal $|R|$.[13]

We claim that R is a linear order. Suppose that for some $a \neq b$, we have neither $(a, b) \in R$ nor $(b, a) \in R$. Let R' be the transitive closure of $R \cup \{(a, b)\}$. Then R' is reflexive (since R is), it is transitive (by definition of transitive closure), and it contains $\leq$ (since R does). We prove that R' is antisymmetric; this would imply that R' is in $\mathcal{R}$ and such that $|R'| > |R|$, which contradicts the assumption about $|R|$ being maximal.

Suppose then for contradiction that there exist $c \neq d$ such that $(c, d), (d, c) \in R'$. Then by the definition of transitive closure there exist paths $e_0, \ldots, e_n$ such that $e_0 = c, e_n = d$ and $(e_i, e_{i+1}) \in R \cup \{(a, b)\}$ for all $i \in \{0, \ldots, n-1\}$, and $e'_0, \ldots, e'_m$ such that $e'_0 = d, e'_m = c$ and $(e'_j, e'_{j+1}) \in R \cup \{(a, b)\}$ for all $j \in \{0, \ldots, m-1\}$.

Suppose that for two distinct $i, i' \in \{0, \ldots, n-1\}$, we have $(e_i, e_{i+1}) = (a, b)$ and $(e_{i'}, e_{i'+1}) = (a, b)$. Without loss of generality, $i < i'$. Then we have a path $b = e_{i+1}, \ldots, e_{i'} = a$ in R, so that $(b, a) \in R$ since R is transitive, a contradiction to the choice of a, b. So we can assume that for all but at most one $i \in \{0, \ldots, n-1\}$ we have $(e_i, e_{i+1}) \in R$. Similarly, for all but at most one $j \in \{0, \ldots, m-1\}$, we have $(e'_j, e'_{j+1}) \in R$.

Moreover, suppose that $(e_i, e_{i+1}) = (a, b)$ for some $i \in \{0, \ldots, n-1\}$, and that $(e'_j, e'_{j+1}) = (a, b)$ for some $j \in \{0, \ldots, m-1\}$. See Figure 17 for an illustration of the situation. Then we have a path $e_{i+1} = b, \ldots, e_n = d = e'_0, e'_1, \ldots, e'_j = a$ in R from b to a, so that $(b, a) \in R$ by transitivity of R. This is a contradiction. So it must be that either (a, b) appears in the e_i's, or that it appears in the e'_i's, but not both.

Without loss of generality, suppose that there exists $i \in \{0, \ldots, n-1\}$ such that $(a, b) = (e_i, e_{i+1})$. Then the sequence $e_{i+1} = b, \ldots, e_n = d = e'_0, \ldots, e'_m = c = e_0, \ldots, e_i = a$ is a path from b to a in R, and since R is transitive this shows that $(b, a) \in R$, again a contradiction.

So there cannot be *any* position in those sequences in which we have a pair (a, b), but then it implies that $(c, d) \in R$ and $(d, c) \in R$, so that by antisymmetry of R, we get

[13] This is the step that is more difficult in the infinite case. For this, one needs a version of the axiom of choice in set theory.

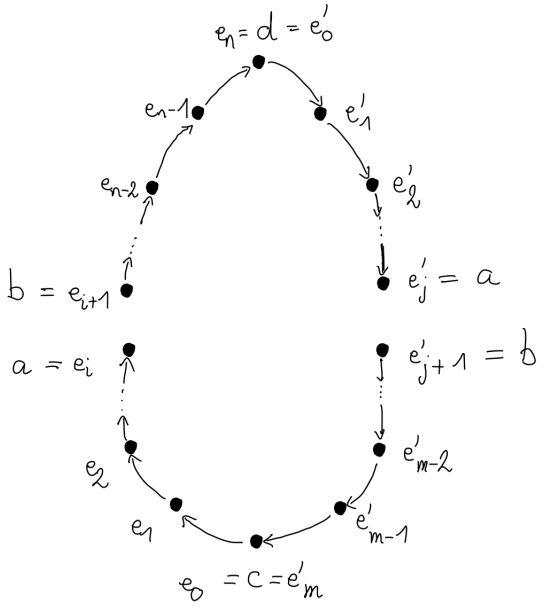


Figure 17: Illustration of the proof concerning the existence of linear completion.

$c = d$, a contradiction.

Therefore, R' is antisymmetric, and we are done. □

	reflexive	transitive	symmetric	antireflexive	antisymmetric
equivalence relation	✓	✓	✓		
quasiorder	✓	✓			
partial order	✓	✓			✓
strict partial order		✓		✓	✓

Figure 18: Summary of the types of relations and the properties that define them.

3.6 Exercises

1. Think about what reflexivity, transitivity, symmetry, antireflexivity, antisymmetry mean when we represent a relation as a directed network, or as lines between potatoes.

2. Suppose that $f\colon A \to B$ is a bijective function. Make sure you understand that f^{-1} is, as a relation, the same thing as $-f$.

3. Find an example of relations $R, S \subseteq A^2$ on a set A of your choice such that $(S + R) - R \neq S$.

4. (*) Let A be an arbitrary set. Find a characterization of the relations $R \subseteq A^2$ satisfying the following property: for all $S \subseteq A^2$, one has $(S + R) - R = S$.

5. Let $R \subseteq T \subseteq A^2$ and suppose that T is symmetric. Show that T contains the symmetric closure of R.

6. Let $R \subseteq T \subseteq A^2$ and suppose that T is reflexive. Show that T contains the reflexive closure of R.

7. Let $R \subseteq A^2$ be a relation. Show that $R + \Delta_A = \Delta_A + R = R$.

8. Show that $R \subseteq A^2$ is reflexive if, and only if, $\Delta_A \subseteq R$.

9. Let $R \subseteq A^2$. Show that R is symmetric if, and only if, $R = -R$.

10. Let $R \subseteq A^2$ be a relation. Show that R is transitive if, and only if, $R + R \subseteq R$. Give an example of a transitive relation where $R + R \neq R$.

11. Let $R \subseteq A^2$ be a reflexive relation on a finite set A. Show that there exists a number $n \geq 1$ such that the transitive closure of R is equal to $n \cdot R$.

12. Prove by induction that for all $n \geq 1$, $-(n \cdot R) = n \cdot (-R)$.

13. (*) Let $R \subseteq A^2$ be a relation on a finite set and let T be the reflexive closure of $R \cup (-R)$. Show that there exists $n \geq 1$ such that $n \cdot T$ is an equivalence relation.

14. Let $R \subseteq A^2$. Show that $(R \cup \Delta_A)^+ = R^+ \cup \Delta_A$.

15. Show that for every equivalence relation $E \subseteq A^2$, we have $\Delta_A \subseteq E$.

16. Show that every equivalence relation on $\{0, 1\}$ is trivial.

17. Compute the equivalence classes of Δ_A and ∇_A (over an arbitrary set A).

18. Recall the equivalence relation from Example 3.10. Determine what are the elements of the equivalence class of 2.

19. Let $R = \{(a, b) \in \mathbb{R}^2 \mid |a - b| \leq 1\}$. Is R an equivalence relation?

20. Explain why the relation $R' = \{(a, b) \in \mathbb{Z}^2 \mid a \text{ divides } b\}$ is *not* an order. Show that it is a quasiorder.

21. Let R be a quasiorder on A. Show that the relation $E = \{(a, b) \in A^2 \mid (a, b) \in R \text{ and } (b, a) \in R\}$ is an equivalence relation on A.

22. Let S be the set of binary strings of finite length. For $s, t \in S$, define $s \preceq t$ if s is a *prefix* of t. So for example $0 \preceq 01, 0110 \preceq 01100$ but $010 \not\preceq 0110$. Show that $\preceq$ is a partial order, and show that it is not a linear order. Draw the Hasse diagram of $(S, \preceq)$ (or some part of it).

23. What is the least upper bound of the set $\{5, 7, 105\}$ in Example 3.32? What about $\{2, 3, 30\}$?

24. Think about the properties of the relation given by your preferred social network: (a, b) is in the relation iff a follows/is a friend of/... b. Think about the properties of this relation, and whether it is an equivalence relation, a quasiorder, an order, ...

4 Cardinals and Counting

Things to remember / to know

- Know the various kind of ways to draw from a set: with and without replacement, and where order matters/does not matter.
- Understand the notion of partition and how to prove recurrence relations about their number (as in Proposition 4.16).
- Know how to compute the size of the intersection or union of two sets.
- Know how to use the principle of inclusion-exclusion to compute the size of a union of three sets.
- Know how to translate an equality of numbers into a statement about bijections.
- Know how to prove an equality between numbers using a combinatorial proof.
- Know the notion of countable sets and be able to prove that a set is countable.

When designing algorithms, one often iterates over a collection of things. For example, if n is an integer,

```
for i in range(n):
  do_something(i)
```

loops over all the integers $0, \dots, n-1$ and does something. It is easy to know how many times the loop is going to be executed, since the cardinality of $\{0, \dots, n-1\}$ is n. In a more advanced code one might find

```
for a in product(range(n),repeat=k):
  do_something(a)
```

which iterates over all the elements of $\{0, \dots, n-1\}^k$ (i.e., all tuples of length k and whose elements are in $\{0, \dots, n-1\}$, or something like

```
for a in combinations(range(n),2):
  do_something(a)
```

which iterates over the set $\mathcal{P}_2(\{0, \dots, n-1\})$ of subsets of $\{0, \dots, n-1\}$ that have two elements.

When writing such loops, it is important to know how long it is going to take for the execution to finish: it is useless to write complicated code that will finish running after 10 years or more!

The size of a set A is called its *cardinal* and is denoted by $|A|$.[14] It is easy to figure out what this means when A has finitely many elements:

cardinality

[14] Alternative notations you might find in some textbooks are $\#A$ and $\mathrm{Card}(A)$.

- $\{0, 1, 2\}$ has size 3,
- $\emptyset$ has size 0,
- $\{0, \{1, 2\}\}$ has size 2 (its only 2 elements are 0 and $\{1, 2\}$).

This chapter is dedicated to the problem of determining the cardinality of such sets; **here, all sets that we consider are finite**. The size of infinite sets is discussed in Chapter 4.6.

Another motivation for knowing how to determine the size of a set A comes from probability theory and random processes. It is sometimes a very good solution to have an algorithm pick an object at random instead of going through all the possible objects under consideration. In order to estimate the likelihood that a desired outcome occurs, one must count on the one hand how many "successful" options exist, and how many options there exist in total. For example, the probability that one rolls an even number with a normal 6-sided die is $3/6$, since there are 3 even numbers in $\{1, 2, \ldots, 6\}$.

In order to count (or estimate) the number of elements of a set A, the fundamental principle of counting is that if $f\colon A \to B$ is a bijective function (resp. injective, surjective), then $|A| = |B|$ (resp. $|A| \leq |B|$, $|A| \geq |B|$). Provided we can count the number elements of B, this gives a way to count the number of elements of A.

This goes the other way, too: if we want to prove that two numbers n and m are equal, it is sometimes easier to construct sets A, B such that $|A| = n$, $|B| = m$, and then show that there is a bijection between A and B. We call this a *combinatorial proof*. combinatorial proof

4.1 Rules for unions and intersections

Definition 4.1. The *disjoint union* of A and B is the set $(A \times \{0\}) \cup (B \times \{1\})$ denoted by $A \uplus B$. disjoint union

Note that the sets $(A \times \{0\})$ and $(B \times \{1\})$ do not intersect, hence the name "disjoint" union. Moreover, there is a bijection $A \to A \times \{0\}$ obtained by defining $f(a) = (a, 0)$ and similarly a bijection $B \to B \times \{1\}$ obtained by defining $g(b) = (b, 1)$. Therefore, by Lemma 4.28, we obtain that $|A \times \{0\}| = |A|$ and $|B \times \{1\}| = |B|$.

Thus, $|A \uplus B|$ is exactly $|A| + |B|$.

Theorem 4.2 (Cardinality and set constructions)**.** *Let A and B be two finite sets. Then the following hold:*

- $|A \cup B| \leq |A| + |B|$,
- $|A \cap B| \leq \min\{|A|, |B|\}$,
- $|A \cup B| = |A| + |B| - |A \cap B|$,

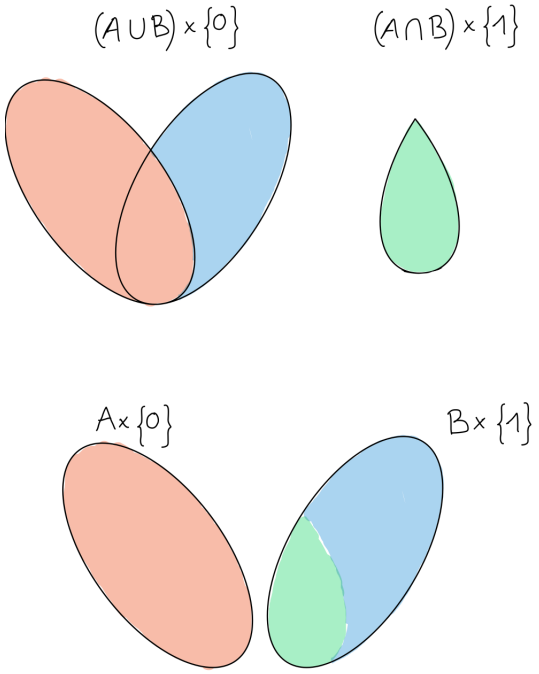


Figure 19: Illustration of the proof of $|A \cup B| + |A \cap B| = |A| + |B|$.

Proof. For the first item, we define an injective function $f\colon A \cup B \to A \uplus B$ as follows. If $a \in A$, then define $f(a) = (a, 0)$, and if $b \in B \setminus A$, define $f(b) = (b, 1)$. This is injective (proof left as an exercise), and therefore $|A \cup B| \leq |A \uplus B| = |A| + |B|$.

Since $A \cap B \subseteq A$ and $A \cap B \subseteq B$, we have $|A \cap B| \leq |A|$ and $|A \cap B| \leq |B|$. Therefore, $|A \cap B|$ must be at most the minimum of $|A|$ and $|B|$.

Intuitively, on the right hand side $|A| + |B|$ counts every element $a \in A \cap B$ exactly twice (once in $|A|$ and once in $|B|$). So by removing $|A \cap B|$, we have counted every element of $A \cup B$ exactly once. Formally, we define a bijection $g\colon (A \cup B) \uplus (A \cap B) \to A \uplus B$. Let $a \in A \cup B$, so that $(a, 0) \in (A \cup B) \uplus (A \cap B)$. Define $g(a, 0) = (a, 0)$ if $a \in A$ or $g(b, 0) = (b, 1)$ if $b \in B \setminus A$. Now given $a \in A \cap B$, define $g(b, 1) = (b, 1)$. A picture corresponding to the function is given in Figure 19. We show that g is a bijection, starting with injectivity. Suppose that $g(a, i) = g(b, j)$. Note that $g(a, i) \in \{(a, 0), (a, 1)\}$ and $g(b, j) \in \{(b, 0), (b, 1)\}$. If $g(a, i) = g(b, j)$, then $a = b$. Suppose for contradiction that $i \neq j$, and up to permuting the two elements we can assume that $i = 0, j = 1$. Therefore, $g(b, 1) = (b, 1)$ and $g(a, 0) = (a, 1)$, which means that $a \in B \setminus A$, but $b \in A \cap B$, a contradiction.

We now turn to prove that g is surjective. Every pair $(a, 0)$ has the preimage $(a, 0)$, and every pair $(b, 1)$ has the preimage $(b, 0)$ or $(b, 1)$ depending on whether $b \in B \setminus A$ or $b \in A \cap B$. Thus, g is surjective. □

Note that in particular, if A and B are disjoint (i.e., $A \cap B = \emptyset$), then $|A \cup B| =$

$|A| + |B|$. This holds for an arbitrary number of sets, if $A_1, \ldots, A_k$ are sets such that $A_i \cap A_j = \emptyset$ whenever $i \neq j$, then we have $\left|\bigcup_{i=1}^{k} A_i\right| = |A_1| + \cdots + |A_k|$.

Notation 4.3. If I is a set and for every i, one has a number a_i, then we write $\sum_{i \in I} a_i$ for the sum over all $i \in I$ of the numbers a_i. It is equivalent to the result of the following code:

```
result = 0
for i in I:
  result = result + a[i]
```

In particular, if I is empty, then $\sum_{i \in I} a_i$ is 0. If I is an interval of numbers $\{a, a+1, \ldots, b\}$, then we can write $\sum_{i=a}^{b} a_i$ or $\sum_{a \leq i \leq b} a_i$ instead.

Using this notation, we have that if $A_1, \ldots, A_k$ are *pairwise disjoint* (i.e., we have $A_i \cap A_j = \emptyset$ whenever $i \neq j$), then $\left|\bigcup_{i=1}^{k} A_i\right| = \sum_{i=1}^{k} |A_i|$. This is sometimes called the *union rule*.

pairwise disjoint

union rule

4.2 Number of possible ways to draw from a set

Suppose we have a set A with n elements, and we want to draw k elements from it. There are several "modes" of drawing and of considering the outcome: once an element is drawn from A, one might put it back into it (so that it can be drawn again later, we call this a draw *with replacement*) or not (called a draw *without* replacement), and one might pay attention to the order in which the elements have been drawn (in which case a draw is a *tuple*) or not (in which case a draw is a *set*[15]).

replacement

4.2.1 Putting back, order matters

We take the elements one by one and we put them back into the set every time, and we just record on a piece of paper the sequence of elements that we see. So if A has the elements $a_1, \ldots, a_n$, and $k = 3$, we might record on our piece of paper (a_1, a_1, a_3) or (a_2, a_3, a_2), or (a_3, a_2, a_2). The possible draws are exactly the elements of A^k. Every time we draw an element, there are n possible values for the element we get, so in total we have $n \times n \times \cdots \times n = n^k$ possible draws, so $|A^k| = |A|^k$. In Python, such elements are enumerated by the construction

```
itertools.product(A, repeat=k)
```

Another way to phrase the same result is to talk about the number of functions from B to A.

Proposition 4.4. *The number of functions $f\colon B \to A$ is $|A|^{|B|}$.*

[15] Or as we will see below rather a multiset, in case we are drawing with replacement.

	0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	2			
3	1	3	6	6		
4	1	4	12	24	24	
5	1	5	20	60	120	120

Figure 20: The falling factorials

Proof. Let k be the number of elements in B, and let $B = \{b_1, \dots, b_k\}$. There is a bijection between the set of possible draws $(s_1, \dots, s_k) \in A^k$ and functions $f\colon B \to A$. The bijection sends a draw $(s_1, \dots, s_k)$ to the function $f\colon b_1 \mapsto s_1, b_2 \mapsto s_2, \dots, b_k \mapsto s_k$. □

Notation 4.5. We sometimes write A^B to denote the set of all functions $B \to A$. Then it is easy to remember that $|A^B| = |A|^{|B|}$.

A^B

It could be that at every step, we draw an element from a different set. If we have sets $A_1, \dots, A_k$, each of size n_i and we draw an element from each set, then a draw is an element of $A_1 \times \cdots \times A_k$. There are n_1 possible elements for the first draw, n_2 elements for the second, ...so that there are in total $n_1 \times n_2 \times \cdots \times n_k$ possible ways to draw. We get the *product rule*, $|A_1 \times \cdots \times A_k| = \prod_{i=1}^{k} |A_i|$, where $\prod$ is to multiplication the same as $\sum$ is to addition.

product rule

4.2.2 Not putting back, order matters

This time, as soon as an element has been drawn we discard it. So every element can be drawn at most once. The draws correspond to tuples $(s_1, \dots, s_k)$ where $s_i \neq s_j$ for every $i \neq j$. The first element can be any of the n elements from the set A, the second time we draw there are $n-1$ possible elements, then $n-2$, etc... So there are $n \times (n-1) \times \cdots \times (n-k+1)$ possible draws. If we define $n! = n \times (n-1) \times \dots 3 \times 2$ to be the *factorial of* n, then this is exactly $n!/(n-k)!$. This number is sometimes also denoted $n^{\underline{k}}$ and called the *falling factorial*. Figure 20 contains the falling factorials $n^{\underline{k}}$, where n indexes the rows and k indexes the columns.

factorial of n

falling factorial

```
itertools.permutations(A,k)
```

In the case where $A = \{0, 1, 2, 3\}$ (so $n = 4$) and $k = 2$, this would generate all the tuples

```
(0,1),(0,2),(0,3),(1,0),(1,2),(1,3),(2,0),(2,1),(2,3),(3,0),(3,1),(3,2)
```

In particular, if $k = n$ then we are counting precisely the number of ways to arrange *all* the elements of A in some order. This is called a *permutation* of A, and by the above there are exactly $n!$ of them.

permutation

Another way to phrase the same result is to talk about injective functions rather than drawing from a set A without putting back the elements.

	0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

Figure 21: Pascal's triangle

Proposition 4.6. *The number of injective functions $f\colon B \to A$ is $n!/(n-k)! = n^{\underline{k}}$, where $n = |A|$ and $k = |B|$.*

Proof. Let $B = \{b_1, \ldots, b_k\}$. Same discussion as before, noting that there is a bijection between the set of possible draws described here, and injective functions $f\colon B \to A$. The bijection sends a draw $(s_1, \ldots, s_k)$ to the function $f\colon b_1 \mapsto s_1, b_2 \mapsto s_2, \ldots, b_k \mapsto s_k$. □

4.2.3 Not putting back, order does not matter

Same as before, but this time we do not care about the order in which the elements came out. The draw a_1, a_2, a_4 is considered the same as the draw a_2, a_1, a_4. In other words, we only consider about the *set* consisting of the elements $\{a_1, a_2, a_4\}$. We define the number of possible k-element subset of an n-element set to be $\binom{n}{k}$, pronounced "n choose k" and called *binomial coefficient*. We have $\binom{n}{k} = n!/(n-k)!k!$, since every set $\{a_1, \ldots, a_k\}$ was counted $k!$ times in the previous case.[16]

binomial coefficient

This kind of object is produced by

```
itertools.combinations(A,k)
```

which for our example with $A = \{0,1,2,3\}$ and $k = 2$ gives[17]

```
(0,1),(0,2),(0,3),(1,2),(1,3),(2,3)
```

We often see the binomial coefficients arranged in a triangle, where each row corresponds to an n and each column corresponding to $k \in \{0, \ldots, n\}$. This is *Pascal's triangle* (see Figure 21), named after Blaise Pascal, a French mathematician and philosopher.

Pascal's triangle

Proposition 4.7. *Let A be a set with n elements. Then $\mathcal{P}(A)$ has size $\sum_{k=0}^{n} \binom{n}{k}$.*

Proof. We have that $\mathcal{P}(A)$ is the union $\bigcup_{k=0}^{n} \mathcal{P}_k(A)$, where $\mathcal{P}_k(A)$ is the set of subsets of A having exactly k elements. By what we just discussed, we have $|\mathcal{P}_k(A)| = \binom{n}{k}$. Then

[16] A formal proof of this is given in Chapter 4.3.1.

[17] Note that Python produces *tuples* instead of sets, but these are essentially the same objects.

by the previous section,

$$\begin{aligned}|\mathcal{P}(A)| &= \left|\bigcup_{k=0}^{n} \mathcal{P}_k(A)\right| \\ &= \sum_{k=0}^{n} |\mathcal{P}_k(A)| \\ &= \sum_{k=0}^{n} \binom{n}{k}.\end{aligned}$$

□

Note that there is another way to count the number of elements of $\mathcal{P}(A)$. Define a function from $\mathcal{P}(\{1,\dots,n\})$ to the set of strings of length n and with symbols 0 or 1 as follows: given $S \subseteq \{1,\dots,n\}$, let $f(S)$ be the string $s_1 \dots s_n$ where $s_i = 1$ if $i \in S$ and $s_i = 0$ otherwise. Then f is a bijective function. A string of length n is obtained by drawing n times with replacement from a size of size 2 and where order matters, therefore by our discussion above there are exactly 2^n such strings. It follows that $|\mathcal{P}(A)| = 2^{|A|}$.

Another identity that one can obtain by a combinatorial proof is the following:

Lemma 4.8 (Vandermonde's identity)**.** *Let n, m be integers and $k \in \{0,\dots,n+m\}$. Then $\binom{n+m}{k} = \sum_{r=0}^{k} \binom{n}{r}\binom{m}{k-r}$.*

Proof. We give a combinatorial proof by giving a set whose size is the number on the left, and showing to partition it into parts $P_0,\dots,P_k$ where each part P_r has size $\binom{n}{r}\binom{m}{k-r}$. For a set A and $i \in \mathbb{N}_0$, let $\mathcal{P}_i(A)$ be the set of all subsets of A that have size exactly i. Let A, B be sets of size n and m, respectively, such that $A \cap B = \emptyset$ (e.g., $A = \{1,\dots,n\}$, and $B = \{n+1,\dots,n+m\}$). Let P_r be the set of all $S \in \mathcal{P}_k(A \cup B)$ such that $|S \cap A| = r$. Note that automatically, if $S \in P_r$ then $|S \cap B| = k - r$ since $|S| = |S \cap A| + |S \cap B|$. Then $\{P_0,\dots,P_k\}$ is a partition of $\mathcal{P}_k(A \cup B)$ and in particular $\binom{n+m}{k} = \sum_{r=0}^{k} |P_r|$.

Let $r \in \{0,\dots,k\}$. We define a bijection $f\colon P_r \to \mathcal{P}_r(A) \times P_{k-r}(B)$ by $f(S) = (S \cap A, S \cap B)$. It is clear that f is injective: if $S \neq T$, then one must have $S \cap A \neq T \cap A$ or $S \cap B \neq T \cap B$. It is also surjective: given a pair of sets $S \in \mathcal{P}_r(A)$ and a set $T \in \mathcal{P}_{k-r}(B)$, then $S \cup T \in \mathcal{P}_k(A \cup B)$ and is such that $f(S \cup T) = (S, T)$.

Thus, we have $|\mathcal{P}_k(A \cup B)| = \sum_{r=0}^{k} |\mathcal{P}_r(A)| \times |\mathcal{P}_{k-r}(B)|$. The quantity on the left is $\binom{n+m}{k}$, while on the right it is $\sum_{r=0}^{k} \binom{n}{r}\binom{m}{k-r}$. □

Giving a pure arithmetic proof of Lemma 4.8 would be a lot more complicated, try it out for yourself!

4.2.4 Putting back, order does not matter

Here, we count possible *multisets* of elements of A. A multiset is like a set (a bag in which elements are not sorted in any order), except that this time an element can appear mul-

multiset

tiple times in it. We often write multisets with double brackets, like $\{\{a_1, a_1, a_3, a_3, a_4\}\}$. The number of times an element appears in a multiset is called the *multiplicity* of that element. Let A be a set of size n and suppose we want to draw k elements from it. In order to count the number of multisets whose elements come from A, we define a bijection f that takes a multiset as input and produces another kind of object that will be easier to count. Since bijections preserve the number of elements, this is legal. Let us consider the set S of strings of length $n+k-1$ and consisting of two possible symbols $\star$ and $|$, and in which there are exactly k times the symbol $\star$ and $n-1$ times the symbol $|$. Given a multiset $\mathcal{M} = \{\{a_1, \ldots, a_1, a_2, \ldots, a_2, \ldots, a_n, \ldots, a_n\}\}$ where a_i has multiplicity m_i, define $f(\mathcal{M})$ to be the string $\star\cdots\star \mid \star\cdots\star \mid \cdots \mid \star\cdots\star$, where the number of $\star$ in the first block is m_1, then m_2, and so on until m_n. Conversely, given a string in S, one can define a multiset $\mathcal{M}$ that is a preimage for this string under f, so that f is a bijection between the two sets. Every element of S is determined by where the $|$ are put, so it is completely determined by a set of k positions within a set of $n+k-1$ possible positions. As counted above, there are exactly $\binom{n+k-1}{k}$ possible such sets, so the number of multisets of elements of A with k elements is exactly $\binom{n+k-1}{k}$.

This kind of objects is produced by

```
itertools.combinations_with_replacement(A,k)
```

4.3 Other counting techniques

4.3.1 Double counting

Double counting refers to the technique of determining the cardinality of a set in two different ways, which gives an equality between different quantities. This is particularly powerful in the following situation. Suppose that we have two sets A, B, and a relation $R \subseteq A \times B$. Suppose furthermore that what we want to count are the elements of R, i.e., we want to know $|R|$.

For $a \in A$, define $R(a)$ to be $\{(a,b) \in R \mid b \in B\}$, and similarly for $b \in B$ define $R(b)$ to be $\{(a,b) \in R \mid a \in A\}$. Let $r(a) = |R(a)|$ and $r(b) = |R(b)|$, which are called the*degrees* of a and b.

degree

Proposition 4.9. *Let $R \subseteq A \times B$. Then $|R| = \sum_{a\in A} r(a) = \sum_{b\in B} r(b)$.*

Proof. We can write R as a union $\bigcup_{a\in A} R(a)$ and $\bigcup_{b\in B} R(b)$. Note moreover that $R(a) \cap R(a') = \emptyset$ and $R(b) \cap R(b') = \emptyset$, for every $a, a' \in A$ with $a \neq a'$ and $b, b' \in B$ with $b \neq b'$. Then we must have $|R| = \left|\bigcup_{a\in A} R(a)\right| = \sum_{a\in A} |R(a)| = \sum_{a\in A} r(a)$, and similarly $|R| = \sum_{b\in B} r(b)$. □

This technique is particularly useful in *graph theory*, where we want to study properties of graphs (or networks, as discussed in Chapter 3). In the following, a *graph* G is a set $V(G)$ of *vertices*, together with some *edges* connecting two distinct vertices. There can be two edges that connect the same two vertices. For $v \in V(G)$, we write $d(v)$ for

graph

vertices, edges

the *degree* of v, which is the number of edges that touch v, also called edges *incident* to v.

degree

Lemma 4.10 ("Handshake lemma")**.** *Let G be a graph with n vertices and m edges. Then $\sum_{v \in V(G)} d(v) = 2m$.*

Proof. Let $A = V(G)$, $B = E(G)$, and let $R \subseteq A \times B$ contain a pair (v, e) if e is incident to v. Then $r(v)$ for $v \in A$ is exactly $d(v)$, while $r(e) = 2$ for every $e \in B$, so that $\sum_{e \in E(G)} r(e) = 2|E(G)| = 2m$. By Proposition 4.9, we get $\sum_{v \in V(G)} d(v) = 2m$. □

We can also use the principle of double counting to rigorously prove that $\binom{n}{k} = n!/(n-k)!k!$. Let $A = \mathcal{P}_k(\{1, \ldots, n\})$ and B be the set of tuples $(a_1, \ldots, a_k)$ such that $a_i \neq a_j$ for all $i \neq j$. Let $R \subseteq A \times B$ contain the pairs $(S, (a_1, \ldots, a_k))$ such that $S = \{a_1, \ldots, a_k\}$. Note that $|B| = n!/(n-k)!$, as B consists of draws without replacement and where order matters, and that $|A| = \binom{n}{k}$ by definition. Moreover, $r(b) = 1$ for each element $b \in B$, while $r(a) = k!$ for every $a \in A$ (this is the number of ways to order a given k-element set, as we have counted above). Since we have $\sum_{a \in A} r(a) = \sum_{b \in B} r(b)$, this gives $|A|k! = |B|$, i.e., $\binom{n}{k}k! = n!/(n-k)!$, so that $\binom{n}{k} = n!/(n-k)!k!$.

4.3.2 The inclusion-exclusion principle

Theorem 4.11. *Let $A_1, \ldots, A_k$ be finite sets. Then*

$$\left| \bigcup_{i=1}^{k} A_i \right| = \sum_{r=1}^{k} (-1)^{r+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq k} |A_{i_1} \cap \cdots \cap A_{i_r}|,$$

where the sum on the right-hand side ranges over all tuples $(i_1, \ldots, i_r) \in \{1, \ldots, k\}^r$ such that $1 \leq i_1 < \cdots < i_r \leq k$.

Proof. We prove the result by induction on k. Note that for $k = 1$ the result is true since the left-hand side is A_1 and the right-hand side is $(-1)^{1+1}|A_1|$. For $k = 2$ the result is also true by Theorem 4.2: the left-hand side is $|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2|$.

So let us assume that $k \geq 3$ and that we already know the result for $k-1$. Let B be $A_1 \cup \cdots \cup A_{k-1}$. Then $|B \cup A_k| = |B| + |A_k| - |B \cap A_k|$, by Theorem 4.2. By induction hypothesis, we also have

$$|B| = \sum_{r=1}^{k-1} (-1)^{r+1} \sum_{1 \leq i_1 < i_2 < \cdots < i_r \leq k-1} |A_{i_1} \cap \cdots \cap A_{i_r}|. \tag{4.1}$$

Moreover, $B \cap A_k = (A_1 \cap A_k) \cup \cdots \cup (A_{k-1} \cap A_k)$. Write C_i for $A_i \cap A_k$, for $i \in \{1, \ldots, k-1\}$.

Then $|B \cap A_k| = |C_1 \cup \cdots \cup C_{k-1}|$ and by induction hypothesis, this is

$$\sum_{r=1}^{k-1}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r\le k-1} |C_{i_1} \cap \cdots \cap C_{i_r}|$$

which can be rewritten as

$$\begin{aligned}
|B \cap A_k| &= \sum_{r=1}^{k-1}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r\le k-1} |A_{i_1} \cap \cdots \cap A_{i_r} \cap A_k| \\
&= \sum_{r=1}^{k-1}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r<i_{r+1}=k} \left|A_{i_1} \cap \cdots \cap A_{i_r} \cap A_{i_{r+1}}\right| \\
&= \sum_{r=2}^{k}(-1)^{r} \sum_{1\le i_1<i_2<\cdots<i_r=k} |A_{i_1} \cap \cdots \cap A_{i_r}|
\end{aligned}$$

Putting everything together, we obtain

$$\begin{aligned}
|B \cup A_k| &= |B| + |A_k| - |B \cap A_k| \\
&= \left[\sum_{r=1}^{k-1}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r\le k-1} |A_{i_1} \cap \cdots \cap A_{i_r}|\right] + |A_k| - \left[\sum_{r=2}^{k}(-1)^{r} \sum_{1\le i_1<i_2<\cdots<i_r=k} |A_{i_1} \cap \cdots \cap A_{i_r}|\right] \\
&= \left[\sum_{r=1}^{k-1}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r\le k-1} |A_{i_1} \cap \cdots \cap A_{i_r}|\right] + |A_k| + \left[\sum_{r=2}^{k}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r=k} |A_{i_1} \cap \cdots \cap A_{i_r}|\right] \\
&= \sum_{r=1}^{k}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r\le k} |A_{i_1} \cap \cdots \cap A_{i_r}|\,.
\end{aligned}$$

To see that this last equality is true, consider the set of sequences $(i_1, \ldots, i_r)$ such that $1 \le i_1 \le \cdots \le i_r \le k$. There are two types of such sequences: those where k is an element of the sequence (and in this case this can only be i_r), and those where k is not. The sequences of the second type already all appear in the first term $\sum_{r=1}^{k-1}(-1)^{r+1} \sum_{1\le i_1<i_2<\cdots<i_r\le k-1} |A_{i_1} \cap \cdots \cap A_{i_r}|$. The sequences of the first type are either of length 1 (which corresponds to the second term $|A_k$ in the sum above), or all the sequences of length greather than 1, which all appear in the third term. □

So far, we have counted the number functions $B \to A$ (equal to $|A|^{|B|}$), the number of injective functions $B \to A$ (which is equal to $|A|^{\underline{|B|}}$), and the number of bijective functions $B \to A$ (equal to $|A|!$ if $|B| = |A|$, and 0 otherwise). With the principle of inclusion-exclusion, we are finally able to compute the number of surjective functions.

Theorem 4.12. *Let A be a set of size n and B be a set of size k. The number of surjective functions $B \to A$ is $\sum_{r=0}^{n}(-1)^r\binom{n}{r}(n-r)^k$.*

Proof. For simplicity of notation, let $A = \{1, \dots, n\}$. For $i \in \{1, \dots, n\}$, let F_i be the set of functions $f\colon B \to A$ whose image does not contain i. Note that the *non-surjective* functions are exactly the elements of the set $\bigcup_{i=1}^{n} F_i$, but the union is not disjoint (an f whose image does not contain 1 and 2 is in both F_1 and F_2). This is our signal to use Theorem 4.11 to determine the size of this set. Given what we have seen above, we know that $|F_i| = (n-1)^k$ for all $i \in \{1, \dots, n\}$. Moreover, $|F_{i_1} \cap F_{i_2} \cap \cdots \cap F_{i_r}| = (n-r)^k$ whenever $i_1, \dots, i_r$ are pairwise distinct elements of A.

Theorem 4.11 gives

$$\begin{aligned}\left|\bigcup_{i=1}^{r} F_i\right| &= \sum_{r=1}^{n}(-1)^{r+1} \sum_{1\le i_1<\cdots<i_r\le n} |F_{i_1} \cap F_{i_2} \cap \cdots \cap F_{i_r}| \\ &= \sum_{r=1}^{n}(-1)^{r+1} \sum_{1\le i_1<\cdots<i_r\le n} (n-r)^k \\ &= \sum_{r=1}^{n}(-1)^{r+1}\binom{n}{r}(n-r)^k \qquad \text{(see Exercise 2).}\end{aligned}$$

Therefore, the number of *surjective* functions is $n^k - \sum_{r=1}^{n}(-1)^{r+1}\binom{n}{r}(n-r)^k$, and by observing that n^k is $(-1)^r\binom{n}{r}(n-r)^k$ for the case $r = 0$, this simplifies to $\sum_{r=0}^{n}(-1)^r\binom{n}{r}(n-r)^k$. □

4.3.3 Showing existence by counting

If $|A| > |B|$ and $f\colon A \to B$ is an arbitrary function, then f is not injective (this can be proved as an exercise, but follows from Proposition 4.13 below) so that there exist $a \neq b$ with $f(a) = f(b)$. The existence of the elements a and b can therefore be achieved only by computing the sizes of the sets A and B. A more quantitative version of this is the following. Given a number q, let $\lceil q \rceil$ be the smallest integer that is at least as large as q.

Proposition 4.13. *Let $f\colon A \to B$ be a function between finite sets. Then there exist at least $K := \left\lceil \frac{|A|}{|B|} \right\rceil$ different elements $a_1, \dots, a_K$ such that $f(a_1) = \cdots = f(a_K)$.*

Proof. We do the proof by contradiction. Consider f as a relation $R \subseteq A \times B$. The negation of the conclusion means that $r(b) \le K - 1$, for all $b \in B$. And $r(a) = 1$ for all $a \in A$, since f is a function. Thus, by Proposition 4.9, $|A| = \sum_{b\in B} r(b) \le (K-1)\cdot|B|$, i.e., $K \ge |A|/|B| + 1$. Since K is an integer, we have $K \ge \left\lceil \frac{|A|}{|B|} + 1 \right\rceil = \left\lceil \frac{|A|}{|B|} \right\rceil + 1$, a contradiction to the choice of K. □

Ramsey's theorem[18] is a version of Proposition 4.13 in "higher dimensions". Let us see the elements of B **??** and proposition 4.13 as *colors*, and the function f as a *coloring*: each element of A receives under f a single color. Then Proposition 4.13 says that if A is large enough compared to B, then there exists a medium-sized subset of A (of size roughly $|A|/|B|$) whose elements all receive the same color.

Ramsey's theorem is a similar statement, except that the function f does not color elements of A, but *subsets* of A.

Theorem 4.14 (Ramsey's theorem)**.** *For every $r, b \geq 1$, there exists a number N with the following property: for every finite set A with $|A| \geq N$, and every function $f\colon \mathcal{P}_2(A) \to \{red, blue\}$, at least one of the following holds:*

- *there exists a subset $R \subseteq A$ with at least r elements such that for every $x, y \in R$ with $x \neq y$, we have $f(\{x, y\}) = red$, or*
- *there exists a subset $B \subseteq A$ with at least b elements such that for every $x, y \in B$ with $x \neq y$, we have $f(\{x, y\}) = blue$.*

We write $R(r, b)$ for the smallest such number N.

Proof. We prove the following by induction: for every $S \geq 2$, and every $r, b \geq 1$ such that $r + b \leq S$, then there exists a number N as in the statement.

The base case of the induction is when $S = 2$. Then we must have $r = 1$ and $b = 1$. Let us take $N = 2$. Then for an arbitrary set A with at least 2 elements x and y, we must have that $f(\{x, y\})$ is either red or blue. Then take $R = A$ or $B = A$ and we are done.

We now suppose that we want to prove the result for a given $S > 2$, assuming the result has been proved for $S - 1$. In particular, we know the existence of a number $R(r - 1, b)$ and $R(r, b - 1)$ satisfying the statement. Let $N \geq R(r - 1, b) + R(r, b - 1)$, and let A be a set with at least N elements. Pick an arbitrary $x \in A$. Divide A into three sets: $\{x\}$, the set $A_r = \{y \in A \mid f(\{x, y\}) = \text{red}\}$ and $A_b = \{y \in A \mid f(\{x, y\}) = \text{blue}\}$. Then we must have $|A_r| \geq R(r - 1, b)$ or $|A_b| \geq R(r, b - 1)$: otherwise, we have $|A| = 1 + |A_r| + |A_b| \leq 1 + R(r - 1, b) - 1 + R(r, b - 1) - 1 < N$, a contradiction.

Suppose that $|A_r| \geq R(r-1, b)$. Then by definition of this number, there exists $R' \subseteq A_r$ of size at least $r - 1$ such that for all $y, z \in R'$ and $y \neq z$, we have $f(\{y, z\}) = \text{red}$). Let $R = R' \cup \{x\}$. Then R has size at least r, and moreover for all $y, z \in R$ with $y \neq z$, we have $f(\{y, z\}) = \text{red}$, so we are done.

The case where $|A_b| \geq R(r, b - 1)$ is proved similarly. □

A usual example of an application of this theorem is that in any group of at least 6 people, either 3 of them do not know each other, or 3 of them know each other. This means that $R(3, 3) \leq 6$, in fact one can show that $R(3, 3) = 6$. We know that

[18] Frank Ramsey was a British logician in the 20th century.

$R(4,4) = 18$, but we only know that $43 \leq R(5,5) \leq 48$. Computing the exact value of $R(5,5)$ would certainly be a big achievement in this field.

4.4 Partitions of a set

Remember that a partition of a set A is a set $P \subseteq \mathcal{P}(A)$ such that for any two $S, T \in P$, either $S \cap T = \emptyset$ or $S = T$, and for every $a \in A$, there exists $S \in P$ such that $a \in S$. Recall also that $\mathrm{Part}(A)$ is the set of partitions of A.

Each $S \in P$ is called a *part* of the partition. In this section, we count how many ways there are to partition a set A.

Definition 4.15. Let $0 \leq k \leq n$. We define $S(n,k)$ to be the number of partitions $P \in \mathrm{Part}(\{1,\ldots,n\})$ that have exactly k parts. This number is called a *Stirling number of the second kind*.

Stirling number of the second kind

We then have, by definition, that $|\mathrm{Part}(A)| = \sum_{k=1}^{n} S(n,k)$.

Proposition 4.16. *For all $1 \leq k \leq n$, one has*

$$S(n,k) = S(n-1,k-1) + kS(n-1,k).$$

Comment about the proof

Since we haven't seen a way to compute $S(n,k)$, the only possible way to prove this is a combinatorial proof. In the proof below, we give a bijection between suitably chosen sets. For this, we need to order the parts of a partition according to their minimum element. For example, $P = \{\{1,3\},\{2,6\},\{5,8\},\{6\},\{7\}\}$ and $P' = \{\{1,3\},\{2,6,8\},\{5\},\{6\},\{7\}\}$ is a partition of $\{1,\ldots,8\}$ into 5 parts, and their parts are numbered $S_1,\ldots,S_5$ in the order given here.
Following the proof, we define $f(P)$ to be $(3,\{\{1,3\},\{2,6\},\{5\},\{6\},\{7\}\})$ and $f(P')$ to be $(2,\{\{1,3\},\{2,6\},\{5\},\{6\},\{7\}\})$. Do you see why the factor k is important in the equality of Proposition 4.16?

Proof. Define $\mathrm{Part}_k(A)$ to be the set of partitions of A having exactly k parts. We want to define a bijective function f from $\mathrm{Part}_k(\{1,\ldots,n\})$ to $\mathrm{Part}_{k-1}(\{1,\ldots,n-1\}) \cup \{1,\ldots,k\} \times \mathrm{Part}_k(\{1,\ldots,n-1\})$. Using the results from this chapter, this would establish the equality between $S(n,k)$ and $S(n-1,k-1) + kS(n-1,k)$, since one observes that the set in the union above are disjoint.

Let P be a partition of an n-element set. We distinguish the following cases:

- suppose that $\{n\}$ is a part in P. Then $P \setminus \{\{n\}\}$ is a partition of $\{1,\ldots,n-1\}$ having exactly $k-1$ parts, so it is an element of $\mathrm{Part}_{k-1}(\{1,\ldots,n-1\})$. We define $f(P) = P \setminus \{\{n\}\}$.

- otherwise, let us enumerate all the parts $S_1, \dots, S_k$ of P in a way that if $i < j$, then the smallest element of S_i is smaller than the smallest element of S_j. Let ℓ be such that S_ℓ is the unique part of P containing n. Define Q to be the partition that has the same parts as P, except that S_ℓ is replaced by $S_\ell \setminus \{n\}$. In symbols, we have $Q = P \setminus \{S_\ell\} \cup \{S_\ell \setminus \{n\}\}$. Note that Q is a partition of $\{1, \dots, n-1\}$: for every $i \in \{1, \dots, n-1\} \setminus S_\ell$, then the part that contained i in P contains i in Q. If $i \in S_\ell$ and $i \neq n$, then $i \in S_\ell \setminus \{n\}$. Moreover, no two parts of Q intersect. Moreover, Q has as many parts as S does (we removed one part and added another). We define $f(P) = (\ell, Q)$.

It remains to show that f is bijective. This time, we give a function $g\colon \operatorname{Part}_{k-1}(\{1, \dots, n-1\}) \cup \{1, \dots, k\} \times \operatorname{Part}_k \to \operatorname{Part}_k(\{1, \dots, n\})$ such that $g \circ f$ and $f \circ g$ are the identity functions. By Lemma 1.21, this is enough.

Define g depending on which set from the union the argument is taken from:

- if the argument is $Q \in \operatorname{Part}_{k-1}(\{1, \dots, n-1\})$, define $g(Q) = P \cup \{\{n\}\}$.
- if the argument is $(\ell, Q) \in \{1, \dots, k\} \times \operatorname{Part}_k(\{1, \dots, n-1\}) \to \operatorname{Part}_k(\{1, \dots, n\})$, we first enumerate the parts $S_1, \dots, S_k$ of Q as above (if $i < j$ then the smallest element of S_i is smaller than the smallest element of S_j). Then we define $g(\ell, Q)$ to be $Q \setminus \{S_\ell\} \cup \{S_\ell \cup \{n\}\}$, i.e., the partition obtained by adding n to the part S_ℓ.

We leave as an exercise to show that $f \circ g$ and $g \circ f$ are identity functions. □

Second proof, using double counting. Define

$$R \subseteq \operatorname{Part}_k(\{1, \dots, n\}) \times (\operatorname{Part}_{k-1}(\{1, \dots, n-1\}) \cup \operatorname{Part}_k(\{1, \dots, n-1\}))$$

as the relation containing the pairs (P, Q), as in the proof above.

Then we have $r(P) = 1$ (since we have defined a function, the partition Q is always uniquely determined by the partition P). Similarly, for $Q \in \operatorname{Part}_{k-1}(\{1, \dots, n-1\})$, we have $r(Q) = 1$ (the only P that is mapped to Q is $Q \cup \{\{n\}\}$). However, for $Q \in \operatorname{Part}_k(\{1, \dots, n-1\})$, we have $r(Q) = k$, since for each of the k parts of Q, one could add n to that part and obtain a different P. Thus, by Proposition 4.9, we must have

$$\sum_P r(P) = \sum_{Q \in \operatorname{Part}_{k-1}(\{1,\dots,n-1\})} r(Q) + \sum_{Q \in \operatorname{Part}_k(\{1,\dots,n-1\})} r(Q)$$

which simplifies to

$$|\operatorname{Part}_k(\{1, \dots, n\})| = |\operatorname{Part}_{k-1}(\{1, \dots, n-1\})| + k \cdot |\operatorname{Part}_k(\{1, \dots, n-1\})|.$$

□

Bell numbers

The *Bell numbers* B_n are the number of ways to partition an n-element set. Thus, $B_n = \sum_{k=1}^{n} S(n, k)$. The Stirling numbers of the second kind can be computed using Proposition 4.16, noting that $S(n, n) = 1$ for $n \geq 0$ and $S(n, 1) = 1$ for $n \geq 1$, see Figure 23.

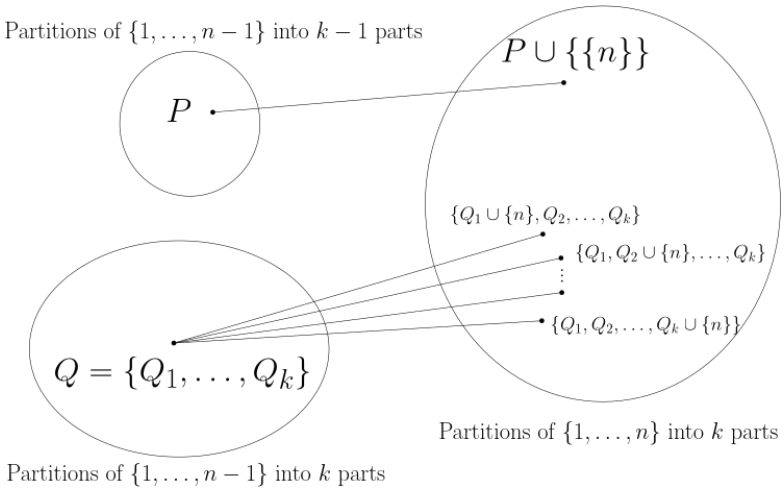


Figure 22: Proof by double counting of Stirling's recurrence formula

	1	2	3	4	5	6	B_n
1	1						1
2	1	1					2
3	1	3	1				5
4	1	7	6	1			15
5	1	15	25	10	1		52
6	1	31	90	65	15	1	203

Figure 23: Stirling numbers of the second kind

Note the following: if $\{S_1, \ldots, S_k\}$ is a partition of $\{1, \ldots, n\}$ into k parts, where the parts are given in some arbitrary order, we can define a function $f\colon \{1, \ldots, n\} \to \{1, \ldots, k\}$ such that $f(i)$ is the unique ℓ such that $i \in S_\ell$. In fact, f is a function such that $\ker(f)$ is exactly the equivalence relation whose equivalence classes are the parts $S_1, \ldots, S_k$! And additionally, we see that f is a surjective function. This gives the following result.

Proposition 4.17. *Let $s_{n,k}$ be the number of surjective functions from a set of size n to a set of size k. Then $s_{n,k} = k! \cdot S(n, k)$.*

Proof. Again, we use double counting. Let A be the set of surjective functions $\{1, \ldots, n\} \to \{1, \ldots, k\}$ and B be the set $\mathrm{Part}_k(\{1, \ldots, n\})$. Let $R \subseteq A \times B$ contain the pairs (f, P) if P is the equivalence classes of $\ker(f)$ are exactly P. Note that every $f \in A$, $\ker(f)$ has exactly k equivalence classes since f is surjective.

We have $r(f) = 1$ for all $f \in A$, since $\ker(f)$ is uniquely determined by f. We have $r(P) = k!$, since the k parts of a given partition can be ordered in $k!$ ways $S_1, \ldots, S_k$ and then associated with a surjective function $f\colon \{1, \ldots, n\} \to \{1, \ldots, k\}$ such that $f(m) = \ell$, where $m \in S_\ell$. By Proposition 4.9, we get $s_{n,k} = k! \cdot S(n, k)$. □

Those numbers get *really* big, even for modest values of n and k. When developing an algorithm, it is usually a bad move to simply iterate over all partitions/surjective functions. However, the idea of partitioning a set in a given number of parts is very natural and used daily in applications: given a set of n points in $\mathbb{R}^d$, a common problem in machine learning/data science is to try to group those points into k parts (in this setting, the parts are called *clusters*) in a way that some properties are satisfied (for example, points in the same cluster should be close together).

Another way to see $S(n, k)$ is as follows: each of the k parts is a bag (often called an *urn* in this context), and $S(n, k)$ counts the number of ways of distributing n different objects into k (indistinguishable) urns (where no urn ends up empty). If the urns have a number and we care about which object goes where, we are counting the number of surjective functions $\{1, \ldots, n\} \to \{1, \ldots, k\}$, which is $k!S(n, k)$, as we saw.

We do not see here the cases where the objects cannot be distinguished.

4.5 Asymptotic growth

One motivation for this chapter was to be able to analyse the runtime of certain loops in algorithms. We have seen so far some sequences of numbers coming from possible ways to draw from a set of partition a set. In order to analyse the performance of algorithms using such construction, we need to be able to compare the growth of these sequences. One of the most important pieces of notation for this is the following.

Notation 4.18. Let $f, g\colon \mathbb{N} \to \mathbb{N}$ be functions. We write:

- $f = O(g)$ if there exist $C \in \mathbb{N}$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $f(n) \leq C \cdot g(n)$

- $f = o(g)$ if for all real $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $f(n) \leq \epsilon \cdot g(n)$
- $f \sim g$ if $f(n)/g(n) \to 1$ as $n \to \infty$.

This is particularly useful in the following situations. If $f = O(g)$ and $f' = O(g)$, then $f + f' = O(g)$ as well, and similarly for $o(g)$. Moreover, $Cn^k = O(n^k)$ (for any constant value C and any $k \geq 0$), so the notation allows to forget about constants. Any function that is $O(1)$ is called *constant*, any function that is $O(n)$ is called *linear*, any function that is $O(n^2)$ is called *quadratic*.

The following estimation of $n!$ is due to Sterling, and without remembering the terms by heart, it helps figuring out the growth of $n!$ relative to polynomials. In the following, e is the natural exponential $\approx 2.71828\ldots$

Theorem 4.19.

$$n! \sim \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$$

This is a much more precise estimation than the trivial $n! \leq n^{n-1}$ one gets by expanding $n!$ intro $n \times (n-1) \times \cdots \times 2$.

The number of ways to draw from a set then have the following estimations, where we consider k to be a *constant*:

- With replacement, order matters: $n^k = O(n^k)$
- Without replacement, order matters: $n^{\underline{k}} = O(n^k)$
- Without replacement, order does not matter: $\binom{n}{k} = n^{\underline{k}}/k! = O(n^k)$
- With replacement, order does not matter: $\binom{n+k-1}{k} = O(n^k)$.

However, as stated in Exercise 11, one has $S(n,k) \geq k^{n-k}$.

Proposition 4.20. *$S(n,k)$ is not $O(n^\ell)$ for any ℓ.*

Proof. We show that $n^\ell = o(S(n,k))$ for *every* ℓ. Indeed, since n^ℓ/k^{n-k} tends to 0 for all constant values of k and ℓ, for every $\epsilon > 0$ there exists some integer n_0 for which $n^\ell/k^{n-k} < \epsilon$ for every $n \geq n_0$. Rearranging, this gives $n^\ell < \epsilon \cdot k^{n-k}$ and therefore $n^\ell = o(k^{n-k})$. Since $k^{n-k} \leq S(n,k)$ (Exercise 11), we also get $n^\ell = o(S(n,k))$.

Suppose now that $S(n,k)$ is $O(n^\ell)$. Then there exists a constant $C \in \mathbb{N}$ such that for large enough $n \geq n_0$, $S(n,k) \leq C \cdot n^\ell$. But from the paragraph above we know that $n^\ell = o(S(n,k))$ and therefore for large enough $n \geq n_1$, we also have $S(n,k) > (C+1)n^\ell$. Thus, for n larger than $\max(n_0, n_1)$, we get $Cn^\ell > (C+1) \cdot n^\ell$, a contradiction. □

4.6 Infinite sets

By the pigeonhole principle for finite sets (Proposition 4.13), if there exists an injective function f from A to B, then $|A| \leq |B|$. This is how we *define* the notion of cardinals for infinite sets.

Definition 4.21. Let A and B be arbitrary sets. We define "$|A| \leq |B|$" to mean "there exists an injective function $f\colon A \to B$." We write "$|A| < |B|$" to mean that $|A| \leq |B|$ and that there does not exist an injective function $B \to A$. We write $|A| = |B|$ if both $|A| \leq |B|$ and $|B| \leq |A|$ are true.

Remark 4.22. Note that in general, $|A|$ is not a number in the classical sense. At our level (for this course), we only care about *comparing* $|A|$ and $|B|$, but we don't have the tools to understand what $|A|$ really *is*. This is one of the topics studied in *set theory*. If you are interested, consult the Wikipedia entry about cardinals.

Since id_A is injective, we have that $|A| \leq |A|$ for every set A. Moreover, we have the natural fact that if $|A| \leq |B|$ and $|B| \leq |C|$, then also $|A| \leq |C|$ (this follows from Lemma 1.18).

Lemma 4.23. *Suppose that A is not the empty set. We have $|A| \leq |B|$ if, and only if, there exists a* surjective *function $g\colon B \to A$.*

Proof. We prove the left-to-right direction: we assume $|A| \leq |B|$, and give a surjective function. By assumption, $|A| \leq |B|$ means that there exists an injective function $f\colon A \to B$. Define g as follows. Let $b \in B$. There are two possible cases:

1. It can be that there exists $a \in A$ such that $f(a) = b$ (in fact, this a is even uniquely determined since f is injective. Then, let us define $g(b) := a$.
2. Otherwise, define $g(b)$ arbitrarily. Here, we use that A is not the empty set, otherwise $g(b)$ could not be defined!

Note that g is surjective: if we pick $a \in A$, we need to find $b \in B$ such that $g(b) = a$. Take $b = f(a)$. Then $g(b)$ was defined by using the first case above, and we have $g(b) = a$.

We now prove the right-to-left direction: we assume that $g\colon B \to A$ is a surjective function, and we give an *injective* function $f\colon A \to B$. For any $a \in A$, there exists at least one $b \in B$ such that $g(b) = a$ (this is because g is surjective). Take any such b and define $f(a) = b$.[19] So we have defined $f\colon A \to B$ and it remains to prove that it is injective. Let $a \neq a'$, and let $b = f(a), b' = f(a')$. By construction, we must have $g(b) = a$ and $g(b') = a'$, so it must be the case that $b \neq b'$, which is what we wanted. □

[19] In proper set theory, this step is controversial and depends on what axioms one is ready to accept (see https://en.wikipedia.org/wiki/Axiom_of_choice). We simply ignore this here.

Let $2\mathbb{N}$ denote the set of even natural numbers. Since $2\mathbb{N} \subseteq \mathbb{N}$, we have $|2\mathbb{N}| \leq \mathbb{N}$. But note that $|\mathbb{N}| \leq |2\mathbb{N}|$ is also true!

$2\mathbb{N}$

The following result is known as *Hilbert's hotel*. David Hilbert (1862–1943) was a prominent German mathematician who was interested at the turn of the century in rigorous foundations of mathematics. He made the following thought experiment: suppose that a hotel has infinitely many rooms, room R1, room R2, ..., and suppose that each room is occupied by a guest (guest G1 is in room R1 and so on). Is the hotel full? No! Assume a new guest arrives. The hotel manager asks all the previous guests to move one room to the right: guest 1 moves to room 2, guest 2 moves to room 3 and so on. So now room 1 is free, and the newly arrived guest can use it. In fact, even if infinitely many new guests arrived, call them G'1,G'2,G'3,..., we could still find rooms for them! Instead of asking every existing guest to move one room to the right, the hotel manager asks them to go to the room whose number is double the existing one: guest G1 goes to room R2, guest G2 goes to room R4, guest G3 goes to room R6, ... Now all the rooms R1, R3, R5, R7, ... are currently unoccupied. So G'1 can go to R1, G'2 can go to R3, G'3 can go to R5, and so on.

In mathematical language, this fun fact translates as follows.

Lemma 4.24 (Hilbert's hotel)**.** $|\mathbb{N}| \leq |2\mathbb{N}|$.

Proof. Let $f\colon \mathbb{N} \to 2\mathbb{N}$ be defined by $f(x) = 2x$. This function is injective: if $x \neq y$, then $2x \neq 2y$. □

Definition 4.25. We say that a set A is *countable* if $|A| \leq |\mathbb{N}|$. We say that A is *uncountable* otherwise.

countable set

uncountable set

Theorem 4.26. *The sets $\mathbb{Z}$ and $\mathbb{Q}$ are countable.*

Proof. We need to prove $|\mathbb{Z}| \leq |\mathbb{N}|$ and $|\mathbb{Q}| \leq |\mathbb{N}|$.

We prove the first statement by giving an injective function $f\colon \mathbb{Z} \to \mathbb{N}$. Define $f(x)$ either as $2x$ if $x \geq 0$, and $f(x) = -2x + 1$ if $x < 0$. We prove that it is injective. Let $x \neq y$ be integer numbers. We distinguish a few cases:

- if $x, y \geq 0$: then $f(x) = 2x, f(y) = 2y$, and since $x \neq y$ we have $2x \neq 2y$.
- if $x \geq 0$, $y < 0$: then $f(x) = 2x$, $f(y) = -2y + 1$. In particular, $f(x)$ is an even number, and $f(y)$ is an odd number, so $f(x)$ and $f(y)$ must be different.
- if $x, y < 0$, then $f(x) = -2x + 1$ and $f(y) = -2y + 1$, so $f(x) \neq f(y)$.

To prove that $|\mathbb{Q}| \leq |\mathbb{N}|$, remember that $\mathbb{Q} = \{p/q \mid p \in \mathbb{Z}, q \in \mathbb{N}, q \neq 0\}$. Every rational number has a unique *irreducible* fraction, where q and p do not share a common divisor. We define an injective function $g\colon \mathbb{Q} \to \mathbb{N}$ by defining $g(a) = 2^{f(p)}3^q$, where p/q

is the irreducible fraction corresponding to a, and f is the injection $\mathbb{Z} \to \mathbb{N}$ constructed above. We prove that g is an injective function by contraposition: we assume that the conclusion is false (i.e., we assume that $g(p/q) = g(p'/q')$), and we prove that the assumption is false (i.e., we prove that $p/q \neq p'/q'$). So, let us suppose that we have $2^{f(p)}3^q = 2^{f(p')}3^{q'}$. The *fundamental theorem of arithmetic* which we will prove later in the course says that every number has a unique decomposition into prime factors. Since 2 and 3 are prime numbers, this implies here that $f(p) = f(p')$ and $q = q'$. Since f is injective, we have $p = p'$, and therefore $p/q = p'/q'$. □

Theorem 4.27 (Cantor-Bernstein-Schröder)**.** *If $|A| = |B|$, then there exists a bijective function $f\colon A \to B$.*

We don't prove Theorem 4.27 now. But it implies that there are bijections between $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. It might seem that in fact, every set is countable! Note that the converse of Theorem 4.27 holds:

Lemma 4.28. *If there exists a bijective function $f\colon A \to B$, then $|A| = |B|$.*

Proof. Since f is an injective function, then $|A| \leq |B|$. Since f is a surjective function, we get by Lemma 4.23 that $|B| \leq |A|$, and thus $|A| = |B|$. □

The definition and what we have proved so far (Definition 4.21, lemmas 4.23 and 4.28, and theorem 4.27) give us the following intuitive way to think about $|A|$; but keep in mind that this dictionary can give us surprises (as in Lemma 4.24):

there exists a...	if, and only if,	A **...**
injection $f\colon A \to B$		is smaller (or equal) in size than B
surjection $f\colon A \to B$		is bigger (or equal) in size than B
bijection $f\colon A \to B$		has the same size as B.

It could *a priori* be the case that any time we have two infinite sets A and B, then in fact $|A| = |B|$. We have seen that this is true for several infinite sets like $\mathbb{N}, \mathbb{N}_0, \mathbb{Z}, \mathbb{Q}$. But there does exist sets that are properly bigger, as we show next.

Theorem 4.29 (Cantor's diagonalization argument (1891))**.** *$\mathcal{P}(\mathbb{N})$ is uncountable.*

Comment about the proof

We do a proof by *contradiction*: we assume that the statement is false, and we derive a contradiction. We will see in Chapter 2 why this is allowed.

contradiction

Proof. Let us assume that $\mathcal{P}(\mathbb{N})$ *is* countable, so that $|\mathcal{P}(\mathbb{N})| \leq |\mathbb{N}|$. By Lemma 4.23, there exists a *surjective* function $g\colon \mathbb{N} \to \mathcal{P}(\mathbb{N})$. We define the following set $T \subseteq \mathbb{N}$: for $n \in \mathbb{N}$, we put n in T only in the case that $n \notin g(n)$. In symbols, $T = \{n \in \mathbb{N} : n \notin g(n)\}$. Since $T \subseteq \mathbb{N}$, we have $T \in \mathcal{P}(\mathbb{N})$. Since g is surjective, there must exist some $m \in \mathbb{N}$ such that $g(m) = T$. Now let us consider whether m is an element of T:

- if $m \in T$, then by definition of T we must have $m \notin g(m)$, a contradiction since $g(m) = T$.
- if $m \notin T$, then again by definition of T it must be that $m \in g(m)$, again a contradiction.

In both cases we reached a contradiction, which must mean that our assumption about g being surjective cannot be true. Thus, $\mathcal{P}(\mathbb{N})$ is uncountable. □

This fact took the mathematical community by surprise: some infinite sets are in fact "bigger" than others. In fact, there is nothing special about $\mathbb{N}$ in Theorem 4.29: for every set A, it is true that $|A| < |\mathcal{P}(A)|$, i.e., $\mathcal{P}(A)$ is strictly bigger than A. So we have an infinite ladder of sizes $|\mathbb{N}| < |\mathcal{P}(\mathbb{N})| < |\mathcal{P}(\mathcal{P}(\mathbb{N}))| < \dots$. The study of such large sets is also an important topic in set theory. A big problem in the 20th century was to determine whether there exists a set A such that $|\mathbb{N}| < |A| < |\mathcal{P}(\mathbb{N})|$. Mathematicians believed that none existed, a belief that was is known as the *continuum hypothesis*. Gödel[20] proved that it is possible that no such set exist (in the sense that it would not create a contradiction in mathematics). But surprisingly, Cohen proved that it is *also* possible that such a set exists! Thus, the truth of the continuum hypothesis cannot be decided in "normal" set theory.

We conclude this chapter by considering the sizes of $|\mathbb{R}|$ and $|\mathbb{C}|$.

Theorem 4.30. $|[0,1]| = |\mathbb{R}| = |\mathcal{P}(\mathbb{N})|$.

Proof. We first prove that $|[0,1]| = |\mathbb{R}|$. Since $[0,1] \subseteq \mathbb{R}$, we have $|[0,1]| \leq |\mathbb{R}|$, so it remains to give an injective function $\mathbb{R} \to [0,1]$. For this, consider for example f defined by $f(x) = \arctan(x)/\pi + 1$.

So it remains to prove that $|[0,1]| = |\mathcal{P}(\mathbb{N})|$. Let us define an injective function $f\colon \mathcal{P}(\mathbb{N}) \to [0,1]$. Let $S \subseteq \mathbb{N}$. For every $i \in \mathbb{N}$, define s_i to be 0 if $i \notin S$, and 1 if $i \in S$. Let $f(S)$ be the number that can be written as $0.s_1s_2s_3\dots$. For example, if $S = \{1,4,5,6,8\}$, then $f(S) = 0.10011101000\dots$, and if T is the set of odd numbers, then $f(T) = 0.1010101010\dots$. This is injective, because if S and T are arbitrary sets such that $S \neq T$, then there exists some $i \in S \Delta T$ (i.e., i is in exactly one of S and T). Then $s_i \neq t_i$, so that $f(S) \neq f(T)$.

The proof that $|[0,1]| \leq |\mathcal{P}(\mathbb{N})|$ is similar and we skip the details: the trick is to write down numbers in $[0,1]$ with their *binary expansion*, and map a number $0.b_1b_2b_3\dots$ to the set $\{i \in \mathbb{N} \mid b_i = 1\}$. One must pay attention to the fact that $0.01111\dots$ (in binary

[20] A famous austrian logician, known for the *Gödel incompleteness theorem*.

representation) is the same number as 0.1 (in binary representation), i.e., $\frac{1}{2}$. It suffices to make a choice for every number of an arbitrary representation for that number. □

Theorem 4.31. $|\mathbb{R}| = |\mathbb{C}|$.

Proof. $|\mathbb{R}| \leq |\mathbb{C}|$ since $\mathbb{R} \subseteq \mathbb{C}$.

For the other inequality, remember that $\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R}\}$. To define an injective function $f\colon \mathbb{C} \to \mathbb{R}$, first consider the set C of all complex numbers $a + ib$ where a and b are in $[0, 1]$. So a and b can be written as $0.a_1a_2\ldots$ and $0.b_1b_2\ldots$, where $a_1, b_1, a_2, b_2, \cdots \in \{0, \ldots, 9\}$. If we insist that a and b do not have an infinite string of 9s in their decimal expansion, then this is unique.[21] Define $f(a + ib)$ as the real number $0.a_1b_1a_2b_2a_3b_3\ldots$. This is an injective function $f\colon C \to \mathbb{R}$. But we also have an injective function $g\colon \mathbb{C} \to C$, see the first paragraph of the proof of Theorem 4.30. Since the composition of injective functions is an injective function by Lemma 1.18, we obtain $|\mathbb{C}| \leq |\mathbb{R}|$. □

A *binary string* is a sequence of 0s and 1s. So 00011, 011000, 010101... are binary strings. The strings that never end are called *infinite*, otherwise a string is *finite*.

binary string

Theorem 4.32. *The set of infinite binary strings has the same cardinal as $\mathcal{P}(\mathbb{N})$. The set of binary strings of finite length has the same cardinal as $\mathbb{N}$.*

Proof. Let S be the set of binary strings. We give a bijection $f\colon S \to \mathcal{P}(\mathbb{N})$, which is enough by Lemma 4.28. For any string $s = s_1s_2s_3s_4\ldots$, where each $s_i \in \{0, 1\}$, define a set $f(s) = \{i \in \mathbb{N} \mid s_i = 1\}$. To prove that f is bijective, let us use Lemma 1.21 and show that f has an inverse. For an arbitrary set $A \subseteq \mathbb{N}$, define $g(A)$ as the string $a_1a_2a_3\ldots$ such that $a_i = 1$ if $i \in A$, and $a_i = 0$ otherwise. Then we have $f(g(A)) = A$ and $g(f(s)) = s$, so $g = f^{-1}$ and f is a bijection.

We leave the case of strings of finite length for later. □

4.7 Exercises

1. Show that if I and J are disjoint, then if a_i is a number for every $i \in I \cup J$, then $\sum_{i\in I} a_i + \sum_{i\in J} a_i = \sum_{i\in I\cup J} a_i$.

2. Show that $\sum_{i\in I} a$, where a is a number that does not depend on i, is equal to $a \cdot |I|$.

3. Suppose we have a collection of numbers $a_{i,j}$ parameterized by $i \in I$ and $j \in J$ (where I and J are finite sets). Show that $\sum_{i\in I} \left(\sum_{j\in J} a_{i,j}\right) = \sum_{j\in J} \left(\sum_{i\in I} a_{i,j}\right)$.

[21] Remember that $1 = 0.999\ldots$, so some numbers have several expansions, but only one that does not have an infinite string of 9s.

4. In general, show that $\sum_{i\in I} a_i + \sum_{i\in J} a_i = \sum_{i\in I\cup J} a_i + \sum_{i\in I\cap J} a_i$. If $a_i = 1$ for all $i \in I \cup J$, what do we get as a corollary?

5. Try to give a proof of Lemma 4.8 by computing, using the fact that $\binom{n}{k} = n!/k!(n-k)!$.

6. Prove that $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ holds for all $1 \le k \le n$.

7. Show that $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$ by computation (and using induction), and by a combinatorial proof.

8. Show that $\binom{2n}{2} = 2 \cdot \binom{n}{2} + n^2$ both by computation and by a combinatorial proof.

9. Compute the number of $a \in \{0, 1, \ldots, 100\}$ that are divisible by 2 *or* 3.

10. A *derangement* of $\{1, \ldots, n\}$ is a bijective function $f\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $f(i) \neq i$ for all $i \in \{1, \ldots, n\}$. Define A_i to be the set of bijective functions $f\colon \{1, \ldots, n\} \to \{1, \ldots, n\}$ such that $f(i) = i$. Determine the number of possible derangements by applying Theorem 4.11.

11. Show that $S(n, k) \ge k^{n-k}$ holds for all $1 \le k \le n$. Thus, even for fixed k, $S(n, k)$ grows *exponentially* in n.

12. Show that $S(n+1, k+1) = \sum_{i=0}^{n} \binom{n}{i} S(i, k)$.

13. Show that $\sum_{k=0}^{m} \binom{n}{k}\binom{n-k}{m-k} = 2^m \binom{n}{m}$ holds for all $0 \le m \le n$.

14. (*) Prove that Ramsey's theorem (Theorem 4.14) holds for an arbitrary number of colors, and not just 2 (use induction on the number of colors).

15. (*) Prove that Ramsey's theorem (Theorem 4.14) also holds for functions $f\colon \mathcal{P}_k(A) \to \{\text{red}, \text{blue}\}$, for any $k \ge 1$ (use induction on k).

16. (**) Prove the following infinite version of Ramsey's theorem: for every $f\colon \mathcal{P}_2(\mathbb{N}) \to \{\text{red}, \text{blue}\}$, there exists an infinite subset M of $\mathbb{N}$ such that for all $x, y \in M$ with $x \neq y$, we have $f(\{x, y\}) = \text{red}$, or for all $x, y \in M$ with $x \neq y$, we have $f(\{x, y\}) = \text{blue}$.
Hint: it is also a proof by induction.

17. Let F be the set of functions $\mathbb{N} \to \mathbb{N}$. Show that the relation $f \preceq g$ on F defined by $f = O(g)$ is a quasiorder, but not a partial order. Is the quasiorder linear?

18. Show that the relation $f \sim g$ on functions $\mathbb{N} \to \mathbb{N}$ defined in Notation 4.18 is an equivalence relation.

19. Determine whether $f = O(g)$ or $g = O(f)$, for functions f, g from the usual functions $n^k, \log(n), \log^k(n), n\log(n), e^n, \ldots$.

20. Show that if $|A| \leq |B|$ and $|B| \leq |C|$, then $|A| \leq |C|$ (make sure to use the definitions!).

21. Prove that if $A \subseteq B$, then $|A| \leq |B|$.

22. Prove that $|[0,1]| = |(0,1)|$.

23. Suppose that A and B are countable. Show that $A \cup B$ and $A \times B$ are countable.

24. Prove that for any $a < b$ and $c < d$, we have $|[a,b]| = |[c,d]|$.

25. (**) Let S be the set of all sequences $(s_1, s_2, s_3, \dots)$, where each s_i is in $\mathbb{N}_0$. Give an injection $S \to \mathcal{P}(\mathbb{N})$. Use this to show that $|S| = |\mathcal{P}(\mathbb{N})|$.

26. (**) Let $A_1, A_2, A_3, \dots$ be countable sets (i.e., for every $i \in \mathbb{N}$, we have a countable set A_i). Show that the union $\bigcup_{i \geq 1} A_i$ (the set containing every a such that for some $i \in \mathbb{N}$, $a \in A_i$) of these sets is countable.

5 Elementary Number Theory

Things to remember / to know

- Euclid's algorithm (Lemma 5.5) and how to apply it to compute $\gcd(a, b)$, as well as numbers u, v such that $ua + vb = \gcd(a, b)$ (Corollary 5.6)
- How to prove the existence of infinitely many prime numbers (Lemma 5.8)
- The fundamental theorem of arithmetic (Theorem 5.13)
- How to decompose a number in a given base (Theorem 5.17)
- How to add, multiply, subtract, invert numbers modulo d (Theorem 5.21, proposition 5.24, and remark 5.25)
- How to apply the Chinese remainder theorem (Theorem 5.27)
- Euler's totient function and Fermat's theorem (Definition 5.28, lemma 5.30, and theorem 5.32)

In this chapter, we study some elementary properties of the natural and integer numbers.

Definition 5.1. Let $a, b \in \mathbb{Z}$. We say that a *divides* b if the remainder in the division of b by a is 0. We also call a a *divisor* of b, or b a *multiple* of a.

divisor

As we already saw, the set $\mathbb{N}$ can be ordered by the divisibility relation: $a|b$ if, and only if, b is a multiple of a. The structure of this poset is what is studied in number theory (see Figure 24).

5.1 Divisibility

Theorem 5.2. *Let $a, b \in \mathbb{Z}$, where $a \neq 0$. There exist integers $q, r \in \mathbb{Z}$ such that the following hold:*

- $b = q \cdot a + r$,
- $r \in \{0, \ldots, |a| - 1\}$.

Moreover, q and r are unique, q is called the quotient *and r is called the* remainder.

quotient, remainder

Lemma 5.3. *Let $a, b, d \in \mathbb{Z}$. If d is a divisor of a and b, then for every $u, v \in \mathbb{Z}$, it is a divisor of $ua + vb$.*

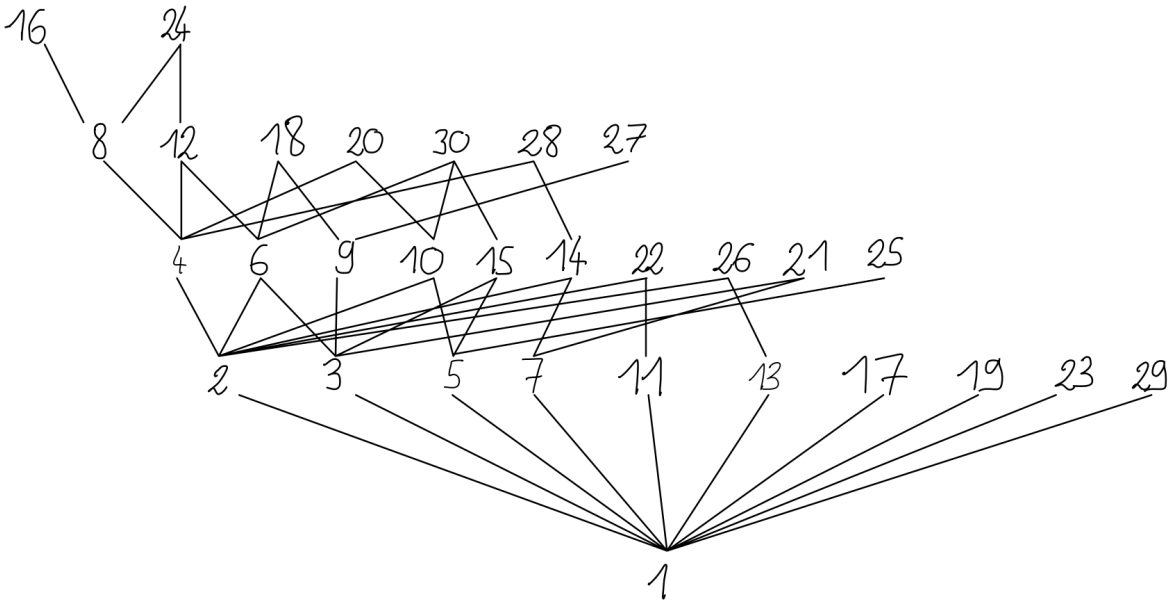


Figure 24: Hasse diagram of the divisibility poset when restricted to the integers in $\{1, \ldots, 30\}$.

Proof. Write $a = qd$ and $b = q'd$. Then $ua + vb = uqd + vq'd = (uq + vq') \cdot d$, so $ua + vb$ is divisible by d. □

5.2 Euclid's Algorithm

Definition 5.4. Two numbers a, b are *coprime* if they do not have a common divisor except for 1 and -1.

coprime

For example, $15, 16$ are coprime since the numbers dividing 15 are $\{\pm 1, \pm 3, \pm 5, \pm 15\}$ and the divisors of 16 are $\{\pm 1, \pm 2, \pm 4, \pm 8, \pm 16\}$, and those divisors have no element in common except for 1 and -1.

When two numbers a, b are *not* coprime, they have a common divisor that is greater than 1. It is a priori not clear that there is a greatest common divisor of a and b in the divisibility poset, but this is what we show in the next result.

Lemma 5.5. *Any two $a, b \in \mathbb{N}$ have a greatest lower bound in $\mathbb{N}$, called $\gcd(a, b)$.*

Proof. a, b definitely have a lower bound in the poset, since 1 divides every number.

The hard part is to prove that there is a *greatest* lower bound. Start by assuming that $b > a$, which we can do by swapping a and b if necessary. If $a = 0$, then b divides both a and b, and there cannot be a divisor of b greater than b, so $\gcd(b, 0) = b$. Similarly, if

$a = 1$, then the only possible divisor of a is 1, so $\gcd(b, 1) = 1$.

If $a > 1$, we follow the following algorithm discovered by Euclid[22] to compute $\gcd(a, b)$ and prove its existence (Algorithm 5.1).

Algorithm 5.1: Euclid's algorithm

Data: $a, b \in \mathbb{N}$ with $b > a > 0$

Result: A decreasing list $L = \langle r_0, r_1, \ldots, r_k \rangle$ of numbers where $r_0 = b, r_1 = a, r_k = \gcd(a, b)$

```
L ← ⟨b, a⟩;
continue ← True;
while continue do
    m, n ← last two elements of L (n is the last, m is the second-to-last);
    r ← remainder of the division of m by n;
    if r ≠ 0 then
        Append r to the end of L;
    else
        continue ← False;
    end
end
return L
```

Note that at every step, we have $0 \leq r < n$, so this process must stop after finitely many rounds (i.e., the *while* loop cannot loop forever). We claim that the last element of the output list L is $\gcd(a, b)$. Let $r_0, r_1, \ldots, r_k$ be the elements of L

For every $i \in \{0, \ldots, k\}$, if d divides a and b, then d divides r_i: it is true for $k \in \{0, 1\}$ (since in that case $r_k \in \{a, b\}$). Assume now that it is true up to some $i > 1$, so that d divides r_{i-1} and r_{i-2}. Let q be the quotient of the division of r_{i-2} by r_{i-1}, i.e., $r_{i-2} = qr_{i-1} + r_i$. Then d divides $r_{i-2} - qr_{i-1}$ by Lemma 5.3, which is r_i. So the claim is true for all $i \in \{0, \ldots, k\}$.

In the other direction, we show that the last element r_k of the list divides a and b. Indeed, r_k divides r_{k-1} (since otherwise the remainder of the division of r_{k-1} by r_k would be nonzero, and the algorithm would have continued at least one more loop). Moreover, since $r_{k-2} = qr_{k-1} + r_k$ for some q, we get by Lemma 5.3 that r_k divides r_{k-2}. By induction, one can prove that r_k then divides every r_{k-j}, for $j \in \{0, \ldots, k\}$, so that for $j = k - 1$ and $j = k$, we get that r_{k-1} divides $r_1 = b$ and $r_0 = a$.

Therefore, r_k divides a, b, and every divisor d of a, b is also a divisor of r_k. In conclusion, r_k is the greatest integer dividing a and b. □

Algorithm 5.1 is in fact efficient. To show this, we must estimate the size k of the list L that is returned. Let $i \in \{1, \ldots, k\}$. Note that if $r_{i-1} < r_{i-2}/2$, then since $r_i \in \{0, \ldots, r_{i-1} - 1\}$, we have $r_i < r_{i-2}/2$. If $r_{i-1} > r_{i-2}/2$, then during this iteration the quotient of the division of r_{i-2} by r_{i-1} must be 1, so $r_i = r_{i-2} - r_{i-1} < r_{i-2}/2$.

[22] A Greek mathematician from around the third century BC.

Therefore, every two iterations of the loop, the size of the current remainder r_i is halved. This means that the loop can only be performed at most $2\log_2(a)$ times. For $a < 10^{10}$, this is only about 60 iterations and would run instantly even on a smartphone.

Corollary 5.6. *For any two $a, b \in \mathbb{N}$, there exist $u, v \in \mathbb{Z}$ such that $\gcd(a, b) = ua + vb$.*

Proof. Let $r_0, \ldots, r_k$ be the elements of the list L computed in Algorithm 5.1. Note that in Algorithm 5.1, it is true that for every $i \in \{0, \ldots, k\}$, there exist $u, v \in \mathbb{Z}$ such that $r_i = ua + vb$. This is proved by induction: it is true for $i \in \{0, 1\}$ by taking $(u, v) = (1, 0)$ or $(u, v) = (0, 1)$, and then for $i > 1$ remember that $r_i = r_{i-2} - qr_{i-1}$, where q is the quotient of the division of r_{i-2} by r_{i-1}. So if u, v, u', v' are such that $r_{i-2} = ua + vb$ and $r_{i-1} = u'a + v'b$, then $r_i = ua + vb - qu'a - qv'b = (u - qu')a + (v - qv')b$. So the property remains true for r_i.

Since $\gcd(a, b) = r_k$ for the last iteration of the algorithm, the proof is finished. □

5.3 Prime numbers

Definition 5.7. A number $p \in \mathbb{Z}$ is called *prime* if $p \notin \{1, -1\}$ and the only numbers dividing p are ± 1 and $\pm p$.

prime

Note that p is prime if, and only if, $-p$ is prime. If $p \in \mathbb{N}$ is prime, then it is on the "second level" in the Hasse diagram of the divisibility poset.

Lemma 5.8. *Every $n \in \mathbb{Z}$ different from $1, -1$ is divisible by a prime number.*

Proof. If $n = 0$, then it is divisible by *all* prime numbers, so we can assume that $n \neq 0$.

Assume that $n > 0$ (if $n < 0$, do the proof for $-n$ instead). We prove the claim by induction on $n \geq 2$.

The base case is $n = 2$, which holds because 2 is itself prime. Suppose now that $n > 2$. If n has no divisor apart from $\pm 1, \pm n$, then n is itself prime so we are done. Otherwise, there exists $m \in \mathbb{N}$ with $m \notin \{n, 1\}$ such that m divides n. Since $m > 1$, we have $n/m < n$, and since $m < n$, we have $n/m > 1$. Thus we can apply the induction hypothesis, which gives that there exists a prime number p that divides n/m, i.e., $n/m = q \cdot p$ for some $q \in \mathbb{N}$. Then $n = qm \cdot p$, so that p divides n and we are done. □

Theorem 5.9. *The set of prime numbers is infinite.*

Proof. Suppose that we have a collection of prime numbers $p_1, \ldots, p_k$, that we assume to be positive. We show how to produce a new prime number p' that is not in $\{p_1, \ldots, p_k\}$. Consider the number $n = 1 + p_1 \times p_2 \times \cdots \times p_k = 1 + \prod_{i=1}^{k} p_i$. Then n is not divisible

by any p_i: indeed, we have $n = (p_1 \cdots p_{i-1} \cdot p_{i+1} \cdots p_k) \cdot p_i + 1$, and $1 \in \{0, \ldots, |n| - 1\}$, so that by uniqueness of quotients and remainder (Theorem 5.2) the division of n by p_i has quotient $p_1 \cdots p_{i-1} \cdot p_{i+1} \cdots p_k$ and remainder 1.

By Lemma 5.8, there exists a prime number p' that divides n, so this number cannot be any of $p_1, \ldots, p_k$. □

Remark 5.10. The proof is constructive: given any set S of primes, it gives a recipe to produce another prime not in S. However, this recipe is not efficient, since we currently do not have a way to find a prime divisor of the number n in the proof. The RSA cryptosystem, one of the oldest encryption schemes in the computer age and still in use to this day, relies on this fact. Due to the advent of quantum computing, it might be the case that the RSA system has to be abandoned for security reasons.

Lemma 5.11. *Suppose that p is a prime number, and that $a, b \in \mathbb{N}$. If p divides ab, then it divides a or it divides b.*

Proof. Suppose that p divides ab and that it does not divide a. By Corollary 5.6, we can find $u, v \in \mathbb{Z}$ such that $up + va = 1$, since $\gcd(a, p) = 1$. Thus, $pub + vab = b$. Since p and ab are divisible by p, we have by Lemma 5.3 that $pub + vab$ is divisible by p. It follows that b is divisible by p. □

Remark 5.12. Note how important the assumption about p being prime is important! Namely 4 divides 2×2, but it does not divide 2. In fact, in a more general algebraic context, a "prime" element is defined as a thing that satisfies Lemma 5.11, and element satisfying the definition Definition 5.7 are called *irreducible*.

Mathematicians are *obsessed* by prime numbers. The following statement is one of the reasons why. It says that the set of numbers $\mathbb{N}$ can be generated by prime numbers using multiplication only, or equivalently that every number can be expressed as a product of prime numbers.

Theorem 5.13. *Let $n \in \mathbb{N}$ be such that $n \geq 2$. There exist prime numbers $p_1, \ldots, p_k \in \mathbb{N}$ and numbers $e_1, \ldots, e_k \in \mathbb{N}$ such that $n = p_1^{e_1} \ldots p_k^{e_k} = \prod_{i=1}^{k} p_i^{e_i}$. Moreover, those numbers are uniquely determined, up to permutation. The numbers $p_1, \ldots, p_k$ are called the* prime factors *of n.*

prime factor

Proof. We do the proof of existence by induction on $n \geq 2$.

The base case $n = 2$ is trivial, since 2 is prime and therefore we can take $p_1 = 2, e_1 = 1$. Assume that $n > 2$. If n is prime, we take $p_1 = n$ and $e_1 = 1$. Otherwise,

by Lemma 5.8 there exists a prime divisor p of n. By induction hypothesis, n/p can be written as $\prod_{i=1}^{k} p_i^{e_i}$ for some primes $p_1, \dots, p_k$ and $e_1, \dots, e_k \in \mathbb{N}$. If $p \notin \{p_1, \dots, p_k\}$, then we let $p_{k+1} = p$ and $e_{k+1} = 1$, and we have $n = n/p \cdot p = (\prod_{i=1}^{k} p_i^{e_i}) \cdot p = \prod_{i=1}^{k+1} p_i^{e_i}$. Otherwise, let i be such that $p_i = p$. Then we let $e'_i = e_i + 1$ and $e'_j = e_j$ for $j \neq i$, and we have $n = \prod_{i=1}^{k} p_i^{e'_i}$.

We now prove the uniqueness of the decomposition. Suppose that the same number n can be written as $p_1^{e_1} \dots p_k^{e_k}$ and as $q_1^{f_1} \dots q_\ell^{f_\ell}$ for prime numbers $p_1, \dots, p_k, q_1, \dots, q_\ell$ and numbers $e_1, \dots, e_k, f_1, \dots, f_\ell \in \mathbb{N}$. Then for each $i \in \{1, \dots, k\}$, p_i divides $q_1^{f_1} \dots q_\ell^{f_\ell}$. By Lemma 5.11, there must exist $j \in \{1, \dots, \ell\}$ such that p_i divides q_j. Since q_j is prime and $p_i \neq 1$, we have that $q_j = p_i$. So every p_i must be one of the q_j, and we prove similarly that for every $j \in \{1, \dots, \ell\}$ there exists $i \in \{1, \dots, k\}$ such that $q_j = p_i$. Therefore, the numbers $p_1, \dots, p_k$ and $q_1, \dots, q_\ell$ are equal up to permutation and we can assume without loss of generality that $p_1 = q_1, \dots, p_k = q_k$.

We finally conclude that the exponents must be identical. Indeed, suppose that $e_i \neq f_i$, and up to symmetry we can suppose $e_i < f_i$. Then $n/p_i^{e_i}$ can be written as a product of $\prod_{j \neq i} p_j^{e_j}$ and $p_i^{f_i - e_i} \cdot \prod_{j \neq i} p_j^{f_j}$, where p_i does not appear in the first product, but it does appear in the right, a contradiction to the previous paragraph. □

Remark 5.14. Note that negative numbers also admit the same kind of decomposition: if $n \in \mathbb{N}$ has decomposition $\prod_{i=1}^{k} p_i^{e_i}$, then $-n$ has the decomposition $(-1) \cdot \prod_{i=1}^{k} p_i^{e_i}$. One could also say that 1 has a prime decomposition: the decomposition into 0 factors. So 0 is the only integer that does not have a prime decomposition.

It is known that the number of prime numbers less than or equal to n, denoted by $\pi(n)$, is such that $\pi(n) \sim \frac{n}{\ln(n)}$.[23] This is the *prime number theorem*, whose proof is outside of the scope of the course. This means that by choosing randomly a number in $\{1, \dots, n\}$, one has a $1/\ln(n)$ chance of finding a prime. To this day, there is no efficient way of finding large prime numbers. Currently, the largest known prime number is $2^{82,589,933} - 1$ and was found essentially by exhaustive search using a computer.

prime number theorem

5.4 How to write numbers

Let us make a quick detour and mention that while in every day life we write numbers in what is called *base 10*, computers handle numbers in a different way.

Definition 5.15. Let $n \in \mathbb{N}$, and let $b \in \mathbb{N}$ with $b \geq 2$ be a number called the

[23] The symbol $\sim$ refers here to the equivalence relation on functions given in Notation 4.18.

base. A *decomposition of n in base-b* is an expression of the form

$$n = \sum_{i=0}^{k} n_i b^i$$

for some $k \geq 0$ and numbers $n_0, \ldots, n_k \in \{0, \ldots, b-1\}$, with $n_k \neq 0$.

decomposition of a number in a base

Example 5.16. Our usual notation for numbers immediately gives us base-10 decompositions of numbers: $314159 = 3{\cdot}10^5 + 1{\cdot}10^4 + 4{\cdot}10^3 + 1{\cdot}10^2 + 5{\cdot}10^1 + 9{\cdot}10^0$. To emphasize that we write a number in a given base, we sometimes write 314159_{10}. For another example, we have $13_{10} = 1 \cdot 10^1 + 3 \cdot 10^0$, but 13 also has the base-2 decomposition $1 \cdot 2^3 + 1 \cdot 2^2 + 1 \cdot 2^0 = 1101_2$. Digital computers store numbers (and everything, actually) in base 2. In that case, the numbers n_i are called the *bits* of the decomposition.

bit

Theorem 5.17. *Every number $n \in \mathbb{N}$ has a unique decomposition in every base $b \geq 2$.*

Proof. We first prove the existence of a decomposition, and we do so by induction on $n \geq 1$.

The base case is $n = 1$: in this case, we can simply take $k = 0, n_0 = 1$. Then $n_0 \cdot b^0 = 1 \cdot 1 = 1$. Let now $n > 1$, and suppose now that every number less than n has a decomposition in base b. Let $k \geq 0$ be the largest natural number such that $b^k \leq n$, and let q, r be obtained by dividing n by b^k (Theorem 5.2). That is, we have $n = qb^k + r$, where $r \in \{0, \ldots, b^k - 1\}$. Note that since k was taken maximal, we must have $q < b$, otherwise we would have $n \geq q \cdot b^k \geq b^{k+1}$. We let n_k be q. By the induction hypothesis, the number r must have a base-b decomposition $\sum_{i=0}^{\ell} r_i b^i$, where $r_0, \ldots, r_\ell \in \{0, \ldots, b-1\}$ and $r_\ell \neq 0$. Since $r < b^k$, we must have $\ell < k$. Let $n_i = r_i$ for all $i \in \{0, \ldots, \ell\}$, and $n_i = 0$ for $i \in \{\ell+1, \ldots, k-1\}$. Then putting everything together we get

$$n = n_k b^k + r = n_k b^k + \sum_{i=0}^{\ell} r_i b^i = \sum_{i=0}^{k} n_i b^i,$$

which is a decomposition of n in base b.

Let us now prove the uniqueness, again by induction on $n \geq 1$. The base case of $n = 1$ is clear: if $1 = \sum_{i=0}^{k} n_i b^i$, it must be that $i = 0$ and $n_0 = 1$.

Suppose now that we have

$$\sum_{i=0}^{k} n_i b^i = n = \sum_{i=0}^{\ell} n'_i b^i$$

for some $n > 1$, with $n_0, \ldots, n_k, n'_0, \ldots, n'_\ell \in \{0, \ldots, b-1\}$ and $n_k \neq 0$, $n'_\ell \neq 0$. We can also assume without loss of generality that $k \geq \ell$ (up to exchanging the roles of the decompositions). Then we must have $k = \ell$: otherwise, if $k > \ell$, then the first decomposition is a number that is at least b^k, while the decomposition on the right gives $\sum_{i=0}^{\ell} n'_i b^i \leq (b-1)\sum_{i=0}^{\ell} b^i = (b-1) \cdot \frac{b^{\ell+1}-1}{b-1} = b^{\ell+1} - 1 < b^k$.

Again, without loss of generality, we can assume that $n_k \geq n'_k$. Then we must have $n_k = n'_k$: otherwise, $n - n'_k b^k$ would have the two decompositions

$$\sum_{i=0}^{k-1} n_i b^i + (n_k - n'_k)b^k = n - n'_k b^k = \sum_{i=0}^{k-1} n'_i b^i$$

in contradiction to the previous paragraph, since the left-hand side has exponents going up to k but on the right-hand side only up to $k-1$.

Finally, by induction hypothesis, we have that the two decompositions

$$\sum_{i=0}^{k-1} n_i b^i = n - n'_k b^k = \sum_{i=0}^{k-1} n'_i b^i$$

of $n - n_k b^k$ are identical, i.e., $n_i = n'_i$ for all i. Thus, n only has one decomposition in base b, which concludes the proof. □

Remark that if n has decomposition $\sum_{i=0}^{k} n_i b^i$, then we can write n as $\left(\sum_{i=1}^{k} n_i b^{i-1}\right) b + n_0$, so that the division of n by b has quotient $\left(\sum_{i=1}^{k} n_i b^{i-1}\right)$ and remainder n_0.

5.5 Modular arithmetic

One of the principles that we saw in Chapters 3.4.1 and 3.4.2, is that in order to understand a set that has some structure (here, the set $\mathbb{Z}$ of integers with the divisibility relation), it can be helpful to define an equivalence relation on $\mathbb{Z}$ and study the factor set. We follow this approach here.

Definition 5.18. Fix an integer $d \in \mathbb{N}$. Let $\equiv_d$ be the relation on $\mathbb{Z} \times \mathbb{Z}$ containing the pairs (a, b) if a and b have the same remainder in the division by d. We also write $a \equiv b \bmod d$ or $a \equiv_d b$ if $(a, b) \in \equiv_d$.

Lemma 5.19. *We have $a \equiv_d b$ if, and only if, d divides $a - b$.*

Proof. Write $a = qd + r$ and $b = q'd + r'$ for $r, r' \in \{0, \ldots, d-1\}$ and $q, q' \in \mathbb{Z}$.

Suppose that $a \equiv_d b$, i.e., we have $r = r'$. Then $a - b = (q - q')d$, so d divides $a - b$.

Conversely, suppose that d divides $a - b$. Suppose that $r > r'$. Then $a - b = (q - q')d + (r - r')$ must be divisible by d. By Theorem 5.2, the remainder in the division by d is unique. Since $r - r' \in \{0, \ldots, d-1\}$, we get $r - r' = 0$, and $r = r'$. The case

where $r' > r$ is similar: if $a - b$ is divisible by d, then $b - a$ is divisible by d as well, so that $r' - r \in \{0, \dots, d-1\}$ is the rest of the division of $b - a$ by d, which must therefore be 0. □

Lemma 5.20. *The relation $\equiv_d$ is an equivalence relation.*

Proof. We need to show reflexivity, symmetry, and transitivity. Definition 5.18 clearly defines a reflexive and symmetric relation.

Suppose that $a, b, c \in \mathbb{Z}$ are such that $a \equiv_d b$ and $b \equiv_d c$. By Lemma 5.19, this means that d divides $a - b$ and d divides $b - c$. By Lemma 5.3 applied with $u = v = 1$, we get that d divides $(a - b) + (b - c) = a - c$. By Lemma 5.19, we finally obtain that $a \equiv_d c$, so that $\equiv_d$ is transitive. □

By Definition 5.18, the equivalence classes of $\equiv_d$ are characterized by the remainder of the elements inside it. We already saw in Example 3.10 the example of the case where $d = 2$, which has exactly two equivalence classes (even numbers, and odd numbers). For $d = 3$, the equivalence classes are $\{\dots, -9, -6, -3, 0, 3, 6, 9, \dots\}$ (the multiples of 3, having a remainder of 0 in the division by 3), $\{\dots, -5, -2, 1, 4, 7, \dots\}$ (numbers of the form $3n + 1$ with $n \in \mathbb{Z}$, having a remainder of 1 in the division by 3), and $\{\dots, -4, -1, 2, 5, 8, 11, \dots\}$ (numbers of the form $3n + 2$ with $n \in \mathbb{Z}$, having a remainder of 2 in the division by 3).

Thus, $\mathbb{Z}/\equiv_d$ has exactly d elements. We also note that all the equivalence classes of $\equiv_d$ are quite similar to each other, they are all of the form $\{\dots, a - 2d, a - d, a, a + d, a + 2d, a + 3d, \dots\}$ for some $a \in \mathbb{Z}$. From this point on, we adopt the notation $\mathbb{Z}/d\mathbb{Z}$ instead of $\mathbb{Z}/\equiv_d$.

Perhaps surprisingly, there is a natural way to add and multiply the equivalence classes of the relation $\equiv_d$. Thus, there is a notion of *arithmetic* in the set $\mathbb{Z}/d\mathbb{Z}$, just like the "classical" arithmetic in the set $\mathbb{Z}$. This arithmetic is called *modular arithmetic*.

modular arithmetic

In the following, recall that $[a]$ denotes the equivalence class of a (for the relation $\equiv_d$).

Theorem 5.21. *Let $a, b \in \mathbb{Z}$ and $d \in \mathbb{N}$. The operations $[a] + [b] := [a + b]$ and $[a] \times [b] := [a \times b]$ defined on $\mathbb{Z}/d\mathbb{Z}$ are well defined.*

Example 5.22. Before seeing the proof, let us see what this means in the case of $d = 3$. As we saw, we have three equivalence classes for $\equiv_3$, which are $[0], [1], [2]$. The addition on these classes is defined by the rules

$$\begin{array}{lll} [0] + [0] = [0] & [1] + [0] = [1] & [2] + [0] = [2] \\ [0] + [1] = [1] & [1] + [1] = [2] & [2] + [1] = [0] \\ [0] + [2] = [2] & [1] + [2] = [0] & [2] + [2] = [1] \end{array}$$

Indeed, since $[0] = [3]$, we have that $[1] + [2]$, which is supposed to be $[1 + 2]$ by

definition, is also equal to $[0]$.
Similarly, multiplication becomes

$$\begin{array}{lll} [0]\times[0]=[0] & [1]\times[0]=[0] & [2]\times[0]=[0] \\ [0]\times[1]=[0] & [1]\times[1]=[1] & [2]\times[1]=[2] \\ [0]\times[2]=[0] & [1]\times[2]=[2] & [2]\times[2]=[1] \end{array}$$

What Theorem 5.21 is saying is that one could also compute with $[6], [-2], [5]$ instead of $[0], [1], [2]$, and obtain the same results: $[-2]+[5]$ gives for example the same result as $[1]+[2]$, because $[-2]=[1]$ and $[5]=[2]$.

Proof. Remember from Theorem 3.22 that when defining a function or operation whose arguments are equivalence classes $[a]$ and $[b]$, one needs to check that the result does not depend on the choice of a, b. Therefore, one needs to prove that if $[a]=[a']$ and $[b]=[b']$, then $[a+b]=[a'+b']$ and $[a\times b]=[a'\times b']$.

By Lemma 5.19, we have that $a-a'$ is divisible by d and $b-b'$ is divisible by d. We have $(a+b)-(a'+b')=(a-a')+(b-b')$, and since both $a-a'$ and $b-b'$ are divisible by d, we have from Lemma 5.3 that $(a+b)-(a'+b')$ is also divisible by d. Thus, $(a+b)\equiv_d(a'+b')$ by Lemma 5.19, so $[a+b]=[a'+b']$.

Concerning multiplication,

$$\begin{aligned} ab-a'b' &= ab+(-a'b+a'b)-a'b' \\ &= (ab-a'b)+(a'b-a'b') \\ &= (a-a')b+a'(b-b') \end{aligned}$$

and by Lemma 5.3 (applied with $u=b$ and $v=a'$), we obtain that $ab-a'b'$ is divisible by d. Therefore, $[a\times b]=[a'\times b']$. □

On top of addition and multiplication, one can also subtract elements in $\mathbb{Z}/d\mathbb{Z}$: if $[a], [b]\in\mathbb{Z}/d\mathbb{Z}$, then $[a]-[b]$ is defined to be $[a-b]$. The proof that this is well defined is the same as for Theorem 5.21.

We will see in our last chapter that $\mathbb{Z}/d\mathbb{Z}$ is an example of a *ring*, an algebraic structure generalizing some notions of arithmetic that we are used to for $\mathbb{Z}$ or $\mathbb{Q}$.

Even if this was not the original intention, we have just discovered an object of great practical applications in computer science!

First, we already mentioned in the previous section that computers store integers in their base-2 decomposition. When programming in low level languages, such decompositions can often have at most 32 bits. On embedded systems (chips with limited memory), this can go as low as 8 bits. This means that the largest integer that can be stored[24] is $2^8=256$. When adding two numbers n, m such that $n+m>2^8$, then all the bits in the decomposition of $n+m$ with exponent greater than 8 are simply ignore, so that for example $100+200=44$. This is called an *integer overflow*, and is also the source

[24] Without additional effort (i.e., without a specific implementation of larger integers).

of many software vulnerabilities. What is going on here is that we are in fact performing addition in $\mathbb{Z}/256\mathbb{Z}$. To analyse such a program mathematically, it then makes sense to use modular arithmetic.

A second application is related to the above one, but this time using this "integer overflow" as a feature instead of a bug. Suppose we want to know if $2^{10^{20}}$ is divisible by 3. One cannot compute $2^{10^{20}}$ or even store this on a hard disk: one would need about a *billion* hard disk drives of 1 To each. But in the end, we only want to know what $[2^{10^{20}}]$ is in $\mathbb{Z}/3\mathbb{Z}$, and by Theorem 5.21, this is the same as $[2^{10^{20}-1}] \times [2]$, which is the same as $[2^{10^{20}-2}] \times [2] \times [2] = [2^{10^{20}-2}] \times [4] = [2^{10^{20}-2}]$, since $[4] = [1]$ in $\mathbb{Z}/3\mathbb{Z}$. By keeping this reasoning, we get that $[2^{10^{20}}] = [2^0] = [1]$. So not only this number is not divisible by 3, but we know that it has a remainder of 1 in the division by 3. More generally, one sees that $[2^n] = [1]$ is n is even, and $[2^n] = [2]$ if n is odd.

Yet another application is in cryptography. The RSA cryptosystem that we already mentioned is based on computations in $\mathbb{Z}/N\mathbb{Z}$ for some large integers N chosen in a specific way.

Even more surprising is that in particular cases, one can do *division* in $\mathbb{Z}/d\mathbb{Z}$. First, let us think about what it means to divide: b/a should be a number such that $(b/a) \times a = b$, and $1/a$ is usually called the inverse of a.

Definition 5.23. Let $d \geq 2$ and let $[a] \in \mathbb{Z}/d\mathbb{Z}$. We say that $[b] \in \mathbb{Z}/d\mathbb{Z}$ is an *inverse* of $[a]$ if $[ab] = [1]$. We also say that b is an *inverse of a modulo d.*

inverse modulo an integer

For example, for $d = 4$, the inverse of 3 modulo 4 is 3 (because $[3 \times 3] = [9] = [1]$), but 2 does not have an inverse modulo 4: $[2 \times 1] = [2], [2 \times 2] = [0], [2 \times 3] = [2]$. However, 2 has an inverse modulo 5: $[2 \times 3] = [1]$, so 3 is the inverse of 2 modulo 5.

Proposition 5.24. *Let $d \geq 2$ and $[a] \in \mathbb{Z}/d\mathbb{Z}$. Then $[a]$ has an inverse if, and only if, a is coprime with d. In particular, if d is prime, then* every *a that is not divisible by d has an inverse modulo d.*

Proof. Suppose that a is coprime with d. By Corollary 5.6, there exist $u, v \in \mathbb{Z}$ such that $ua + vd = 1$. Therefore, $[ua + vd] = [1]$, and $[ua] + [vd] = [ua] + [0] = [ua]$. Thus, the inverse of $[a]$ is $[u]$.

Conversely, suppose that $[a]$ has the inverse $[b]$. Then $[ba] = [1]$, so the division of ba by d has a quotient v and remainder 1. It follows that $ba = vd + 1$, and by rearranging $ba - vd = 1$. It follows that a and d are coprime: any common factor of a, d is a factor of $ba - vd$ by Lemma 5.3, so the only common factor is ± 1.

The last sentence of the statement follows from the fact that if d is prime, then every a that is not divisible by d is coprime with d. □

Remark 5.25. The previous proof also gives us a way to compute an inverse modulo d: start with a, use the Euclidean algorithm to compute the coefficients u, v as in Corollary 5.6 such that $ua + vd = 1$. Then u is an inverse of a modulo d.

5.6 The Chinese Remainder Theorem

We have discussed so far a single relation $\equiv_m$ for a fixed m. It is natural to look at how different relations $\equiv_m, \equiv_n$ relate with respect to each other. Many things can be said, but an important result for us is the following. Here, we consider simple system of "equations" of the form

$$\begin{aligned} x &\equiv a \bmod m \\ x &\equiv b \bmod n \end{aligned} \tag{†}$$

where x is a variable representing an integer. In particular we would like to understand if such systems have solutions, and also how to find them.

Lemma 5.26. *If x is a valid solution to (†), then $x + k \cdot mn$ is a valid solution as well, for any $k \in \mathbb{Z}$.*

Proof. Simply follows from the fact that $x + k \cdot mn \equiv x \bmod m$ and $x + k \cdot mn \equiv x \bmod n$. □

So such a system either has no solution, or it has infinitely many. What the so-called Chinese remainder theorem says is that if m and n are coprime, there exists a solution and it is essentially unique.

Theorem 5.27 (Chinese remainder theorem)**.** *Let $m, n \geq 2$ be integers that are coprime, and let $a, b \in \mathbb{N}$. Let $S_{a,b} \subseteq \mathbb{Z}$ be the set of all solutions to (†). Then there exists $r \in \{0, \dots, mn - 1\}$ such that for all $x \in S$, we have $x \equiv r \bmod mn$.*

Proof. Since m and n are coprime, we know by Corollary 5.6 that there exist some integers u, v such that $u \cdot m + v \cdot n = 1$. Let c be defined as $bum + avn$. Then $c \equiv bum + avn \equiv avn \equiv a \bmod m$ and $c \equiv bum + avn \equiv bum \equiv b \bmod n$, so that $c \in S$.

Let r be the remainder of c in the division by mn, i.e., we have $c = q(mn) + r$ with $r \in \{0, \dots, mn - 1\}$. By Lemma 5.26, we have that $r \in S$ and in fact all integers of the form $kmn + r$ are in S. We show that the other inclusion holds as well. For this, we use the following trick, which we explain in words first and prove formally below. We know that for *every* $a \in \{0, \dots, m - 1\}$ and *every* $b \in \{0, \dots, n - 1\}$, the system has a non-empty set of solutions $S_{a,b}$. Suppose that for some a, b, we would even obtain multiple solutions in $\{0, \dots, mn - 1\}$. Since there are mn possible choices for a, b, there

would be (using double counting) some $r \in \{0, \dots, mn-1\}$ that ends up being a solution for several pairs (a, b), which cannot be the case.

Here is the formal argument. Consider the relation $R \subseteq (\{0, \dots, m-1\} \times \{0, \dots, n-1\}) \times \{0, \dots, mn-1\}$ containing all elements $((a, b), r)$ such that $r \in S_{a,b}$. We know that each pair (a, b) has degree $\deg(a, b)$ at least 1 (this is what we proved above). Note that every $r \in \{0, \dots, mn-1\}$ has degree $\deg(r)$ *at most* 1. Indeed, for a fixed r, there is a single a such that $r \equiv a \bmod m$ and a single b such that $r \equiv b \bmod n$.

By applying Proposition 4.9, we get

$$\begin{aligned} mn &\leq \sum_{a=0}^{m-1} \sum_{b=0}^{n-1} \deg(a, b) && \deg(a, b) \geq 1 \text{ for all } a, b \\ &= \sum_{r \in \{0, \dots, mn-1\}} \deg(r) && \text{by Proposition 4.9} \\ &\leq mn && \deg(r) \leq 1 \text{ for all } r. \end{aligned}$$

So $\sum_{a=0}^{m-1} \sum_{b=0}^{n-1} \deg(a, b) = mn$. Thus it must be that every $\deg(a, b)$ is 1 for all a, b, i.e., for each a, b, there exists a unique solution $r \in \{0, \dots, mn-1\}$. This concludes the proof. □

5.7 Euler's totient function

Definition 5.28. Let $d \geq 2$. We define $\varphi(d)$ to be the size of the set

$$\{a \in \{0, \dots, d-1\} \mid \gcd(a, d) = 1\},$$

i.e., to be the number of numbers in $\{0, \dots, d-1\}$ that are coprime with d. This function is called *Euler's totient function*.

Euler's totient function

Lemma 5.29. *Suppose that $p \geq 2$ is a prime number, and let $k \geq 1$. Then $\varphi(p^k) = (p-1) \cdot p^{k-1}$.*

Proof. Note that for $k = 1$, this should be clear: p being prime, every number in $\{0, \dots, p-1\}$ apart from 0 is coprime with p, so $\varphi(p) = p - 1$.

Now let us consider the general case. The only prime factor of p^k is p by Theorem 5.13. Thus, a is *not* coprime with p^k, if, and only if, its remainder in the division by p is 0. There are p^{k-1} such integers, they are of the form $n \cdot p$ with $n \in \{0, \dots, p^{k-1} - 1\}$. So $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$. □

More generally, one can compute φ of a number given its decomposition into prime factors:

Lemma 5.30. *Let $p_1, \dots, p_n \in \mathbb{N}$ be primes, and let $k_1, \dots, k_n \geq 1$. Let $N := \prod_{i=1}^n p_i^{k_i}$. Then $\varphi(N) = (p_1 - 1) \cdots (p_n - 1) p_1^{k-1} \cdots p_n^{k_n - 1} = \prod_{i=1}^n (p_i - 1) p_i^{k_i - 1}$.*

Comment about the proof

This proof is a bit technical and relies on an application of the inclusion-exclusion principle, as well as on the following fact: if $p_1, \dots, p_n$ are any numbers, then $(p_1 - 1) \times \cdots \times (p_n - 1)$ can be expressed as $\sum_{r=0}^n (-1)^r \sum_{S \in \mathcal{P}_r(\{1,\dots,n\})} \prod_{s \notin S} p_s$. This can be seen as follows: when developping the product, one needs to "take" one of the numbers p_i or -1 from each of the factors. One can first choose an $r \in \{0, \dots, n\}$ that tells us how many -1's we want to take, then we choose an r-element subset of $\{1, \dots, n\}$ to decide *which* of the -1's to take, and for the other $n - r$ positions $s \notin S$ we must take p_s.

Proof. Note that a number $a \in \{0, \dots, N-1\}$ is coprime with N if, and only if, a is not divisible by any p_i (Lemma 5.11). Let D_i be the set of numbers in $\{0, \dots, N-1\}$ that are divisible by p_i. Then $\varphi(N)$ is $N - |\bigcup_{i=1}^n D_i|$, so this begs for an application of the inclusion-exclusion principle (Theorem 4.11).

For this, we would like to know the size of $D_{i_1} \cap \cdots \cap D_{i_r}$ for $1 \leq i_1 < \cdots < i_r \leq n$. This is exactly $N/(p_{i_1} \cdots p_{i_r})$, which is $\prod_{j=1}^n p_j^{\ell_j}$, where $\ell_j = k_j$ if j is not one of $i_1, \dots, i_r$, or $\ell_j = k_j - 1$ if $j \in \{i_1, \dots, i_r\}$. Finally, note that we can write this as $\left(\prod_{i=1}^n p_i^{k_i - 1}\right) \times \prod_{j \notin \{i_1,\dots,i_r\}} p_j$.

Then we compute:

$$\begin{aligned}
\varphi(N) &= N - \sum_{r=1}^n (-1)^{r+1} \sum_{1 \leq i_1 < \cdots < i_r \leq n} N/(p_{i_1} \cdots p_{i_r}) \\
&= N + \left(\sum_{r=1}^n (-1)^r \left(\prod_{i=1}^n p_i^{k_i - 1} \right) \sum_{S \in \mathcal{P}_r(\{1,\dots,n\})} \prod_{s \notin S} p_s \right) && \text{see previous paragraph} \\
&= \sum_{r=0}^n (-1)^r \left(\prod_{i=1}^n p_i^{k_i - 1} \right) \sum_{S \in \mathcal{P}_r(\{1,\dots,n\})} \prod_{s \notin S} p_s && \text{the term } N \text{ is the case } r = 0 \text{ of the sum} \\
&= \left(\prod_{i=1}^n p_i^{k_i - 1} \right) \times \prod_{i=1}^n (p_i - 1) && \text{see comment before the proof}
\end{aligned}$$

which we see is the desired result. □

Corollary 5.31. *If n and m are coprime, then $\varphi(nm) = \varphi(n)\varphi(m)$. We say that φ is an* arithmetic function.

arithmetic function

Proof. Consider the decompositions of n and m into prime factors. Since n and m are coprime, no prime appears in both. By Lemma 5.30, we see that $\varphi(nm)$ is simply $\varphi(n) \times \varphi(m)$. □

The final piece of knowledge about φ that we need to understand RSA is the following result.

Theorem 5.32 (Fermat's little theorem)**.** *Let $a \in \mathbb{Z}, d \geq 2$, and suppose that a and d are coprime. Then $a^{\varphi(d)} = 1 \bmod d$.*

We will see how to prove Theorem 5.32 as a corollary of Theorem 6.13.

5.8 Application: the RSA cryptosystem

We have seen in Lemma 5.30 that if we are given a prime decomposition of n, then one can compute $\varphi(n)$ easily. But as far as we know, it is hard to compute φ without that.

The security of the RSA cryptosystem relies on the hypothesis that φ is hard to compute. In this system, two parties Alice and Bob can communicate using something called *asymmetric* encryption. Alice generates numbers $N = pq$ and e, where p, q are prime numbers and e is coprime with $\varphi(N)$. Note that since Alice knows the prime factorization of N, she can also compute $\varphi(N) = (p-1)(q-1)$. She gives (e, N) to Bob; this is her *public key*. She then computes the inverse d of e modulo $\varphi(N)$: she can do this using Euclid's algorithm since e is coprime with $\varphi(N)$ (Proposition 5.24). Note that since $ed = 1 \bmod \varphi(N)$, there exists $s \in \mathbb{N}$ such that $ed = s\varphi(N)+1 = s(p-1)(q-1)+1$. Alice keeps d for herself; the pair (d, N) is her *private key*.

public key

private key

When Bob wants to send a message (a number $a \in \{0, \ldots, N-1\}$) to Alice, he computes $a^e \bmod N$, the element of $\mathbb{Z}/N\mathbb{Z}$. He can do this efficiently, because division by N can be done easily. And since $[a^e] = [a]^e$, no number $\geq N$ needs to be stored even for large e. To decrypt the element $[b] \in \mathbb{Z}/N\mathbb{Z}$ that she receives, Alice computes $b^d \bmod N$.

We now prove that Alice decrypts an encrypted message correctly.

Lemma 5.33. *If $a \in \{0, \ldots, N-1\}$, then $a^{ed} = a \bmod N$.*

Proof. It suffices to prove that $a^{ed} = a \bmod p$ and $a^{ed} = a \bmod q$: indeed, suppose that $a^{ed} - a = p \cdot k$ and $a^{ed} - a = q \cdot \ell$ for some $k, \ell \in \mathbb{N}_0$ (we have used here Lemma 5.19 to obtain that $a^{ed} = a \bmod p$ iff p divides $a^{ed} - a$, and similarly for q). Then $p \cdot k = q \cdot \ell$, and since p does not divide q it must divide ℓ by Lemma 5.11. So pq divides $q \cdot \ell$, and therefore $N = pq$ divides $a^{ed} - a$. By Lemma 5.19, we get that $a^{ed} = a \bmod N$.

Now we prove that $a^{ed} = a \bmod p$. Suppose first that a is coprime with p. Then we

get:

$$\begin{aligned} a^{ed} \bmod p &\equiv a^{s(p-1)(q-1)+1} \bmod p && \text{since } d \text{ is the inverse of } e \text{ modulo } \varphi(N) = (p-1)(q-1) \\ &\equiv (a^{p-1})^{s(q-1)} \times a \bmod p \\ &\equiv 1^{s(q-1)} \times a \bmod p && \text{by Theorem 5.32 and since } \varphi(p) = p-1 \\ &\equiv a \bmod p \end{aligned}$$

Otherwise, a is divisible by p. But then a^{ed} is divisible by p, too, so we have $a^{ed} = 0 \bmod p$ and $a = 0 \bmod p$, so $a^{ed} = a \bmod p$. □

This "simple" procedure has been used since the 1970's until this day as an encryption scheme for establishing secure server connections on the internet. A procedure to "break" RSA has been made public by Peter Shor in 1994, who found an algorithm to decompose a number N into its prime factors efficiently, but this procedure can only be implemented with a *quantum* computer.[25]

5.9 Exercises

1. Show that the divisibility poset $(\mathbb{N}_0, \mid)$ has a unique maximal element and a unique minimal element.

2. Compute $\gcd(8, 13)$ by following Euclid's algorithm, and compute the coefficients $u, v \in \mathbb{Z}$ whose existence is asserted in Corollary 5.6.

3. Let $a, b \geq 1$. We know from Corollary 5.6 that $\gcd(a, b)$ can be expressed as $ua+vb$ for some $u, v \in \mathbb{Z}$. Show that $\gcd(a, b)$ is the smallest number $d \geq 1$ that can be expressed as $ua + vb$, for $u, v \in \mathbb{Z}$.

4. Let $F_0 = F_1 = 1$, and for $n \geq 2$, let $F_n = F_{n-1} + F_{n-2}$. The numbers F_n are called *Fibonacci numbers*. Show that for all $n \geq 0$, F_n and F_{n+1} are coprime.

5. Show that a and $a + 1$ are coprime, for every $a \in \mathbb{N}$.

6. Show that $ab/\gcd(a, b)$ is the *lowest common multiple* of a and b, denoted by $\operatorname{lcm}(a, b)$.

lowest common multiple

7. Let $a, b, c \in \mathbb{N}$. Show that if a, b are coprime, and c is a multiple of a and of b, then it is a multiple of ab.

8. Show that the remainder of 2^n in the division by 3 is 1 when n is even, and 2 when n is odd.

9. Show that if $[a]$ has an inverse $[b]$ in $\mathbb{Z}/d\mathbb{Z}$, then this inverse is unique. I.e., if $[ab] = [ac] = [1]$, then $[b] = [c]$.

[25] To date, and according to Wikipedia, the largest integer that has been factorized by a quantum computer is 21.

10. Compute $\varphi(d)$ for $d \in \{6, 10, 18, 28\}$.

11. (*) Show that for all $n \geq 1$, we have $\sum_{d|n} \varphi(d) = n$, where $d \in \{1, \ldots, n\}$ ranges over all divisors of n (including n).
(Hint: consider the following combinatorial proof. Define C_d to be the set of numbers in $\{0, \ldots, d-1\}$ that are coprime with d (so that by definition $\varphi(d) = |C_d|$). Define the function $f\colon \{1, \ldots, n\} \to \bigcup_{d|n} C_d \times \{d\}$ by $f(k) = (k/\gcd(k,n), n/\gcd(k,n))$ and show that it is well-defined (i.e., that $k/\gcd(k,n)$ belongs to $C_{n/\gcd(k,n)}$ for all k, and that it is injective and surjective.) Conclude that it gives the desired equality.)

12. Let $\mu\colon \mathbb{N} \to \mathbb{N}$ be defined as follows. We define $\mu(1) = 1$. For $n \geq 2$, let $\prod_{i=1}^{k} p_i^{e_i}$ be its prime factorization. Then we define $\mu(n)$ to be 0 if $e_i > 1$ for some $i \in \{1, \ldots, k\}$, and otherwise $\mu(n) = (-1)^k$. This function is also called the *Möbius function*. Show that μ is an arithmetic function.

Möbius function

6 Algebraic Structures

Things to remember / to know

- Know what associativity and commutativity mean and how to prove that the property holds for a given operation
- Know what a monoid and a group are and be able to prove whether a composition law satisfies the monoid or group properties
- Know what the order of an element is and how to prove that in a finite group every element has an order (Theorem 6.13)
- Know what distributivity is, and be able to prove whether a given $(R, +, \times)$ is a ring
- Be able to determine whether a ring is a field, and know the classification of finite fields (Theorem 6.24)
- Be able to compute with polynomials: addition, multiplication, division with quotient and remainder (Theorem 6.33)
- Know what a Boolean algebra is and know the classification of finite Boolean algebras (Corollary 6.52)

We have seen in the previous chapter a new notion of arithmetic, namely arithmetic in $\mathbb{Z}/d\mathbb{Z}$, and seen an example of how this arithmetic can be useful in computer science. In this final chapter we study an abstraction of this: instead of caring about what the objects that we are adding and multiplying are, we simply care about the laws of multiplication and addition and the properties that they should satisfy.

Definition 6.1. Let A be a set. An *internal composition law* is a function $f\colon A^2 \to A$.

internal composition law

A composition law is often denoted by a symbol like $\cdot, \times, *, \otimes, +, \oplus, \ldots$ and we write $a \times b$ instead of $\times(a, b)$ for $a, b \in A$.

Remark 6.2. The word *internal* refers to the fact that the two arguments of the operation and the result of the operation all come from the same set. For example, the division operation is an internal composition law in $\mathbb{Q} \setminus \{0\}$ because for all $a, b \in \mathbb{Q} \setminus \{0\}$, we have $a/b \in \mathbb{Q} \setminus \{0\}$, but it is not an internal composition law in $\mathbb{Z} \setminus \{0\}$ because $1/2 \notin \mathbb{Z} \setminus \{0\}$.

The usual operations of addition, multiplication are internal composition laws on

sets like $\mathbb{N}, \mathbb{Z}, \mathbb{R}, \ldots$. Exponential (the operation defined by $a \otimes b := a^b$) is an internal composition law on $\mathbb{N}$.

We can study abstract properties of such composition laws, in particular the following properties are important.

Definition 6.3. Let $\otimes$ be a composition law on a set A. We say that $\otimes$ is:

- *associative* if for all $a, b, c \in A$, we have $(a \otimes b) \otimes c = a \otimes (b \otimes c)$;
- *commutative* if for all $a, b \in A$, we have $a \otimes b = b \otimes a$.

associative

commutative

As always, it is nice to have a way to represent the mathematical objects graphically. For operations, this is done by writing the *Cayley table*: it is a table where rows and columns are indexed by the elements in A, and the cell indexed by $(a, b) \in A^2$ (in row a and column b) contains the element $c \in A$ such that $c = a \otimes b$.

Cayley table

Example 6.4. Let $\otimes$ be the operation on $\mathbb{Z}/3\mathbb{Z}$ defined by $[a] \otimes [b] = [ab + a + 1]$. Then the Cayley table of $\otimes$ is the following, where the brackets [and] are omitted for readability:

$\otimes$	0	1	2
0	1	1	1
1	2	0	1
2	0	2	1

Besides the usual arithmetic operations, we have already seen other examples of composition laws:

- Given two subsets $B, C \subseteq A$, then $B \cup C$, $B \cap C$, and $B \Delta C$ are also subsets of A, giving three internal composition laws on $\mathcal{P}(A)$.
- Given two binary relations $R, S \subseteq A \times A$, then $R + S \subseteq A \times A$ is another relation. This gives a composition law on $\mathcal{P}(A \times A)$.
- Given two functions $f, g \colon A \to A$, then $f \circ g$ is another function $A \to A$. This gives a composition law on A^A.
- Given two numbers $a, b \in \mathbb{Z}$, then $\gcd(a, b) \in \mathbb{Z}$.
- Given two Boolean values $a, b \in \{\text{True}, \text{False}\}$, then $a \wedge b$ and $a \vee b$ are also Boolean values given by the truth tables of $\wedge$ and $\vee$. This gives two composition laws on $\{\text{True}, \text{False}\}$.

Studying composition laws in an abstract way allows us to understand properties of all the examples above in a unified way.

6.1 Groups and Monoids

Definition 6.5 (Monoid)**.** A *monoid* is a pair $(M, \otimes)$ where M is a set and $\otimes$ is an internal composition law on M satisfying the following properties:

- $\otimes$ is associative,
- there exists $e \in M$ such that for all $a \in M$, $e \otimes a = a \otimes e = a$.

monoid

We call an element $e \in M$ *neutral* if it satisfies the second property in Definition 6.5. A natural[26] example of a monoid is $(\mathbb{N}_0, +)$, since 0 is a neutral element for $+$ and $+$ is associative. Another example of a monoid is the following, very important in computer science.

neutral element

Example 6.6. Let Σ be a set of letters (say, 0 and 1). Let Σ^* be the set of all *finite strings*, e.g. 000110101 and 1101011. Given two strings s, t, one can define the *concatenation* of s and t as the string starting with the letters of s, followed by the letters of t. We usually denote the concatenation by $s \cdot t$. In the example above, the concatenation is 0001101011101011. Then $(\Sigma^*, \cdot)$ is a monoid, since concatenation is associative and since the *empty string* is a neutral element for $\cdot$.

concatenation

empty string

Another example of a monoid is as follows. Let A be a set. Remember that given two binary relations $R, S \subseteq A \times A$, one can form the composition $R + S$ of R and S as $\{(a, c) \in A \times A \mid \exists b \in A : (a, b) \in R \text{ and } (b, c) \in A\}$. This composition is associative, and moreover the equality relation Δ_A is a neutral element. Then the set of all binary relations on A, $(\text{BinRel}, +)$ is a monoid.

The definition of a monoid only asks that there exists at least one neutral element. As it turns out, there can only be one.

Lemma 6.7. *Let $(M, \otimes)$ be a monoid. Then there exists a unique neutral element, denoted by e_M.*

Proof. Let $e, f \in M$ be neutral elements. Then $e \otimes f = e$ (since f is neutral) and $e \otimes f = f$ (since e is neutral), so $e = f$. □

There are many other examples of monoids, e.g. $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, the monoid of all column vectors $\mathbb{R}^n$ of size n, ... Those latter structures actually satisfy an important additionally property that makes them extremely useful in many areas of mathematics and computer science.

[26] Pun intended.

Definition 6.8 (Group)**.** A *group* is a pair $(G, \otimes)$ where G is a set and $\otimes$ is a composition law on G satisfying the following properties:

- $\otimes$ is associative;
- there exists $e \in G$ such that for all $a \in G$, $e \otimes a = a \otimes e = a$;
- for every $a \in G$, there exists $b \in G$ such that $a \otimes b = b \otimes a = e$.

group

We call b an *inverse of* a if $a \otimes b = b \otimes a = e$. In other words, a group is a monoid where every $a \in G$ has an inverse element.

inverse of an element

We have seen several examples of groups already, for example $(\mathbb{Z}, +), (\mathbb{Q}, +), (\mathbb{R}, +)$. However, one can see that $(\mathbb{N}, +)$ is not a group: the only possible neutral element for $+$ is 0, and there does not exist a $b \in \mathbb{N}$ such that $1 + b = 0$, contradicting the existence of an inverse for 1. Another group we have encountered is $(\mathbb{Z}/d\mathbb{Z}, +)$, for any $d \geq 2$.

Proposition 6.9. *Let A be a set, and let G be the set of functions $f \colon A \to A$ that are bijections. Then $(G, \circ)$ is a group, where $\circ$ denotes the composition of functions.*

Proof. The composition of bijections is a bijection, so $\circ$ is an internal composition law.

We prove the three properties that a group must satisfy. Associativity means that if $f, g, h \in G$, then $(f \circ g) \circ h = f \circ (g \circ h)$. Let $a \in A$. Then

$$\begin{aligned}((f \circ g) \circ h)(a) &= (f \circ g)(h(a)) \\ &= f(g(h(a))) \\ &= f((g \circ h)(a)) \\ &= (f \circ (g \circ h))(a)\end{aligned}$$

so that $(f \circ g) \circ h = f \circ (g \circ h)$ is indeed true.

For the neutral element, we pick $e = \mathrm{id}_A$, for which we have $f \circ \mathrm{id}_A = f = \mathrm{id}_A \circ f$.

Finally, the existence of an inverse follows from Proposition 1.19 applied to the case of $A = B$. □

The group from Proposition 6.9 is called the *symmetric group* on A, sometimes denoted by $\mathfrak{S}_A$. If $A = \{1, \ldots, n\}$ for some $n \geq 1$, then we write $\mathfrak{S}_n$.

symmetric group

Proposition 6.10. *Let $d \geq 2$, and let $(\mathbb{Z}/d\mathbb{Z})^\times$ be the set of elements $[a] \in \mathbb{Z}/d\mathbb{Z}$ where a is coprime to d. Then $((\mathbb{Z}/d\mathbb{Z})^\times, \times)$ is a group.*

Proof. First, we need to show that $\times$ is an *internal* composition law, that is, if $[a], [b] \in \mathbb{Z}/d\mathbb{Z}$ with a, b coprime with d, then also ab is coprime with d. Assume for contradiction that there exists a common factor $p \geq 2$ of ab and d; without loss of generality, this

common factor can be taken to be prime by Lemma 5.8. Then by Lemma 5.11 and since $p \mid ab$, we have that $p \mid a$ or $p \mid b$. So a or b is not coprime with d, a contradiction. Associativity of $\times$ follows from the fact that multiplication in the integers is associative: we have $[a] \times ([b] \times [c]) = [a(bc)] = [(ab)c] = ([a] \times [b]) \times [c]$. The neutral element is $[1]$, since $[a] \times [1] = [a \cdot 1] = [a]$ and $[1] \times [a] = [1 \cdot a] = [a]$. Finally, every $[a]$ has an inverse by Proposition 5.24. □

We have seen in Proposition 1.19 that in the symmetric group $\mathfrak{S}_A$, every $f \in \mathfrak{S}_A$ has a unique inverse. One of the nice things about abstraction (going from studying particular cases of groups to studying the general *concept* of a group) is that it allows to prove this in more generality, once and for all: in a group, every element has a unique inverse.

Lemma 6.11. *Let $(A, \otimes)$ be a group. Then there exists a unique neutral element, and every $a \in A$ has a unique inverse.*

Proof. Let $a \in A$ and let $b, c \in A$ be inverses for a. Then

$$\begin{aligned} b &= b \otimes e && e \text{ is the neutral element} \\ &= b \otimes (a \otimes c) && c \text{ is an inverse of } a \\ &= (b \otimes a) \otimes c && \text{associativity of } \otimes \\ &= e \otimes c && b \text{ is an inverse of } a \\ &= c \end{aligned}$$

so $b = c$. □

Notation 6.12. Let $(G, \otimes)$ be a group. We often write e_G for the neutral element of G (it is unique by Lemma 6.11). For $a \in G$, we write a^{-1} for the inverse of a; again by Lemma 6.11, the inverse of a is unique and thus a^{-1} is uniquely defined. For $a \in G$ and $n \geq 1$ we write a^n for $a \otimes a \otimes \cdots \otimes a$, with n times the factor a. If $n = 0$, then $a^0 = e_G$, by definition. If n is negative, then we write a^n for $(a^{-1})^{-n}$.

A group is called *commutative* if its composition law is commutative. We also say that the group is *abelian*. In an abelian group, we sometimes prefer to use the symbol $+$ or variants like $\oplus$, while in a general group we prefer to use $\times$ or variants like $\otimes, *, \ldots$. And similarly, in an abelian group, rather than using a^n we can use $n \cdot a$ instead.

commutative group

6.2 Order of elements and Fermat's little theorem

Using the concept of groups, it is possible to prove Theorem 5.32, saying that $a^{\varphi(d)} = 1 \bmod d$ for all a, d that are coprime. The following is a generalization of it. Given a group G, $a \in G$, and $n \geq 1$, we say that a has *order* n if $a^n = e_G$ and n is the smallest integer verifying this. Note that not every element needs to have an order in a group: for example, in the group $(\mathbb{Z}, +)$ (whose neutral element is 0), then 1 does not have an order.

order of an element

Theorem 6.13 (Lagrange's theorem)**.** *Let G be a finite group and $a \in G$. Then a has a finite order n, and n divides $|G|$.*

Proof. To prove the existence, consider the function $f\colon \mathbb{N} \to G$ defined by $f(n) = a^n$. By the pigeonhole principle (Proposition 4.13), there must exist numbers n, m such that $f(n) = f(m)$, i.e., $a^n = a^m$. Without loss of generality, we can assume that $n < m$. Then by multiplying by a^{-n} on both sides, we get $e_G = a^{m-n}$, so a has an order smaller or equal to $m - n$.

We now prove that the order m of a divides $|G|$. Define an equivalence relation $\sim$ on G by $x \sim y$ iff there exists $p \in \mathbb{Z}$ such that $x \otimes a^p = y$ (the fact that this is an equivalence relation is left as an exercise). Note that $a^n \sim e_G$ for all n, and if $b \sim a^n$ then by definition there exists p such that $b = a^n \otimes a^p = a^{n+p}$. So the equivalence class of a consists exactly of the elements of the form a^n for some n. By definition of the order of a, $[a]$ contains exactly m elements $e_G, a, a^2, \ldots, a^{m-1}$.

We show that every equivalence class $[b]$ also contains exactly m elements. Let the function $f\colon [a] \to [b]$ be defined by $a^i \mapsto b \otimes a^i$. This map is injective: if $ba^i = ba^j$, then $a^i = a^j$. Moreover, it is surjective: let $c \in [b]$. By definition, there exists $p \in \mathbb{Z}$ such that $c = b \otimes a^p$, and therefore $f(a^p) = c$ so f is surjective. Thus, f is a bijection and we have $m = |[a]| = |[b]|$.

Therefore, G can be partitioned into equivalence classes $[b_1], \ldots, [b_k]$ (see Lemma 3.15 and the paragraphs following it), and every equivalence class has size m. Thus $|G| = |[b_1] \cup \cdots \cup [b_k]| = |[b_1]| + \cdots + |[b_k]|$. Since every term in the sum is divisible by m (in fact, equal to m), then $|G|$ is divisible by m, too (Lemma 5.3). □

Proof of Theorem 5.32. By Proposition 6.10, the set $(\mathbb{Z}/d\mathbb{Z})^\times$ forms a group under multiplication, where $[1]$ is the neutral element. By definition of $\varphi(d)$, the set contains precisely $\varphi(d)$ elements.

Let $a \in \{0, \ldots, d-1\}$ be coprime with d. By Theorem 6.13, $[a]$ has a finite order m that divides $\varphi(d)$. Then $[a^{\varphi(d)}] = [a]^{\varphi(d)} = ([a]^m)^{\varphi(d)/m} = [1]^{\varphi(d)/m} = [1]$. Thus, $a^{\varphi(d)} = 1 \bmod d$. □

6.3 Morphisms between groups

Definition 6.14 (Homomorphism)**.** Let $(G, \otimes_G)$ and $(H, \otimes_H)$ be two groups. A function $f\colon G \to H$ is a *homomorphism of groups* from $(G, \otimes_G)$ to $(H, \otimes_H)$ if the following holds:

group homomorphism

- $f(e_G) = e_H$,
- for all $a, b \in G$, we have $f(a \otimes_G b) = f(a) \otimes_H f(b)$,
- for all $a \in G$, we have $f(a^{-1}) = f(a)^{-1}$ (the first inverse is taken for an element in $(G, \otimes_G)$, the second inverse is for an element in $(H, \otimes_H)$).

We write $f\colon (G, \otimes_G) \to (H, \otimes_H)$ to say that f is a homomorphism between the two groups.

The canonical inclusion maps $\mathbb{Z} \to \mathbb{Q} \to \mathbb{R} \to \mathbb{C}$ are all group homomorphisms between the respective groups whose addition is the composition law. There are of course other examples.

Example 6.15. Let $f\colon \mathbb{Z}/6\mathbb{Z} \to \mathbb{Z}/3\mathbb{Z}$ be defined by $f([x]_6) = [2x]_3$, where $[x]_6$ refers to the equivalence class of $x \in \mathbb{Z}$ under the relation $\equiv_6$, while $[\cdot]_3$ refers to the equivalence class under the relation $\equiv_3$. We first check that this is well-defined: suppose that $[x]_6 = [y]_6$, i.e., $x = y \bmod 6$. Then there exists $k \in \mathbb{Z}$ such that $x - y = 6k$, by Lemma 5.19. So $2x - 2y = 3 \cdot (4k)$ and $2x = 2y \bmod 3$, so $[2x]_3 = [2y]_3$.
We now check that f is a homomorphism from $(\mathbb{Z}/6\mathbb{Z}, +)$ to $(\mathbb{Z}/3\mathbb{Z}, +)$. To check this, note that the neutral element in $(\mathbb{Z}/6\mathbb{Z}, +)$ is $[0]_6$, while the neutral element in $(\mathbb{Z}/3\mathbb{Z}, +)$ is $[0]_3$. Since $[2 \cdot 0]_3 = [0]_3$, we have $f([0]_6) = [0]_3$. Let now $x, y \in \mathbb{Z}$. Then $f([x]_6 + [y]_6) = f([x+y]_6) = [2x+2y]_3 = [2x]_3 + [2y]_3 = f([x]_6) + f([y]_6)$, so the second property from Definition 6.14 holds. Finally, $f(-[x]_6) = f([-x]_6) = [2 \cdot (-x)]_3 = -[2x]_3 = -f([x]_6)$ so the third property also holds.

If a homomorphism $f\colon (G, \otimes_G) \to (H, \otimes_H)$ is a bijection, we say that it is an *isomorphism*, and we say that $(G, \otimes_G)$ and $(H, \otimes_H)$ are *isomorphic*. When two groups are isomorphic, we can consider that they are essentially the same: the isomorphism is an operation that renames the elements of G and assigns them to elements of H, but otherwise does not change anything to the group. In particular if two groups are isomorphic and one of them is Abelian, then so is the other.

isomorphism, isomorphic groups

Example 6.16. Consider the set G consisting of the following 3 matrices:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad A = \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} \quad B = \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}$$

We leave as an exercise to prove that $\times$ (the multiplication of matrices) is an internal composition law on G, and in fact that $(G, \times)$ is a group. Then $(G, \times)$ is isomorphic to $(\mathbb{Z}/3\mathbb{Z}, +)$, where an isomorphism can be defined by $f(I) = [0], f(A) = [1], f(B) = [2]$.

The notion of isomorphism really should be seen as "those two objects are really just the same, only the elements are called differently." We will see what this notion means for other structures.

Many of the statements that we proved for functions between sets also hold for homomorphisms between groups. For example homomorphisms can be composed, and

every bijective homomorphism $f\colon (G,\otimes_G)\to(H,\otimes_H)$ admits an inverse f^{-1} that is a homomorphism from $(H,\otimes_H)$ to $(G,\otimes_G)$. Moreover, if $f\colon (G,\otimes_G)\to(H,\otimes_H)$ is a homomorphism, then there exists a composition law $*$ on $G/\ker(f)$ such that $(G/\ker(f),*)$ is a group and such that Theorem 3.22 is true when considering homomorphisms instead of functions.

Theorem 6.17. *Let $(G,\otimes_G)$ and $(H,\otimes_H)$ be groups with a homomorphism $f\colon (G,\otimes_G)\to(H,\otimes_H)$. The law $[a]*[b]=[a\otimes_G b]$ on $G/\ker(f)$ is a well-defined internal composition law, and $(G/\ker(f),*)$ is a group.*

Proof. We first have to show that $*$ is well-defined. Suppose that $(a,a')\in\ker(f)$ and that $(b,b')\in\ker(f)$. We have to show that $(a\otimes_G b, a'\otimes_G b')\in\ker(f)$. We have

$$\begin{aligned} f(a\otimes_G b) &= f(a)\otimes_H f(b) && \text{since } f \text{ is a homomorphism}\\ &= f(a')\otimes_H f(b') && \text{since } f(a)=f(a') \text{ and } f(b)=f(b')\\ &= f(a'\otimes_G b') \end{aligned}$$

so $(a\otimes_G b, a'\otimes_G b')\in\ker(f)$.

We now show that $*$ satisfies the group properties. Associativity follows from the fact that $\otimes_G$ is associative: $[a]*([b]*[c])=[a]*[b\otimes_G c]=[a\otimes_G(b\otimes_G c)]=[(a\otimes_G b)\otimes_G c]=([a*b])*[c]$. Moreover $[e_G]$ is a neutral element, since $[e_G]*[a]=[e_G\otimes a]=[a]$ and similarly $[a]*[e_G]=[a]$. Finally, the inverse of any element $[a]$ is $[a^{-1}]$: we have $[a]*[a^{-1}]=[a\otimes_G a^{-1}]=[e_G]$. □

In many applications, knowing properties about $G/\ker(f)$ and H then allows us to derive properties about G itself. This is a generalization of the same concept that we saw for sets.

Remark 6.18. Note that if we know $[e_G]_{\ker(f)}$, the equivalence class of the neutral element of G, then we know exactly the kernel. Indeed, we have $(a,b)\in\ker(f)$ if, and only if, $ab^{-1}\in[0]_{\ker(f)}$: $f(a)=f(b)$ is equivalent to $f(a)f(b)^{-1}=e_H$, which is the same as $f(ab^{-1})=e_H=f(e_G)$ and $(ab^{-1},e_G)\in\ker(f)$. This is why in the context of groups $\ker(f)$ refers rather to the equivalence class of e_G rather than to the equivalence relation. See how this relates to the notion of kernel that you have studied in linear algebra in the case where f is a linear map.

6.4 Rings, Fields

Let us move on with our program of finding useful abstractions of the classical structures that we know.

We have seen that in $\mathbb{Z}/d\mathbb{Z}$, not only one can add $[a]$ and $[b]$ but also one can multiply them. Same in all the usual arithmetic structures we are familiar with: $\mathbb{Z},\mathbb{Q},\mathbb{R},\mathbb{C},\dots$. In particular in the case of $\mathbb{Z}$ we could also find some interesting notions like divisibility, prime numbers, prime decompositions, ... based on the multiplication operation.

In all the examples above, multiplication and addition obey a particular property called *distributivity* and formally defined as follows.

distributivity

Definition 6.19 (Distributivity)**.** Let $\oplus$ and $\otimes$ be internal composition laws on the same set A. We say that $\otimes$ *distributes over* $\oplus$ if for all $a, b, c \in A$, we have

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

This allows us to define the notion of ring.

Definition 6.20 (Ring)**.** Let R be a set with two internal composition laws $\oplus$ and $\otimes$. Then $(R, \oplus, \otimes)$ is a *ring* if the following properties hold:

ring

- $(R, \oplus)$ is an Abelian group whose neutral element is written 0 (or 0_R),
- $\otimes$ is associative and has a neutral element denoted by 1 (or 1_R) that we assume to be different from 0,
- $\otimes$ distributes over $\oplus$.

Some natural rules of addition and multiplication are not part of this definition, for example the fact that $0 \otimes r = 0$ is true for all $r \in R$, but it is something one can prove from the given definition (Exercise 23).

Thus, $(\mathbb{Z}, +, \times)$ is an example of a ring, and so is $(\mathbb{Q}, +, \times)$.

Proposition 6.21. *Let $d \geq 2$. Then $(\mathbb{Z}/d\mathbb{Z}, +, \times)$ is a ring.*

Proof. We already know that $(\mathbb{Z}/d\mathbb{Z}, +)$ is a group with neutral element $[0]$. It is clear that $\times$ is associative: $([a] \times [b]) \times [c] = [abc] = [a] \times ([b] \times [c])$. Moreover, $[1]$ is a neutral element for $\times$. The fact that $\times$ distributes over $+$ follows from the fact that this is true in $\mathbb{Z}$: $[a] \times ([b] + [c]) = [a(b + c)] = [ab + ac] = [ab] + [ac] = [a] \times [b] + [a] \times [c]$. □

Some confusion might arise from the fact that we have two composition laws in a ring. We always use the additive notation for the law $\oplus$, meaning that if $r \in R$ and $n \geq 1$, then $n \cdot r$ denotes the element $r \oplus \cdots \oplus r$, while r^n denotes the element $r \otimes \cdots \otimes r$. We use $-r$ for the inverse of r in the group $(R, \oplus)$ (the unique element such that $(-r) + r = 0 = r + (-r)$), also called *additive inverse* of r. We use the notation r^{-1} for the inverse of an element r in $(R, \otimes)$ if it has one, i.e., we have $r \otimes r^{-1} = 1 = r^{-1} \otimes r$. This element is called a *multiplicative inverse*. But since $(R, \otimes)$ is not a group, not every element has a multiplicative inverse: in fact, 0 (the neutral element of $\oplus$) *never* has one.

additive inverse

multiplicative inverse

Just like in $\mathbb{Z}$, we say that a *divides* b if there exists $q \in R$ such that $b = q \otimes a$. The notion of divisibility in $\mathbb{Q}, \mathbb{R}, \ldots$ is useless: every non-zero $a \in \mathbb{R}$ divides every $b \in \mathbb{R}$ (take $q = b/a$). But we saw that over $\mathbb{Z}$ the notion of divisibility leads to interesting

divisibility in a ring

concepts such as greatest common divisors, Euclid's algorithm, ... We will see how this generalizes in the case of some particular rings below.

For now, let us look at the case where the divisibility notion is "useless," where every non-zero element divides everything else. As long as a has an inverse a^{-1}, then a divides any b by taking $q = b \otimes a^{-1}$. Conversely, if a divides everything then in particular it divides 1, so that it has a multiplicative inverse a^{-1}. This leads to the following definition.

field

Definition 6.22 (Field, *Körper* in German)**.** A ring $(R, \oplus, \otimes)$ is called a *field* if $\otimes$ is commutative and every element different from 0 (the neutral element of $\oplus$) admits a multiplicative inverse. In other words, $(R \setminus \{0\}, \otimes)$ is an Abelian group.

As described above, we already know many fields: $\mathbb{R}, \mathbb{Q}, \mathbb{C}$ are all fields (when equipped with the usual addition and multiplication laws). Those fields happen to be infinite, but this does not have to be.

Theorem 6.23. *Let $d \geq 2$. Then $(\mathbb{Z}/d\mathbb{Z}, +, \times)$ is a field if, and only if, d is prime.*

Proof. Suppose that d is not prime, so that there exist $a, b \in \{2, \dots, d-1\}$ with $ab = d$. Then $[a] \times [b] = [d] = [0]$. So $[a]$ cannot have a multiplicative inverse: otherwise, we would have $[a]^{-1} \times [a] \times [b] = [b]$ and $[a]^{-1} \times [0] = 0$, so $[b] = 0$.

Now, suppose that d is prime. Then all the elements $[1], \dots, [d-1]$ have a multiplicative inverse (Proposition 5.24). Moreover since $\times$ is commutative we have that $(\mathbb{Z}/d\mathbb{Z} \setminus \{[0]\}, \times)$ is an Abelian group. □

There exists a complete characterization of the fields with a finite number of elements. We say that a map $f : (R, \oplus_R, \otimes_R) \to (S, \oplus_S, \otimes_S)$ is a *homomorphism of rings* if for all $r_1, r_2 \in R$ it satisfies that $f(r_1 \oplus_R r_s) = f(r_1) \oplus_S f(r_2)$, $f(r_1 \otimes_R r_2) = f(r_1) \otimes_S f(r_2)$ and $f(1_R) = 1_S$. An isomorphism is a homomorphism that is bijective.

ring homomorphism

Theorem 6.24. *Let n be a number. There exists a field of size n if, and only if, $n = p^k$ for some prime $p \geq 2$ and arbitrary exponent $k \geq 1$. Moreover, this field is unique (up to isomorphism of rings) and denoted by $\mathbb{F}_n$ or $\mathrm{GF}(n)$.*

The proof of Theorem 6.24 is slightly beyond the scope of this course.

Recall that by definition, if $(R, \oplus, \otimes)$ is a field then $(R \setminus \{0\}, \otimes)$ is a group. It turns out that if R is finite, then this group is in fact very easy to understand.

Theorem 6.25. *Let $(R, \oplus, \otimes)$ be a finite field. Then $(R \setminus \{0\}, \otimes)$ is isomorphic to $(\mathbb{Z}/n\mathbb{Z}, +)$ for $n = |R| - 1$.*

Proof. We know by Theorem 6.13 that every element a in $(R \setminus \{0\}, \otimes)$ has a finite order d, and that d divides n. Let $\psi(d)$ be the number of $a \in R \setminus \{0\}$ that have order d. In particular $\sum_{d|n} \psi(d) = n$ (this is a translation of the fact that every element has some order, and that this order is unique).

Fix $d \mid n$. Either $\psi(d) = 0$, or there exists $a \in R \setminus \{0\}$ that has order d. Then the set of elements $\{a^i \mid i \in \{0, \dots, d-1\}\}$ consists of d elements and they all satisfy the equation $x^d = 1$ (in the field $(R, \oplus, \otimes)$). Since an equation of degree d in a field has at most d solutions,[27] it must be that every solution of this equation is of the form a^i for some $i \in \{0, \dots, d-1\}$. In particular, every element b of order d is of the form a^i for $i \in \{0, \dots, d-1\}$, and i must be coprime with d: if $\gcd(i, d) > 1$, then $d/\gcd(i, d) < d$ and $b^{d/\gcd(i,d)} = a^{d \cdot i/\gcd(i,d)} = (a^d)^{i/\gcd(i,d)} = 1$, a contradiction to the order of b being d. So there are at most $\varphi(d)$ elements of order d.

Suppose for contradiction that $\psi(n) = 0$. Then $\sum_{d|n} \psi(d) < \sum_{d|n} \varphi(d)$. It can be checked that $\sum_{d|n} \varphi(d) = n$ (Exercise 11), so that $\sum_{d|n} \psi(d) < n$, a contradiction.

So $\psi(n) > 0$, and in particular there exists an element a of order n, and therefore simply because $|R \setminus \{0\}| = n$ by definition of n, we must have $R \setminus \{0\} = \{a^0, a^1, \dots, a^{n-1}\}$. Define a function $f \colon (\mathbb{Z}/n\mathbb{Z}, +) \to (R \setminus \{0\}, \otimes)$ by $f(i) = a^i$. One checks that this is a homomorphism of groups and moreover it is surjective since $R \setminus \{0\} = \{a^0, a^1, \dots, a^{n-1}\}$. Since $\mathbb{Z}/n\mathbb{Z}$ has size n, it must be that f is injective, too. This means that f is an isomorphism of groups and concludes the proof. □

Any element $a \in R \setminus \{0\}$ with order $|R|-1$ in $(R \setminus \{0\}, \otimes)$ is called *primitive*. Primitive elements are important in applications to computer science such as in Fourier analysis (which appears for example in signal processing) or error-correcting codes, which we will see in Chapter 6.6.2. Before seeing this application, we need to do one more detour.

primitive element

6.5 Polynomials

In the following section, we **fix a ring** $(R, +, \times)$ **that is either a field or the ring** $(\mathbb{Z}, +, \times)$ **of integers.**

Definition 6.26. A *polynomial* P with coefficients R is a sequence $(p_0, p_1, \dots)$ where for all large enough n, we have $p_n = 0$, and each $p_i \in R$. The elements $p_0, p_1, \dots$ are called the *coefficients* of the polynomial; p_i is the coefficient *of degree* i. The set of all polynomials with coefficients in R is denoted by $R[X]$.

polynomial

coefficients

The *zero polynomial* is $(0, 0, \dots)$, the sequence containing only 0s. We also denote this polynomial by 0. If one wants to make a difference, one can write 0_R for the additive neutral element in the ring $(R, +, \times)$, and $0_{R[X]}$ for the zero polynomial in $R[X]$. The *degree* of a polynomial $P \neq 0_{R[X]}$ is the largest n such that $p_n \neq 0_R$. It is denoted by $\deg(P)$. A polynomial of degree 0 is called *constant*, a polynomial of degree 1 is called *linear*. We write 1 (or $1_{R[X]}$ if we really want to avoid confusion) for the polynomial

zero polynomial, $0_{R[X]}$

degree

$1_{R[X]}$

[27] See Corollary 6.38 for a proof of this fact.

$(1, 0, \dots)$ of degree 0. Every element $r \in R$ can be seen as a constant polynomial, namely the polynomial $(r, 0, 0, \dots)$.

Definition 6.27. One can add and multiply polynomials with coefficients in R:

$$(p_0, p_1, \dots) +_{R[X]} (q_0, q_1, \dots) = (p_0 + q_0, p_1 + q_1, \dots)$$

and

$$(p_0, p_1, \dots) \times_{R[X]} (q_0, q_1, \dots) = (r_0, r_1, \dots)$$

where

$$r_k = \sum_{i=0}^{k} p_i \times q_{k-i},$$

where the summation notation $\sum \dots$ refers to addition with $+$.

Note that it is not clear that the resulting $(r_0, r_1, \dots)$ is a polynomial, since it could be that r_k is non-zero for infinitely many k. We first see an example and prove that this does not happen.

One way of illustrating how the coefficient r_k in the product is computed is given in Figure 25.

Figure 25: A representation of the product of the polynomials $P = (1, 1, 1, 0, \dots)$ and $Q = (1, 2, 3, 0, \dots)$. By listing Q from right to left and aligning its kth coefficient q_k with p_0, computing the kth coefficient r_k of the product amounts to multiplying term by term and then summing up the results. Any product involving 0 is ignored. In the picture, this is shown for $k = 2$ (left) and $k = 3$ (right). Here, $P \times Q$ is $(1, 3, 6, 5, 3, 0, \dots)$.

Example 6.28. Let us consider the ring of integers $(\mathbb{Z}, +, \times)$. The polynomial $(1, -2, 4, -8, 0, 0, 0, \dots) \times_{\mathbb{Z}[X]} (1, 2, 3, 0, 0, 0, 0, \dots)$ is the sequence $(r_0, r_1, \dots)$ where:

- $r_0 = 1 \times 1 = 1$

- $r_1 = -2 \times 1 + 1 \times 2 = 0$
- $r_2 = 4 \times 1 + (-2) \times 2 + 1 \times 3 = 3$,
- $r_3 = -8 \times 1 + 4 \times 2 + (-2) \times 3 + 1 \times 0 = -6$
- $r_4 = 0 \times 1 + (-8) \times 2 + 4 \times 3 + (-2) \times 0 + 1 \times 0 = -4$
- $r_5 = 0 \times 1 + 0 \times 2 + (-8) \times 3 + 4 \times 0 + (-2) \times 0 + 1 \times 0 = -24$,
- $r_6 = 0 \times 1 + 0 \times 2 + 0 \times 3 + (-8) \times 0 + 4 \times 0 + (-2) \times 0 + 1 \times 0 = 0$

From this point on, one sees that $r_7 = r_8 = \cdots = 0$, so $(r_0, r_1, \dots)$ is indeed a polynomial in $\mathbb{Z}[X]$, equal to $(1, 0, 3, -6, -4, -24, 0, \dots)$ of degree 5.

Proposition 6.29. *Let P and Q be polynomials in $R[X]$. Then $P +_{R[X]} Q$ is a polynomial that is either 0 or of degree at most $\max(\deg(P), \deg(Q))$. Moreover, $P \times_{R[X]} Q$ is also a polynomial, and if $P \times_{R[X]} Q$ is non-zero then it is of degree at most $\deg(P) + \deg(Q)$.*

Proof. Let $(r_0, r_1, \dots)$ be $P +_{R[X]} Q$. It is clear that this is a polynomial: if k is larger than $\max(deg(P), \deg(Q))$, then $r_k = p_k + q_k = 0 + 0 = 0$.

Let now $(r_0, r_1, \dots)$ be $P \times_{R[X]} Q$ and let $k > \deg(P) + \deg(Q)$. Then the coefficient r_k is defined to be $\sum_{i=0}^{k} p_i \times q_{k-i}$. For any $i > \deg(P)$, we have $p_i = 0_R$. Otherwise, if $i \leq \deg(P)$, then we must have $k - i \geq k - \deg(P) > (\deg(P) + \deg(Q)) - \deg(P) = \deg(Q)$, so that $q_{k-i} = 0_R$. Therefore, all the terms in the sum are 0_R, and $r_k = 0_R$. □

Notation 6.30. We *define* X to be the polynomial $(0, 1, 0, \dots)$ of degree 1. Then for all $n \geq 1$, the polynomial X^n (obtained as the multiplication of X with itself, n times) is equal to $(0, \dots, 0, 1, 0, \dots)$ where the 1 appears at position $n + 1$. Therefore, every polynomial $(p_0, p_1, \dots, p_n, 0, \dots)$ of degree n can be written as a finite sum $p_0 + p_1 X + p_2 X^2 + \cdots + p_n X^n$.

For example, $(1, -2, 4, -8, 0, \dots)$ is an element of $\mathbb{Z}[X]$ that we denote by $1 - 2X + 4X^2 - 8X^3$.

Remark 6.31. Note the familiarity with the polynomial functions over $\mathbb{R}$ that you know, like the functions that map $x \in \mathbb{R}$ to x^2 or $x \in \mathbb{R}$ to $-x^3 + x^2 - 3x + 1$. However these are very different objects: while in a polynomial function, the symbol x stands for an element of $\mathbb{R}$, the symbol X is itself a polynomial. There is a natural bijection between all the polynomial functions over $\mathbb{R}$ and the elements of $\mathbb{R}[X]$, the proof is left as an exercise.
However, there are only finitely many polynomial functions over $\mathbb{Z}/2\mathbb{Z}$. Indeed,

there are only finitely many functions $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$ in the first place, but there are infinitely many polynomials in $(\mathbb{Z}/2\mathbb{Z})[X]$.
Consider the following concrete example. As *functions* $\mathbb{Z}/2\mathbb{Z} \to \mathbb{Z}/2\mathbb{Z}$, $x \mapsto x$ and $x \mapsto x^2$ are equal (remember that a number is odd iff its square is odd), but the *polynomials* $P = X$ and $Q = X^2$ in $(\mathbb{Z}/2\mathbb{Z})[X]$ are different.

Theorem 6.32. *$(R[X], +_{R[X]}, \times_{R[X]})$ is also a ring. The neutral element for $+_{R[X]}$ is $0_{R[X]}$, and the neutral element for $\times_{R[X]}$ is $1_{R[X]}$.*

Comment about the proof

The proof is a bit tedious in computation. However it is a good exercise to understand it (or even better, to try to write it yourself), since it involves using all the properties defining rings.

Proof. We already know from Proposition 6.29 that $+_{R[X]}$ and $\times_{R[X]}$ are internal composition laws on $R[X]$.

We sketch a proof that $(R[X], +_{R[X]})$ is a group. Associativity of $+_{R[X]}$ comes from the fact that $+$ is associative. The fact that $0_{R[X]}$ is a neutral element for $+_{R[X]}$ is clear since 0_R is a neutral element for $+$. Finally, the additive inverse of a polynomial $(p_0, p_1, \dots)$ is $(-p_0, -p_1, \dots)$.

The fact that $1_{R[X]}$ is a neutral element for $\times_{R[X]}$ should also be clear: let $P = (p_0, p_1, \dots)$ be a polynomial, let $1_{R[X]} = (q_0, q_1, \dots)$ and let $(r_0, r_1, \dots)$ be the polynomial $P \times 1_{R[X]}$. Then $r_k = \sum_{i=0}^{k} p_i \times_R q_{k-i}$. Since q_{k-i} is not zero only for $i = k$, in which case $q_{k-i} = 1_R$, we get $r_k = p_k$, so $P = P \times 1_{R[X]}$. The proof for $1_{R[X]} \times P = P$ is similar.

We now show that $\times_{R[X]}$ is associative. Let A, B, C be polynomials with coefficients $(a_k)_k, (b_k)_k, (c_k)_k$ respectively. Then $B \times_{R[X]} C$ is a polynomial with coefficients $d_k = \sum_{j=0}^{k} b_{k-j} \times c_j$. Let e_k be the coefficient of degree k of $A \times_{R[X]} (B \times_{R[X]} C)$. Then we

get

$$
\begin{aligned}
e_k &= \sum_{i=0}^{k} a_i \times d_{k-i} && \text{definition of the product} \\
&= \sum_{i=0}^{k} a_i \times \left(\sum_{j=0}^{k-i} b_{k-i-j} \times c_j \right) && \text{definition of the product for } B \times_{R[X]} C \\
&= \sum_{i=0}^{k} \sum_{j=0}^{k-i} a_i \times (b_{k-i-j} \times_R c_j) && \text{distributivity of } \times \text{ over } + \\
&= \sum_{i=0}^{k} \sum_{j=0}^{k-i} (a_i \times b_{k-i-j}) \times c_j && \text{associativity of } \times \\
&= \sum_{j=0}^{k} \sum_{i=0}^{k-j} (a_i \times_R b_{k-i-j}) \times c_j && \text{commutativity of } + \text{ to exchange the order of the sums} \\
&= \sum_{j=0}^{k} \left(\sum_{i=0}^{k-j} a_i \times b_{k-i-j} \right) \times c_j && \text{using distributivity to factor out } c_j \\
&= \sum_{j=0}^{k} f_{k-j} \times c_j
\end{aligned}
$$

where f_{k-j} is the coefficient of degree $k-j$ of $A \times_{R[X]} B$. This is the expression for the coefficient of degree k of $(A \times_{R[X]} B) \times_{R[X]} C$ and therefore $A \times_{R[X]} (B \times_{R[X]} C)$ and $(A \times_{R[X]} B) \times_{R[X]} C$ have the same coefficients, i.e., they are the same polynomial.

It remains to prove that $\times_{R[X]}$ distributes over $+_{R[X]}$. Let A, B, C be polynomials and denote their coefficients as above. Then $A \times_{R[X]} (B +_{R[X]} C)$ has coefficients $\sum_{i=0}^{k} a_i \times (b_{k-i} + c_{k-i})$, which is $\sum_{i=0}^{k} (a_i \times b_{k-i} + a_i \times c_{k-i})$ since $\times$ distributes over $+$. Since $+$ is commutative, the order of the terms can be exchanged and the terms can be regrouped as $\left(\sum_{i=0}^{k} a_i \times b_{k-i}\right) +_R \left(\sum_{i=0}^{k} a_i \times c_{k-i}\right)$, which is the sum of the coefficients of degree k of $A \times_{R[X]} B$ and of $A \times_{R[X]} C$, so $A \times_{R[X]} (B +_{R[X]} C) = A \times_{R[X]} B +_{R[X]} A \times_{R[X]} C$. The proof for $(B +_{R[X]} C) \times_{R[X]} A = B \times_{R[X]} A +_{R[X]} C \times_{R[X]} A$ is symmetric. □

As it turns out, we can also *divide* polynomials with remainders just like in $\mathbb{Z}$, if we assume that their coefficients are from a field.

Theorem 6.33 (Polynomial Division)**.** *Let $\mathbb{K}$ be a field. Let $A, B \in \mathbb{K}[X]$, and suppose that $A \neq 0$. There exist unique polynomials $Q, R \in \mathbb{K}[X]$ such that the following hold:*

- $B = Q \times A + R$,
- *if $R \neq 0_{\mathbb{K}[X]}$, then* $\deg(R) < \deg(A)$.

Remark 6.34. The theorem fails in $\mathbb{Z}[X]$: taking $B = 2X, A = 3X$, either $R = 0$ or since $\deg(R) < \deg(A) = 1$ then R must be a constant. Similarly, Q must also be a constant otherwise $\deg(QA) \geq 2 > \deg(B)$. But since there is no integer q such that $3 \cdot q = 2$, we cannot find a quotient and a remainder in the division of B by A.

Proof. This is similar to division of integers as you know it (Theorem 5.2). Algorithm 6.1 is an algorithm computing Q, R. In the iterations of Line 3, we always keep the property that $B = QA+R$. It is clear that this holds before the first time, since $Q = 0$ and $R = B$. Otherwise, let Q, R be the polynomials at the beginning of a loop and $R' = R - AS$ and $Q' = Q+S$. Then we have $Q'A+R' = (Q+S)A+R-AS = AQ+R+SA-AS = AQ+R$.

We simply have to check that the loop cannot go on forever, so that the result is well-defined. For this, observe that if $\deg(R) = n$, then the coefficient of degree n in $R - AS$ is 0. Moreover, $\deg(R - AS) \leq \deg(R)$, which overall gives $\deg(R - AS) < \deg(R)$. Thus, the degree of R decreases at every iteration, meaning that at some point either we reach $R = 0$ or $\deg(R) < \deg(A)$. □

Algorithm 6.1: Polynomial division in $\mathbb{K}[X]$ when $\mathbb{K}$ is a field

Data: Polynomials $A, B \in \mathbb{K}[X]$, with $A \neq 0$
Result: Polynomials $Q, R \in \mathbb{K}[X]$ such that $B = QA + R$ and such that if $R \neq 0$, then $\deg(R) < \deg(A)$

1 $Q \leftarrow 0$;
2 $R \leftarrow B$;
3 $a \leftarrow$ coefficient of highest degree in A;
4 **while** $R \neq 0$ *and* $\deg(R) \geq \deg(A)$ **do**
5 $\quad r \leftarrow$ coefficient of highest degree in R;
6 $\quad S \leftarrow r/a \cdot X^{\deg(R)-\deg(A)}$;
7 $\quad R \leftarrow R - AS$;
8 $\quad Q \leftarrow Q + S$;
9 **end**
10 **return** (Q, R)

Note that the division of integers was the only thing that we needed to define Algorithm 5.1. This gives us that any two polynomials A, B have a greatest common divisor, just like integers do! The only difference is that it is only defined up to multiplication by a non-zero constant.

Example 6.35. Let $A = X^2 - 3X + 2$ and $B = X^2 - 4X + 3$ be elements of $\mathbb{R}[X]$. We follow Euclid's algorithm to compute their greatest common divisor. The division of B by A gives $Q = 1$ and $R_2 = -X + 1$. The division of A by R_2 gives

$Q' = -X + 2$ and $R_3 = 0$. Therefore the algorithm stops here, and $\gcd(A, B)$ equals the last non-zero remainder, in this case $-X+1$. But note that $X-1$ also divides A and B in $\mathbb{R}[X]$. Both $X-1$ and $-X+1$ can be called $\gcd(A, B)$.

Moreover, Algorithm 5.1 gives us an analog of Corollary 5.6, lemma 5.11, and theorem 5.13 for polynomials.

Given a polynomial $P \in R[X]$ (say $p_0 + p_1X + \cdots + p_nX^n$) and $r \in R$, one can *evaluate* P at r, by computing $p_0 + p_1r + \cdots + p_nr^n$ (which is an element of R). We denote this element $P(r)$. We say that r is a *root* if $P(r) = 0_R$, the (additive) neutral element of R.

evaluation of a polynomial

root

Proposition 6.36. *The map $ev_r \colon R[X] \to R$ defined by $P \mapsto P(r)$ is a homomorphism of rings, that is, it satisfies*

$$ev_r(P +_{R[X]} Q) = ev_r(P) + ev_r(Q) \quad \text{and} \quad ev_r(P \times_{R[X]} Q) = ev_r(P) \times ev_r(Q)$$

as well as $ev_r(1_{R[X]}) = 1_R$.

Proof. Let $n = \max(\deg(P), \deg(Q))$. Let $P = (p_0, p_1, \dots)$ and $Q = (q_0, q_1, \dots)$. Then

$$\begin{aligned}
\mathrm{ev}_r(P) +_R \mathrm{ev}_r(Q) &= \left(\sum_{i=0}^{n} p_i \times r^i\right) + \left(\sum_{i=0}^{n} q_i \times r^i\right) \\
&= \sum_{i=0}^{n} (p_i \times r^i +_R q_i \times r^i) && + \text{ is commutative} \\
&= \sum_{i=0}^{n} (p_i + q_i) \times r^i && \times \text{ distributes over } + \\
&= \mathrm{ev}_r(P +_{R[X]} Q).
\end{aligned}$$

Let s_k be the kth coefficient of $P \times_{R[X]} Q$, i.e., $s_k = \sum_{i=0}^{k} p_i \times q_{k-i}$. Note that

$P \times_{R[X]} Q$ has degree at most $2n$. Then

$$\begin{aligned}
\mathrm{ev}_r(P \times_{R[X]} Q) &= \sum_{k=0}^{2n} s_k \times r^k \\
&= \sum_{k=0}^{2n} \sum_{i=0}^{k} (p_i \times q_{k-i}) \times r^k \\
&= \sum_{k=0}^{2n} \sum_{i=0}^{k} (p_i \times q_{k-i}) \times r^i \times r^{k-i} \\
&= \sum_{k=0}^{2n} \sum_{i=0}^{k} (p_i \times r^i) \times (q_{k-i} \times r^{k-i}) \qquad \text{since } \times \text{ is commutative} \\
&= \left(\sum_{i=0}^{n} p_i \times r^i \right) \times \left(\sum_{i=0}^{n} q_i \times r^i \right) \\
&= \mathrm{ev}_r(P) \times \mathrm{ev}_r(Q).
\end{aligned}$$

□

Lemma 6.37. *Let $\mathbb{K}$ be a field. Let $P \in \mathbb{K}[X]$ and let $r \in \mathbb{K}$ be a root of P. Then $(X - r)$ divides P.*

Proof. If P is the zero polynomial, then it is clear: $(X - r) \times_{\mathbb{K}[X]} 0_{\mathbb{K}[X]} = 0_{\mathbb{K}[X]}$ so $(X - r)$ divides P. Suppose now that P is nonzero and has degree n. We want to find a polynomial Q such that $Q \times_{\mathbb{K}[X]} (X - r) = P$.

Let us use Theorem 6.33: there exist $Q, R \in \mathbb{K}[X]$ such that $P = Q \times (X - r) + C$, and if $C \neq 0_{\mathbb{K}[X]}$ then $\deg(C) < \deg(X - r) = 1$, so $\deg(C) = 0$. So C is either the zero polynomial or a constant. If we evaluate P at r, we get 0 by assumption. Therefore, $((Q \times (X - r)) + C)(r) = 0$. This is equal to $Q(r) \times (r - r) + C(r)$, by Proposition 6.36. Since $r - r = 0$, $Q(r) \times (r - r) = 0$, so we get $C(r) = 0$. Since we know that $C = c_0$ for some $c_0 \in \mathbb{K}$, $C(r) = 0$ implies $c_0 = 0$, and therefore C is 0. It follows that $P = Q \times (X - r)$, as desired. □

Corollary 6.38. *Let $\mathbb{K}$ be a field. Let $P \in \mathbb{K}[X]$ be non-zero and have degree n. Then P has at most n roots in $\mathbb{K}$.*

Proof. Suppose that P has d distinct roots $r_1, \ldots, r_d \in \mathbb{K}$. Then by Lemma 6.37, the polynomials $(X - r_1), \ldots, (X - r_d)$ all divide P. Moreover, these polynomials are all coprime (i.e., the greatest common divisor of any two of them must have degree 0 and be a constant). Just like for integers, this means that their product $Q = (X - r_1) \times \cdots \times (X - r_d)$ also divides P. We have $\deg(Q) = d$ and $\deg(Q) \leq \deg(P)$, so $d \leq n$. □

6.6 Application: Error-correcting codes

6.6.1 Motivation

Nowadays, every piece of information that is stored or transmitted from one device to another is represented in binary, using 0s and 1s. One might think that the technology is very robust and that if a message $0001001111010101011\ldots$ is stored on a hard drive or sent over the internet, then this message is exactly what will be read later from the drive or received from the other side. However this is very much not the case, the communication mediums that we use are not 100% reliable and errors might slip inside the data.

For devices in space this is very common: energetic rays coming from the sun or outside of our solar system carry enough energy that when they hit a transistor (the physical device that holds or 0 or a 1 in a classical computer) then the transistor can change value. This is why processors/memory units/... that are sent in space must be shielded against such radiations. But one doesn't even need to get to space to observe such an event of a "spontaneous bit flip." A famous example where this went absolutely wrong is in the case of a Belgian election in 2003. At the end of the day and while counting the votes in two different ways, one candidate ended up with a disparity of 4096 votes between the two counts. Since $4096 = 2^{12}$, it is assumed that the bit corresponding to 2^{12} in memory was flipped by a cosmic ray. The Belgian system of having the possibility to count the votes in two different ways is what saved the day. But in the event that the corresponding bits flip on the two counters, there would have been no chance of understanding what had happened.

An *error-correcting code* is a system to encode messages in a way to make them robust against such random errors. In the weak version (called *error-detecting code*), one simply wants to know that the message was altered, while in the strong version one wants to be able to recover it. Of course, if for some reason *many* bits of the messages that you want to send get changed without you knowing, then it is impossible to recover the original message. So we build such families of codes that are more or less resilient to errors.

One trivial way of building a code would be to repeat each bit of the message a certain number of times: the message $b_0b_1b_2\ldots$ is encoded as $b_0b_0b_0b_0b_0b_1b_1b_1b_1b_1b_2b_2b_2b_2b_2\ldots$. This way, even if every bit has a 40% chance of flipping while stored/transmitted, then on average in each group $b_ib_ib_ib_ib_i$, 2 of the bits would flip so that the majority of bits would retain their proper value. On the receiving end, it suffices to decode the message and take the majority of the values in each group to recover the proper message. The drawback of this method is that the message length has been multiplied by 5, which is obviously not good (imagine wanting to download the 1Gb file `barbie.mp4`, and having to download 5Gb instead just because of the error-correcting code). We describe here a family of codes called *Reed-Solomon codes* that use finite fields and polynomials to encode data in a more succinct way.

6.6.2 Reed-Solomon codes

By Theorem 6.24, for every $s \geq 1$ there exists a field of size 2^s denoted by $GF(2^s)$ (recall that for $s \geq 2$, this field is *not* $(\mathbb{Z}/2^s\mathbb{Z}, +, \times)$, which is definitely not a field by Theorem 6.23). For brevity we simply write $\mathbb{K}$ for this field.

In this section, define a *packet* to be any group of s bits (i.e., a string made of 0s and 1s of length s). Since there are exactly 2^s such strings, one can see such strings as elements of $\mathbb{K}$ by picking an arbitrary bijection between $\{0,1\}^s$ and $\mathbb{K}$. Let $k, t \geq 1$ be two parameters. We show how Reed-Solomon encodes a group of k packets.

First, pick a primitive element α in $\mathbb{K}$, i.e., $\alpha \in \mathbb{K} \setminus \{0\}$ is an element of order $2^s - 1$ in the multiplicative group. Yet another formulation is that $\{\alpha^i \mid i \in \{0, \dots, 2^s - 2\}\}$ equals $\mathbb{K} \setminus \{0\}$. We know that such an α exists by Theorem 6.25.

Let G be the polynomial in $\mathbb{K}[X]$ defined by $\prod_{i=1}^{2t}(X - \alpha^i)$, where $\prod \dots$ refers to the multiplication in $\mathbb{K}[X]$. If one tried to develop this expression, one would get something of the form $X^{2t} - (\sum_i \alpha^i)X^{2t-1} + (\sum_{i \neq j} \alpha^{i+j})X^{2t-2} - \dots - \alpha^1 \cdots \alpha^{2t}$. This is called the *generator polynomial*, and it has degree $2t$. Every α^i is a root of G for $i \in \{1, \dots, 2t\}$, and therefore these are exactly the roots of G by Corollary 6.38.

Now let $c_0, \dots, c_{k-1} \in \mathbb{K}$, that we see as k packets that we want to encode. Let C be the polynomial $(c_0, \dots, c_{k-1}, 0, \dots)$, i.e., $c_0 + c_1 X + \dots + c_{k-1}X^{k-1}$. The encoded data is obtained by computing the polynomial $D = GC$ whose coefficients are $(d_0, d_1, \dots, d_{k+2t}, 0, \dots)$ of degree at most $\deg(G) + \deg(C) = 2t + k - 1$. Each d_i is an element of $\mathbb{K}$ and therefore can be seen as a packet, which is what we store/transmit. Thus, we started with a message of length $k \cdot s$ (k packets, each consisting of s bits) and ended up storing/sending $(k + 2t) \cdot s$ bits. Instead of having multiplied the message length by 5 like in the example at the beginning of the section, we multiplied by $(k + 2t) \cdot s/(k \cdot s) = 1 + 2t/k$. Thus if we can keep the ratio $2t/k$ small, this is not a big overhead. We will see how robust this code is in Theorem 6.39.

Supposing that the transmission went without problem and that one receives the polynomial D correctly, it is simple to decode the message and recover C. For this, it suffices to apply Theorem 6.33 and divide D by G; the quotient is then C.

We now prove that this encoding is robust against errors.

Theorem 6.39 (Error-correcting and detecting power of the Reed-Solomon code)**.** *Let $k, t \geq 1$ be such that $k + 2t \leq 2^s - 1$. Let $c_0, \dots, c_{k-1} \in \mathbb{K}$ be packets, and let $d_0, \dots, d_{k+2t-1} \in \mathbb{K}$ be the obtained encodings. Let $f_0, \dots, f_{k+2t-1} \in \mathbb{K}$ be the data after transmission/retrieval from storage. If $f_i \neq d_i$ for at most t indices, then one can recover the whole original message $c_0, \dots, c_{k-1}$. If $f_i \neq d_i$ for at most $2t$ indices, then one can detect whether the data was corrupted.*

Comment about the proof

In the following proof, we use some basic methods from linear algebra and in particular we consider solutions of equations of the form

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Remember that a solution to this equation (i.e., values for the variables $x_1, \ldots, x_n$) corresponds to taking a combination of the various columns whose result is the desired vector:

$$x_1 \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} + \cdots + x_n \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

If we have a solution $(x_1, \ldots, x_n)$ and $x_1 = 0$, then this means that the column 1 in the matrix does not appear in the combination, and therefore by forgetting about position 1 (removing x_1 from the solution and removing the column 1 from the matrix) we obtain a solution to

$$\begin{pmatrix} a_{12} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m2} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

Of course this works for any i such that $x_i = 0$, not just x_1. This observation is used extensively in the coming proof.

Proof. Let $e_j = f_j - d_j$ for all $j \in \{0, \ldots, k + 2t - 1\}$. Thus, we have $f_j = e_j + d_j$. Moreover, $e_j = 0$ for all j if, and only if, the transmitted packet is received correctly.

Let E be the polynomial $e_0 + e_1 X + \cdots = \sum_{i=0}^{k+2t-1} e_i X^i$ and F be $f_0 + f_1 X + \cdots = \sum_{i=0}^{k+2t-1} f_i X^i$. Note that $F = E + D$, and therefore $F(\alpha^i) = E(\alpha^i) + D(\alpha^i)$ for all $i \in \{1, \ldots, 2t\}$ (Proposition 6.36). Moreover, α^i is by construction a root of G for every i (indeed, $X - \alpha^i$ divides G). So $D(\alpha^i) = G(\alpha^i)C(\alpha^i) = 0$ by Proposition 6.36, from which we get

$$F(\alpha^i) = E(\alpha^i) = \sum_{j=0}^{k+2t-1} e_j \alpha^{i \cdot j} \tag{6.1}$$

If the message was sent without error, then we have $e_j = 0$ for all j, and therefore $F(\alpha^i) = 0$ for all i. We now prove that in the other direction, supposing that $F(\alpha^i) = 0$ for all $i \in \{1, \ldots, 2t\}$, then either $e_j = 0$ for all $j \in \{0, \ldots, k + 2t - 1\}$, or $e_j \neq 0$ for *many* indices, i.e., for at least $2t + 1$ indices $j \in \{0, \ldots, k + 2t - 1\}$.

One can write the right-hand side expression of Equation (6.1) as a row-column multiplication, namely $(\alpha^0, \alpha^i, \ldots, \alpha^{i(k+2t-1)}) \cdot \begin{pmatrix} e_0 \\ e_1 \\ \vdots \\ e_{k+2t-1} \end{pmatrix}$. Doing this for all $i \in \{1, \ldots, 2t\}$, we arrive at the matrix-column multiplication

$$\begin{pmatrix} \alpha^0 & \alpha^1 & \ldots & \alpha^{k+2t-1} \\ \alpha^0 & \alpha^2 & \ldots & \alpha^{2(k+2t-1)} \\ \vdots & \vdots & & \vdots \\ \alpha^0 & \alpha^{2t} & \ldots & \alpha^{2t(k+2t-1)} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{k+2t-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{6.2}$$

and therefore the errors $e_0, \ldots, e_{k+2t-1}$ (that we do not know) are solutions of a system of linear equations that we *do* know, since we know α and F.

Let us write V for the $(2t) \times (k+2t)$ matrix above. Suppose that $e_j \neq 0$ for at most $2t$ indices $j_1, \ldots, j_{2t}$. Thus in Equation (6.2), only the columns $j_1, \ldots, j_{2t}$ of V are "used" in the linear combination to get the 0 vector, as explained before the start of the proof above. In other words, $(e_{j_1}, \ldots, e_{j_{2t}})$ is a solution to the equation

$$\begin{pmatrix} \alpha^{j_1} & \alpha^{j_2} & \ldots & \alpha^{j_{2t}} \\ (\alpha^{j_1})^2 & (\alpha^{j_2})^2 & \ldots & (\alpha^{j_{2t}})^2 \\ \vdots & \vdots & & \vdots \\ (\alpha^{j_1})^{2t} & (\alpha^{j_2})^{2t} & \ldots & (\alpha^{j_{2t}})^{2t} \end{pmatrix} \cdot \begin{pmatrix} x_{j_1} \\ x_{j_2} \\ \vdots \\ x_{j_{2t}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{6.3}$$

Since $\mathbb{K}$ is a field, all the classical results from linear algebra that you know hold for systems of linear equations over $\mathbb{K}$ just the same as for $\mathbb{R}$.[28] We then know that if this matrix is invertible, then the only solution to this equation is the 0 vector (see Proposition 3.70, page 82 in the script of your Math 1 course). Note that $\alpha^0, \alpha^1, \ldots, \alpha^{k+2t-1}$ are all different by the choice of α. They are also all different from 0. By a classical result in linear algebra known as Vandermonde's identity (see Lemma 6.40), this means that the matrix in Equation (6.3) is invertible. Therefore the only solution to Equation (6.3) is the 0 vector, and we must have $e_{j_1} = \cdots = e_{j_{2t}} = 0$.

Summarizing, we have the following possible outcomes:

- $F(\alpha^i) \neq 0$ for some i: then there must be some error $e_j \neq 0$, but we do not necessarily know for which j, and what is e_j. We can only *detect* that some error occurred.
- $F(\alpha^i) = 0$ for all i: either there were no errors at all in the transmission, or there were at least $2t+1$ errors.

[28] Look into your linear algebra course and observe that all one needs for defining determinants, ranks, kernels, ... only uses properties about addition and multiplication.

Finally, we show that if there are at most t errors, then one can in fact recover D. Let us consider again Equation (6.2), this time without the assumption that $F(\alpha^i) = 0$ for all i (since we also want to consider the case where there are some errors):

$$\begin{pmatrix} \alpha^0 & \alpha^1 & \dots & \alpha^{k+2t-1} \\ \alpha^0 & \alpha^2 & \dots & \alpha^{2(k+2t-1)} \\ \vdots & \vdots & & \vdots \\ \alpha^0 & \alpha^{2t} & \dots & \alpha^{2t(k+2t-1)} \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{k+2t-1} \end{pmatrix} = \begin{pmatrix} F(\alpha^1) \\ F(\alpha^2) \\ \vdots \\ F(\alpha^{2t}) \end{pmatrix}$$

Let $J_{err} = \{j \in \{0, \dots, k+2t-1\} \mid e_j \neq 0\}$ be the set of positions where the errors happened. By assumption on the number of errors, we have $|J_{err}| \leq t$. Let $J = \{j_1, \dots, j_t\}$ be a set of indices in $\{0, \dots, k+2t-1\}$ of size t. We say that J is *good* if the equation

$$\begin{pmatrix} \alpha^{j_1} & \alpha^{j_2} & \dots & \alpha^{j_t} \\ (\alpha^{j_1})^2 & (\alpha^{j_2})^2 & \dots & (\alpha^{j_t})^2 \\ \vdots & \vdots & & \vdots \\ (\alpha^{j_1})^{2t} & (\alpha^{j_2})^{2t} & \dots & (\alpha^{j_t})^{2t} \end{pmatrix} \cdot \begin{pmatrix} x_{j_1} \\ x_{j_2} \\ \vdots \\ x_{j_t} \end{pmatrix} = \begin{pmatrix} F(\alpha^1) \\ F(\alpha^2) \\ \vdots \\ F(\alpha^{2t}) \end{pmatrix} \tag{6.4}$$

has a solution. We claim that J is good if, and only if, $J_{err} \subseteq J$.

Suppose first that $J_{err} \subseteq J$. Then $(e_{j_1}, \dots, e_{j_t})$ is a solution to Equation (6.4), so J is good.

Conversely, suppose that J is good and let $(a_{j_1}, \dots, a_{j_t})$ be a solution to Equation (6.4). Let $J' = \{j'_1, \dots, j'_t\}$ be a set of size t containing J_{err}. We know from the above that J' is good, and has a solution $(e_{j'_1}, \dots, e_{j'_t})$ to Equation (6.4). Let $J'' = \{j''_1, \dots, j''_{2t}\}$ be a set of size $2t$ containing $J \cup J'$. Such a set exists since $|J| = |J'| = t$ and therefore $|J \cup J'| \leq 2t$. Then

$$\begin{pmatrix} \alpha^{j''_1} & \alpha^{j''_2} & \dots & \alpha^{j''_{2t}} \\ (\alpha^{j''_1})^2 & (\alpha^{j''_2})^2 & \dots & (\alpha^{j''_{2t}})^2 \\ \vdots & \vdots & & \vdots \\ (\alpha^{j''_1})^{2t} & (\alpha^{j''_2})^{2t} & \dots & (\alpha^{j''_{2t}})^{2t} \end{pmatrix} \cdot \begin{pmatrix} x_{j''_1} \\ x_{j''_2} \\ \vdots \\ x_{j''_{2t}} \end{pmatrix} = \begin{pmatrix} F(\alpha^1) \\ F(\alpha^2) \\ \vdots \\ F(\alpha^{2t}) \end{pmatrix} \tag{6.5}$$

has the solutions b, c defined by

$$b_j = \begin{cases} a_j & j \in J \\ 0 & j \notin J \end{cases} \quad \text{and} \quad c_j = \begin{cases} e_j & j \in J_{err} \\ 0 & j \notin J_{err} \end{cases}$$

To see that b and j are solutions, remember the explanation before the proof and observe that b takes a combination of the columns $\{j_1, \dots, j_t\}$ from the matrix in Equation (6.5), which then coincides with the matrix in Equation (6.4) and we had assumed that $(a_{j_1}, \dots, a_{j_t})$ was a solution to this equation. The same reasoning applies to c.

But as we discussed, the matrix in Equation (6.5) is invertible and therefore Equation (6.5) has a unique solution, therefore $b = c$. If $J_{err} \not\subseteq J$, then there exists $j \in J_{err} \setminus J$.

Then $b_j = 0$ since $j \notin J$ and $c_j = e_j \neq 0$ since $j \in J_{err}$, so $b_j \neq c_j$, a contradiction. Thus, we must have $J_{err} \subseteq J$.

Therefore, to decode the message, one can iterate over all possible sets $J \subseteq \{0, \dots, k+2t-1\}$ with $|J| = t$, check for each of them whether they are good (by solving Equation (6.4) using e.g. Gauß elimination). As soon as we have found a good J, then we know it contains J_{err} by the previous paragraphs, and therefore the corresponding solution to Equation (6.4) gives us all the values for e_j. To obtain the original message, it suffices to compute $d_j = f_j - e_j$ for all $j \in \{0, \dots, k+2t-1\}$. □

The field of classical error correction is not at all modern (even in the digital age, Hamming developed the theory of error-correcting codes in the 1940's). However there is now a renewed interest in error correction in relation to quantum computing, since the *qubits* building a quantum computers are even less reliable for retaining their value than the classical transistors.

6.6.3 Vandermonde's identity

For completeness, we include a proof of Vandermonde's identity that we used in the proof of Theorem 6.39.

Lemma 6.40 (Vandermonde)**.** *Let $\mathbb{K}$ be a field, let $x_1, \dots, x_n \in \mathbb{K}$. Then*

$$V(x_1, \dots, x_n) = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & & \vdots \\ x_1^n & x_2^n & \dots & x_n^n \end{pmatrix}$$

has determinant $x_1 \dots x_n \prod_{1 \leq i < j \leq n} x_j - x_i$. Therefore it is invertible if, and only if, $x_i \neq x_j$ for all $i \neq j$ and $x_i \neq 0$ for all i.

Proof. The proof is by induction on $n \geq 1$. If $n = 1$ the matrix is (x_1) whose determinant is indeed x_1 (note that the product in the formula is empty for $n = 1$ and therefore has value 1).

Suppose that $n > 1$ and the lemma has been proved for smaller values of n. We know that performing row operations does not change the determinant. Therefore, by subtracting to every row i the row $i-1$ multiplied by x_1, we obtain that $V(x_1, \dots, x_n)$ has the same determinant as

$$\begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 0 & x_2^2 - x_1 x_2 & \dots & x_n^2 - x_1 x_n \\ \vdots & \vdots & & \vdots \\ 0 & x_2^n - x_1 x_2^{n-1} & \dots & x_n^n - x_1 x_n^{n-1} \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ 0 & x_2(x_2 - x_1) & \dots & x_n(x_n - x_1) \\ \vdots & \vdots & & \vdots \\ 0 & x_2^{n-1}(x_2 - x_1) & \dots & x_n^{n-1}(x_n - x_1) \end{pmatrix}$$

and using the determinant expansion formula we get that the determinant of this matrix

is equal to

$$x_1 \det \begin{pmatrix} x_2(x_2 - x_1) & \dots & x_n(x_n - x_1) \\ \vdots & & \vdots \\ x_2^{n-1}(x_2 - x_1) & \dots & x_n^{n-1}(x_n - x_1) \end{pmatrix}$$

Finally, by multilinearity of the determinant, we can factor out all the factors $(x_2 - x_1), \dots, (x_n - x_1)$ which gives

$$x_1(x_2 - x_1) \cdots (x_n - x_1) \det \begin{pmatrix} x_2 & \dots & x_n \\ \vdots & & \vdots \\ x_2^{n-1} & \dots & x_n^{n-1} \end{pmatrix}$$

We recognize the matrix in this expression to be $V(x_2, \dots, x_n)$ whose determinant is by induction hypothesis $x_2 \cdots x_n \prod_{2 \le i < j \le n} x_j - x_i$. By rearranging the terms we get the desired formula. □

6.7 Boolean Algebras

There still exists one kind of algebraic structures that we have not discussed yet. We remarked at the beginning of this chapter that $\wedge, \vee$ are internal composition laws on the set $\mathbb{B} = \{\text{True}, \text{False}\}$. However $(\mathbb{B}, \wedge)$ and $(\mathbb{B}, \vee)$ are not groups, so nothing that we saw so far can express properties of these algebraic structures. Still, we know that $\wedge$ distributes over $\vee$ and that $\vee$ and $\wedge$ have neutral elements, for example, so there is still *something* interesting going on. Let us take all the properties that we can easily see about $(\mathbb{B}, \vee, \wedge)$ and create a definition from that.

Definition 6.41 (Boolean Algebra)**.** Let B be a set, let $\wedge, \vee$ be arbitrary internal composition laws on B, and let $\neg \colon B \to B$ be a function. We say that $(B, \vee, \wedge, \neg)$ is a *Boolean algebra* if the following properties are satisfied:

Boolean algebra

- $\vee$ and $\wedge$ are associative, commutative, and have neutral elements denoted by 0 and 1,
- $\vee$ distributes over $\wedge$ *and* $\wedge$ distributes over $\vee$,
- for every $b \in B$, $b \vee \neg b = 1$ and $b \wedge \neg b = 0$.

Note that the properties we ask about $\vee$ and $\wedge$ are symmetric. Thus $(B, \vee, \wedge, \neg)$ is a Boolean algebra if, and only if, $(B, \wedge, \vee, \neg)$ is a Boolean algebra.

Let us see that $(\mathbb{B}, \vee, \wedge, \neg)$ is a Boolean algebra. The first two items in Definition 6.41 are clearly true as we have already seen, where the neutral element for $\vee$ is False and the neutral element for $\wedge$ is True. The last item can be seen since $\text{True} \vee \neg\text{True} = \text{True} \vee \text{False} = \text{True} = 1$ and $\text{True} \wedge \neg\text{True} = \text{True} \wedge \text{False} = \text{False} = 0$.

Proposition 6.42. *Let S be a set. Then $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$ is a Boolean algebra, where $S \setminus \_$ is the function that sends A to its complement $S \setminus A$.*

This gives a family of examples of Boolean algebra. In the following, we will obtain a complete list of the finite Boolean algebras (just like Theorem 6.24 gives a complete list of the finite fields). For this, we need to define a notion of homomorphisms between Boolean algebras.

Definition 6.43. Let $(B, \vee_B, \wedge_B, \neg_B)$ and $(C, \vee_C, \wedge_C, \neg_C)$ be two Boolean algebras. A function $f\colon B \to C$ is a *homomorphism of Boolean algebras* if it satisfies $f(a \wedge_B b) = f(a) \wedge_C f(b)$, $f(a \vee_B b) = f(a) \vee_C f(b)$, and $f(\neg_B a) = \neg_C f(a)$ for all $a, b \in B$. As for groups and rings, a homomorphism is called an isomorphism if the function is bijective.

homomorphism of Boolean algebras

For example, the function $f\colon \emptyset \mapsto \text{False}, \{\star\} \mapsto \text{True}$ is an isomorphism of Boolean algebras from the algebra $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$ to $(\mathbb{B}, \vee, \wedge, \neg)$: namely, it satisfies

$$f(A \cup B) = f(A) \vee f(B) \quad f(A \cap B) = f(A) \wedge f(B) \quad f(S \setminus A) = \neg f(A)$$

Thus, in fact all the Boolean algebras that we have seen so far are of the form $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$ for some set S. We will see that as far as finite Boolean algebras are concerned, these are in fact the only examples.

Proposition 6.42 allows us to draw many parallels between the algebra on True/False that we know, and computations with sets that we saw in the very first chapter. Any property that we can prove using only the properties in Definition 6.41 will be true in every Boolean algebra, for example in $(\mathbb{B}, \vee, \wedge, \neg)$ and $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$.

Another less formal parallel is the following. For example, remember that $A \Rightarrow B$ is logically equivalent to $\neg A \vee B$. This expression, interpreted in $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$, becomes $(S \setminus A) \cup B$. Note that $(S \setminus A) \cup B$ is equal to S if, and only if, $A \subseteq B$. Since in the Boolean algebra $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$ the neutral element for $\cap$ is S and in the Boolean algebra $(\mathbb{B}, \vee, \wedge, \neg)$ it is "True," we can see a connection between the predicates $A \Rightarrow B$ and $A \subseteq B$.

Let us now see some properties of Boolean algebras in general. We proved in the case of groups that a neutral element for the group operation must be unique (Lemma 6.11); the same proof shows that the neutral element for $\vee$ and $\wedge$ is also unique. Moreover, the "complement" element of any $b \in B$ is uniquely defined:

Lemma 6.44. *Let $(B, \vee, \wedge, \neg)$ be a Boolean algebra, let $a, b \in B$. Suppose that $a \wedge b = 0$ and $a \vee b = 1$. Then $a = \neg b$.*

Proof. We have

$$\begin{aligned} a &= a \vee 0 && 0 \text{ is neutral for } \vee \\ &= a \vee (b \wedge \neg b) && b \wedge \neg b = 0 \text{ by definition} \\ &= (a \vee b) \wedge (a \vee \neg b) && \text{distributivity of } \vee \text{ over } \wedge \\ &= 1 \wedge (a \vee \neg b) && \text{assumption that } a \vee b = 1 \\ &= a \vee \neg b && 1 \text{ is neutral for } \wedge. \end{aligned}$$

Similarly, we get

$$\begin{aligned} \neg b &= 0 \vee \neg b && 0 \text{ is neutral for } \vee \\ &= (b \wedge a) \vee \neg b && \text{assumption that } b \wedge a = 0 \\ &= (b \vee \neg b) \wedge (a \vee \neg b) && \text{distributivity of } \vee \text{ over } \wedge \\ &= 1 \wedge (a \vee \neg b) && b \vee \neg b = 1 \text{ by definition} \\ &= a \vee \neg b && 1 \text{ is neutral for } \wedge. \end{aligned}$$

Thus $a = a \vee \neg b = \neg b$ and we are done. □

In particular, we obtain that $\neg(\neg a) = a$ and $\neg 0 = 1$ in every Boolean algebra (see Exercise 35).

Lemma 6.45. *In every Boolean algebra, $\wedge$ and $\vee$ are* idempotent*, that is, $a \wedge a = a$ and $a \vee a = a$ for all a.*

idempotent operation

Proof. Let a be an element of the algebra. Then $a \wedge a = (a \wedge a) \vee 0 = (a \wedge a) \vee (a \wedge \neg a) = a \wedge (a \vee \neg a) = a \wedge 1 = a$. The proof for $a \vee a = a$ is symmetric. □

Lemma 6.46. *In every Boolean algebra, $a \wedge 0 = 0$ for all a.*

Proof. We have $a \wedge 0 = a \wedge (a \wedge \neg a) = (a \wedge a) \wedge \neg a = a \wedge \neg a = 0$. □

Another important property that we saw for $(\mathbb{B}, \vee, \wedge, \neg)$ are the DeMorgan laws (Lemma 2.10): $\neg(a \wedge b) = \neg a \vee \neg b$ and $\neg(a \vee b) = \neg a \wedge \neg b$. The proof that we gave at that time simply looked at all the possibility for $a, b \in \mathbb{B}$. But in fact one can prove this for all Boolean algebras in general.

Lemma 6.47. *Let $(B, \vee, \wedge, \neg)$ be a Boolean algebra. Then the De Morgan laws hold in B.*

Proof. By Lemma 6.44, to prove $\neg(a \wedge b) = \neg a \vee \neg b$, it suffices to prove that $(\neg a \vee \neg b)$ satisfies the properties of the complement of $a \wedge b$, that is, it should satisfy $(\neg a \vee \neg b) \wedge (a \wedge b) = 0$ and $(\neg a \vee \neg b) \vee (a \wedge b) = 1$. But $(\neg a \vee \neg b) \wedge (a \wedge b) = (\neg a \wedge a \wedge b) \vee (\neg b \wedge a \wedge b) = 0$ and $(\neg a \vee \neg b) \vee (a \wedge b) = (\neg a \vee \neg b \vee a) \wedge (\neg a \vee \neg b \vee b) = 1$, so we are done. □

6.7.1 Orders from Boolean algebras

Proposition 6.48. *Let $(B, \vee, \wedge, \neg)$ be an arbitrary Boolean algebra. Define $a \leq b$ by $a \wedge b = a$. Then $\leq$ is a partial order.*

Proof. The reflexivity of $\leq$ follows from $\wedge$ being idempotent (Lemma 6.45). The antisymmetry follows from $\wedge$ being commutative: if $a \leq b$ and $b \leq a$, then $a = a \wedge b = b \wedge a = b$, so $a = b$. Finally, suppose that $a \leq b$ and $b \leq c$. Then $a \wedge c = (a \wedge b) \wedge c = a \wedge (b \wedge c) = a \wedge b = a$, so $a \leq c$. □

Let us see what this order is in the Boolean algebra that we know. In $\mathcal{P}(S)$, we have $A \cap B = A$ if, and only if, $A \subseteq B$. Thus, the order associated with the Boolean algebra $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$ is simply $(\mathcal{P}(S), \subseteq)$. In the Boolean algebra $(\mathbb{B}, \vee, \wedge, \neg)$, this is the partial order where False $\leq$ True.

This order always has a unique minimal element: since $0 \wedge a = 0$ for every element a, then $0 \leq a$ for all a. Moreover, it has a unique maximal element as well, since $a \wedge 1 = a$ for all a, we get $a \leq 1$ for all a.

Let $(B, \leq)$ be a partial order with a unique minimal element 0. Define an *atom* to be an $a \in B \setminus \{0\}$ such that $0 \leq a$ and for every b such that $0 \leq b \leq a$, either $b = 0$ or $b = a$. In words, this means that an atom is an element on the second level of the Hasse diagram. The atoms of $(\mathbb{N}_0, \mid)$ are exactly the prime numbers. Note that for two atoms a, b, one has that $a \wedge b = 0$ if $a \neq b$, and $a \wedge b = a$ if $a = b$ (Exercise 39).

atom

Lemma 6.49. *Let $(B, \leq)$ be a finite partial order with a unique minimal element 0 and let $b \in B$ be an element that is not minimal. Then there exists an atom a such that $a \leq b$.*

Proof. Suppose that no such atom exists. We prove that for all $n \geq 0$, there exists a sequence $b = b_0, b_1, b_2, \ldots, b_n$ such that $0 < b_{i+1} < b_i$ for all i. Since B is finite, for $n = |B| + 1$ this is a contradiction.

The proof is by induction on n, the base case being $n = 0$. Then b is a sequence satisfying the statement.

Suppose now that a sequence $b = b_0, \ldots, b_n$ has already been constructed. Note that b_n cannot be an atom, since if it were then it would be an atom smaller than b, a contradiction to our assumption that no such atom exists. Thus, there exists $b_{n+1} \in B \setminus \{0, b_n\}$ such that $0 \leq b_{n+1} \leq b_n$, which proves that a sequence also exists for $n+1$. □

Remark 6.50. Note that this gives an alternative proof of the fact that every natural number $n > 1$ is divisible by a prime number (Lemma 5.8). Indeed, prime numbers are the atoms in the divisibility poset $(\mathbb{N}, \mid)$. While this poset is infinite and therefore Lemma 6.49 does not apply "out-of-the-box," we see that the proof of Lemma 6.49 that we gave works for any poset where there are no infinite

descending chains, which is the case in the poset $(\mathbb{N}_0, \mid)$. This is yet another opportunity to see the benefits of abstractions: by introducing a more general and abstract concept (i.e., atoms in an partial order) with a unique minimal element, we can derive known results about special cases and the proof can often be made simpler since we can get rid of all the unnecessary notions (in this case, it turns out that Lemma 5.8 does not have anything to do with numbers, it is purely because of the divisibility relation having the right structural properties).

6.7.2 Classification of finite Boolean algebras

Proposition 6.51. *Let $(B, \vee, \wedge, \neg)$ be a finite Boolean algebra, and let $b \in B \setminus \{0\}$. Let $a_1, \dots, a_n$ be a complete list of all the atoms of B such that $a_i \leq b$ for all i. Then $b = a_1 \vee \cdots \vee a_n = \bigvee_{i=1}^n a_i$.*

Proof. Let $c = \bigvee_{i=1}^n a_i$. Note that $c \wedge b = (\bigvee_{i=1}^n a_i) \wedge b = \bigvee_{i=1}^n (a_i \wedge b) = \bigvee_{i=1}^n a_i = c$, so $c \leq b$. Let $d = b \wedge \neg c$. Note that $\neg c$ is the same as $\bigwedge_{i=1}^n \neg a_i$ by Lemma 6.47.

We prove that $d \neq 0$ or $b = c$. We have $b = b \wedge 1 = b \wedge (c \vee \neg c) = (b \wedge c) \vee (b \wedge \neg c) = c \vee d$. So if $d = 0$, then we obtain $b = c$ and we are done. Otherwise, $d \neq 0$.

Since $d \neq 0$, it must be by Lemma 6.49 that there exists an atom a' such that $0 \leq a' \leq d$. We show that this atom cannot be in $\{a_1, \dots, a_n\}$. Indeed, $d \wedge a_i = (b \wedge \neg c) \wedge a_i = b \wedge (\neg c \wedge a_i) = b \wedge (\bigwedge_{j=1}^n \neg a_j \wedge a_i)$ and since $\neg a_i \wedge a_i = 0$, this expression cancels to 0 by Lemma 6.46. So $a_i \not\leq d$ for every i, and therefore $a' \notin \{a_1, \dots, a_n\}$. But since $d \leq b$ and $a' \leq b$, we get $a' \leq b$, so a' is an atom smaller than b and not in $\{a_1, \dots, a_n\}$, contradicting the choice of $a_1, \dots, a_n$. □

Corollary 6.52. *Let $(B, \vee, \wedge, \neg)$ be a finite Boolean algebra with atoms $S = \{a_1, \dots, a_n\}$. Then $(B, \vee, \wedge, \neg)$ is isomorphic to $(\mathcal{P}(S), \cup, \cap, S \setminus \_)$. In particular, B has size 2^n.*

Proof. Define a map $f \colon \mathcal{P}(S) \to B$ by mapping $\{a_{i_1}, \dots, a_{i_n}\}$ to $a_{i_1} \vee \cdots \vee a_{i_n}$ (in particular $\emptyset$ is mapped to 0). This function is surjective by Proposition 6.51: every $b \in B \setminus \{0\}$ can be written as $a_{i_1} \vee \cdots \vee a_{i_n}$, where $a_{i_1}, \dots, a_{i_n}$ list all the atoms that are smaller than b, and therefore $f(\{a_{i_1}, \dots, a_{i_n}\}) = b$. Moreover $f(\emptyset) = 0$, so every element in B has a preimage under f.

The function is also injective: suppose that $a_{i_1} \vee \cdots \vee a_{i_n} = a_{j_1} \vee \cdots a_{j_m}$, for some atoms $a_{i_1}, \dots, a_{i_n}, a_{j_1}, \dots, a_{j_m}$. Denote this element of B by b. For every $a \in \{a_{i_1}, \dots, a_{i_n}\}$, we get $a \wedge b = a \wedge (a_{i_1} \vee \cdots \vee a_{i_n}) = a$, and by distributivity also $a \wedge b = \bigvee_{r=1}^m (a \wedge a_{j_r})$. If $a \notin \{a_{j_1}, \dots, a_{j_m}\}$, then $a = \bigvee_{r=1}^m (a \wedge a_{j_r}) = 0$, a contradiction to the fact that a is an atom and therefore is non-zero. Thus, $a \in \{a_{j_1}, \dots, a_{j_m}\}$ and we conclude that $\{a_{i_1}, \dots, a_{i_n}\} \subseteq \{a_{j_1}, \dots, a_{j_m}\}$. Doing a similar reasoning in the other direction, we get $\{a_{i_1}, \dots, a_{i_n}\} = \{a_{j_1}, \dots, a_{j_m}\}$. This means that f is injective.

It remains to prove that f is an isomorphism, that is, that $f(A \cap B) = f(A) \wedge f(B)$, $f(A \cup B) = f(A) \vee f(B)$, and $f(S \setminus A) = \neg f(A)$. We have $f(A) \wedge f(B) = \left(\bigvee_{a \in A} a\right) \wedge \left(\bigvee_{b \in B} b\right)$ which by distributivity is equal to $\bigvee_{a \in A, b \in B}(a \wedge b)$. Since $a \wedge b = 0$ for any two different atoms a, b, and $a \wedge b = a$ if $a = b$, the only non-zero terms are the ones where $a = b \in A \cap B$. Thus we get $\bigvee_{a \in A \cap B} a$, which is $f(A \cap B)$.

The fact that $f(A \cup B) = f(A) \vee f(B)$ is simpler: $f(A) \vee f(B) = \left(\bigvee_{a \in A} a\right) \vee \left(\bigvee_{b \in B} b\right) = \bigvee_{a \in A \cup B} a = f(A \cup B)$.

Finally, note that $f(\emptyset) = 0$ and $f(S) = 1$. Thus, $f(A) \wedge f(S \setminus A) = f(A \cap (S \setminus A)) = f(\emptyset) = 0$ and $f(A) \vee f(S \setminus A) = f(A \cup (S \setminus A)) = f(S) = 1$, so by Lemma 6.44, we have that $f(S \setminus A)$ is the complement of $f(A)$, i.e., $f(S \setminus A) = \neg f(A)$. This concludes the proof. □

6.8 Exercises

1. Show that exponentiation, defined as a function $f(a, b) := a^b$, is not an internal composition law on the set $\mathbb{Q}$ of rational numbers.

2. Determine whether the operation $\otimes$ from Example 6.4 is associative or commutative.

3. Show that the usual laws of exponentiation hold in a group G: $a^p \otimes a^q = a^{p+q}$ and $(a^p)^q = a^{pq}$ for all $p, q \in \mathbb{Z}$ and $a \in G$.

4. Let $(G, \otimes)$ be a group with neutral element e_G. Show that $(e_G)^{-1} = e_G$.

5. Let $(G, \otimes)$ be a group. Show that $(a \otimes b)^{-1} = b^{-1} \otimes a^{-1}$.

6. Let $(\mathbb{Z}/3\mathbb{Z})^\times$ be the set of elements $[a] \in \mathbb{Z}/3\mathbb{Z}$ where a is coprime with 3. Let $+$ be the addition of classes: $[a] + [b] = [a + b]$. Determine whether $((\mathbb{Z}/3\mathbb{Z})^\times, +)$ is a group.

7. Compute the Cayley table for the group $((\mathbb{Z}/10\mathbb{Z})^\times, \times)$.

8. Show that the relation $\sim$ defined in the proof of Theorem 6.13 is indeed an equivalence relation.

9. Let $n \geq 1$, and let $M_n(\mathbb{R})$ be the set of $n \times n$ matrices with coefficients in $\mathbb{R}$. Is $(M_n(\mathbb{R}), +)$ a group? Is $(M_n(\mathbb{R}), \times)$ a group? Are these operations associative or commutative?

10. Show that $(ab)^n$ is not always equal to $a^n b^n$ in a group (i.e., give an example of a group G, elements $a, b \in G$, and $n \geq 2$ for which the equality fails).

11. Show that if $(G, \otimes)$ is an *Abelian* group (i.e., $\otimes$ is commutative) then $(ab)^n = a^n b^n$ for all $a, b \in G$ and $n \in \mathbb{Z}$.

12. Let A be a nonempty set. Is $(\mathcal{P}(A), \cup)$ a group? What about $(\mathcal{P}(A), \cap)$ and $(\mathcal{P}(A), \Delta)$?

13. Let $f\colon A \to A$ be a bijection, where A is a finite set of size $n \geq 1$. Show that the composition $f \circ f \circ \cdots \circ f$ of f with itself $n!$ times is equal to id_A.

14. Show that the composition of two homomorphisms is a homomorphism.

15. Enumerate all the homomorphisms from $(\mathbb{Z}/3\mathbb{Z}, +)$ to $(\mathbb{Z}/2\mathbb{Z}, +)$. Similarly, enumerate all the homomorphisms from $(\mathbb{Z}/3\mathbb{Z}, +)$ to itself.

16. (*) Show that if $a \in G$ has order n and $h\colon (G, \otimes_G) \to (H, \otimes_H)$ is a homomorphism, then $h(a)$ has an order m that divides n.

17. (*) Let $(G, \otimes_G)$ and $(H, \otimes_H)$ be two finite groups such that $|G|$ and $|H|$ are coprime. Show that the only homomorphism $(G, \otimes_G) \to (H, \otimes_H)$ is the map such that $f(x) = e_H$ for all $x \in G$.

18. Show that in Example 6.16, $(G, \times)$ forms a group.

19. Find a set G of 4 matrices such that $(\mathbb{Z}/4\mathbb{Z}, +)$ is isomorphic to $(G, \times)$, where $\times$ is the multiplication of matrices.

20. Find a set of complex numbers $z_0, z_1, z_2 \in \mathbb{C}$ such that $(\mathbb{Z}/3\mathbb{Z}, +)$ is isomorphic to $(\{z_0, z_1, z_2\}, \times)$.

21. Suppose that $f\colon (G, \otimes_G) \to (H, \otimes_H)$ is a homomorphism. Prove that the map g obtained from Theorem 3.22 is a group homomorphism $(G/\ker(f), *) \to (H, \otimes_H)$.

22. (*) Let G be a finite subset of $\mathbb{C}$, the set of complex numbers. Suppose that $(G, \times)$ is a group, where $\times$ denotes multiplication of complex numbers. Show that for every $g \in G$, one has $|g| = 1$, where $|\cdot|$ denotes the complex modulus. Show that there exists $n \in \mathbb{N}$ such that every g is of the form $e^{2ki\pi/n}$ for some $k \in \mathbb{N}$.

23. Show that if $(R, \oplus, \otimes)$ is a ring, then $0_R \otimes r = 0_R$ for every $r \in R$, and that $(-r) \otimes s = -(r \otimes s)$ for all $r, s \in R$.

24. Show that if $(R, \oplus, \otimes)$ is a ring with neutral elements 0 and 1, then 0 does not have an inverse in $(R, \otimes)$. (Use the fact that we assume $0 \neq 1$.)

25. Suppose that $(R, \oplus_R, \otimes_R)$ and $(S, \oplus_S, \otimes_S)$ are rings, and that $f\colon (R, \oplus_R, \otimes_R) \to (S, \oplus_S, \otimes_S)$ is a homomorphism of rings. Prove that f is also a group homomorphism $(R, \oplus_R) \to (S, \oplus_S)$.

26. Suppose that $(R, \oplus_R, \otimes_R)$ and $(S, \oplus_S, \otimes_S)$ are *fields*, and that $f\colon (R, \oplus_R, \otimes_R) \to (S, \oplus_S, \otimes_S)$ is a homomorphism. Prove that f must be injective.

27. Let $(\mathbb{K}, +, \times)$ be a field. Show that if $a \times b = 0_{\mathbb{K}}$, then $a = 0_{\mathbb{K}}$ or $b = 0_{\mathbb{K}}$.

28. Let $f\colon (R, \oplus_R, \otimes_R) \to (S, \oplus_S, \otimes_S)$ be an isomorphism between rings. Show that if $(R, \oplus_R, \otimes_R)$ is a field, then $(S, \oplus_S, \otimes_S)$ is a field, too.

29. (*) Let $(\mathbb{K}, +, \times)$ be a finite field, and let n be the order of 1 in the group $(\mathbb{K}, +)$, that is, n is the smallest integer such that $1 + 1 + \cdots + 1 = 0$, where 1 appears n times in the sum. (Such an n is known to exist by Theorem 6.13.) Prove that n must be a prime number called the *characteristic* of $\mathbb{K}$.

characteristic

30. (**) Let $\mathbb{K}$ be a finite field of characteristic p, and remember that since p is prime $\mathbb{Z}/p\mathbb{Z}$ is a field. For $\lambda \in \mathbb{Z}/p\mathbb{Z}$ and $a \in \mathbb{K}$, define $\lambda \cdot a$ to be $(1+1+\cdots+1) \times a$, where 1 appears λ times in the sum (although strictly speaking λ is an equivalence class of numbers modulo p, the resulting expression $(1 + \cdots + 1) \times a$ does not depend on the choice of a particular number in that class, by definition of the characteristic). Convince yourself that this gives $\mathbb{K}$ a structure of vector space over the field $\mathbb{Z}/p\mathbb{Z}$. Conclude that $\mathbb{K}$ must have size p^n, where n is the dimension of this vector space.

31. Let $R = \{0, 1\}^2$. Consider the composition laws on R defined by $(a, b) + (c, d) = (a +_2 c, b +_2 d)$ and $(a, b) \times (c, d) = (ac +_2 bd, bc +_2 ad +_2 bd)$ (where $+_2$ is addition taken modulo 2, i.e. $1 +_2 1 = 0$). Show that $(R, +, \times)$ is a field.

32. Show that $X^n = (0, \ldots, 0, 1, 0, \ldots, 0)$ (where the 1 is in position $n + 1$), where X is the polynomial $(0, 1, 0, \ldots) \in R[X]$.

33. Prove Theorem 6.32.

34. Prove Proposition 6.36.

35. Let $(B, \vee, \wedge, \neg)$ be a Boolean algebra and let $a \in B$. Show that $\neg(\neg a) = a$. Show that $\neg 0 = 1$. (Hint: use Lemma 6.44.)

36. Show that the following holds in any Boolean algebra and for any a, b, c: if $a \wedge b = a \wedge c$ and $\neg a \wedge b = \neg a \wedge c$, then $b = c$.

37. Suppose that f is a homomorphism of Boolean algebras $(B, \vee_B, \wedge_B, \neg_B) \to (C, \vee_C, \wedge_C, \neg_C)$, and let $0_B, 1_B, 0_C, 1_C$ be the respective neutral elements. Show that $f(0_B) = 0_C$ and $f(1_B) = 1_C$.

38. Let $(B, \vee, \wedge, \neg)$ be a Boolean algebra and let $a, b \in B$. Show that $a \wedge b = a$ if, and only if, $a \vee b = b$.

39. Let $(B, \vee, \wedge, \neg)$ be a Boolean algebra. Prove that if a, b are two different atoms, then $a \wedge b = 0$.

40. Let $B = [0, 1]$ (the closed interval in $\mathbb{R}$), and define $\neg x = 1 - x$ for all $x \in [0, 1]$. Is $(B, \max, \min, \neg)$ a Boolean algebra?

Index

List of Figures